Received: 6 May 2022 | Revised: 13 September 2022 | Accepted: 16 September 2022

DOI: 10.1002/psp4.12870

MINI-REVIEW

Comparing the applications of machine learning, PBPK, and population pharmacokinetic models in pharmacokinetic drug–drug interaction prediction

Jaidip Gill[1] | **Marie Moullet**[1] | **Anton Martinsson**[2] | **Filip Miljković**[2] | **Beth Williamson**[3] | **Rosalinda H. Arends**[4] | **Venkatesh Pilla Reddy**[1]

[1]Clinical Pharmacology and Quantitative Pharmacology, Clinical Pharmacology and Safety Sciences, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK

[2]Imaging and Data Analytics, Clinical Pharmacology & Safety Sciences, Biopharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[3]Oncology DMPK, Oncology R&D, AstraZeneca, Cambridge, UK

[4]Clinical Pharmacology and Quantitative Pharmacology, Clinical Pharmacology and Safety Sciences, Biopharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, USA

Correspondence
Venkatesh Pilla Reddy, Clinical Pharmacology and Quantitative Pharmacology, Clinical Pharmacology and Safety Sciences, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK.
Email: venkatesh.reddy@astrazeneca.com

Present address
Beth Williamson, DMPK, UCB, Surrey, UK

Rosalinda H. Arends, Bioinformatics & Data Science, Exelixis, Alameda, CA, USA

Abstract

The gold-standard approach for modeling pharmacokinetic mediated drug–drug interactions is the use of physiologically-based pharmacokinetic modeling and population pharmacokinetics. However, these models require extensive amounts of drug-specific data generated from a wide variety of in vitro and in vivo models, which are later refined with clinical data and system-specific parameters. Machine learning has the potential to be utilized for the prediction of drug–drug interactions much earlier in the drug discovery cycle, using inputs derived from, among others, chemical structure. This could lead to refined chemical designs in early drug discovery. Machine-learning models have many advantages, such as the capacity to automate learning (increasing the speed and scalability of predictions), improved generalizability by learning from multicase historical data, and highlighting statistical and potentially clinically significant relationships between input variables. In contrast, the routinely used mechanistic models (physiologically-based pharmacokinetic models and population pharmacokinetics) are currently considered more interpretable, reliable, and require a smaller sample size of data, although insights differ on a case-by-case basis. Therefore, they may be appropriate for later stages of drug–drug interaction assessment when more in vivo and clinical data are available. A combined approach of using mechanistic models to highlight features that can be used for training machine-learning models may also be exploitable in the future to improve the performance of machine learning. In this review, we provide concepts, strategic considerations, and compare machine learning to mechanistic modeling for drug–drug interaction risk assessment across the stages of drug discovery and development.

THE SIGNIFICANCE OF DRUG–DRUG INTERACTION MODELING

With an aging population, polypharmacy has become widespread, with an average prevalence of ~32% in elderly adults across Europe.[1] The direct impact is that patients are at increased risk of drug–drug interactions (DDIs). The affected drug is known as the "victim," whereas the drug modulating a change in the other is known as the "perpetrator." Pharmacokinetic (PK) DDIs are a type of DDI caused by shared absorption, distribution, metabolism, or excretion (ADME) pathways, leading to changes in drug exposure.[2] This altered effect can lead to negative consequences in health care. For example, the exposure may fall below efficacious levels. If the exposure increases, then there is an increased risk of adverse drug events (ADEs). Therefore, prediction of PK DDIs is important in preventing loss of efficacy and avoiding ADEs.

One approach to PK DDI prediction is mechanistic modeling.[3] This approach can either be static (snapshot in time and incorporate in vitro data and predicted [or measured] clinical exposure) or dynamic (integration of drug-dependent and system-dependent parameters over time). Currently, two of the standard approaches to dynamic modeling of DDIs are physiologically-based pharmacokinetic (PBPK) and population pharmacokinetic (Pop-PK) models. These use differential equations to describe drug PKs based on known biological processes. Another approach is machine learning (ML), which utilizes statistical relationships between variables to make predictions. The use of ML in drug development has increased over recent years and has already been applied to various parts of the drug discovery pipeline,[4] such as target identification and validation, small molecule design, biomarker discovery, and pathology prediction. The two approaches will be compared in this review to determine their respective applications in DDI modeling.

PBPK/POP-PK MODELING

PBPK models predict the changes in PK profiles of a drug prior to clinical use, making it useful in identifying the key DDI risks before first use in humans. The input parameters describe properties which may be intrinsic to the drug (e.g., solubility, fraction unbound in the blood [f_{ub}], in vitro intrinsic clearance, and permeability) or extrinsic (e.g., patient renal impairment).[3] These parameters are then used to make predictions on the PK characteristics of the drug, representing a mechanistic, bottom-up approach to modeling based on the underlying pharmacological properties of the drug. The parameters and predictions of the model can then be verified in the clinic and data from clinical studies can then be used to refine the model, representing a middle-out approach.[3]

In contrast, Pop-PK models utilize a top-down approach, where sources of clinical data are required with the appropriate sample numbers to be able to understand the variability in PKs between and within target populations identified. For example, a clinical nested DDI study is typically evaluated using Pop-PK analysis.[5] In some cases, the Pop-PK analysis plan for DDI assessment must be established prior to conducting the clinical study. The DDI is investigated as a categorical covariate (e.g., comedication effect on drug's clearance or volume of distribution). The categorical value can change over the time course of the DDI, so the status of the DDI over time must also be considered.[6] The precision of a covariate effect depends on the number of subjects in the study. Both approaches enable simultaneous modeling of DDIs and other PK properties.

MACHINE-LEARNING MODELING

Supervised ML is an approach to create mathematical algorithms to predict an outcome based on input variables. In the context of PK DDI modeling, supervised ML is often utilized where the outcome label is exposure and the predictors are in vitro and in vivo ADME data or physicochemical properties. Features are the predictor (independent) variables, whereas the label is a dependent variable that we predict. These models can be separated into regression and classification models: regression models predict continuous labels; and classification models predict discrete labels. The key difference between ML and PBPK/Pop-PK is that ML models automatically re-establish mathematical relationships between training samples with each data update, whereas PBPK/Pop-PK requires manual derivation of differential equations to relate different PK parameters.

ML modeling requires choosing a function/algorithm that connects the input features to the outcome of the dataset, in this case, DDI ratios.[7] For example, linear regression models define a linear function that calculates the labels (e.g., area under the drug concentration curve [AUC] and maximum drug concentration [C_{max}]), in terms of the features (e.g., drug chemical properties and system parameters). On the other hand, decision trees can be used to make predictions for each input sample using rule-based decisions on the features to predict the label. In all cases, the model must be fit to the training data by selecting parameter values that lead to optimal performance. The optimal parameters are those that lead to the lowest loss (prediction error). These algorithms also have hyperparameters – user-defined parameters. The choice of hyperparameters also depends on minimizing the "loss."[7] Therefore, the most accurate

prediction of DDIs depends partially on optimal selection of the parameters and hyperparameters used for a given ML algorithm. The strengths/weaknesses of ML will be compared to PBPK/Pop-PK models below.

ML has previously been applied to DDI prediction using features describing chemical structure. For example, a deep neural network model used structural similarity profiles to predict 86 distinct types of DDIs, including the prediction of possible causal mechanisms for the DDI.[8] Another method improved upon this by accounting for the mechanism of interaction between drugs, meaning the model represents DDIs as a function of the relevant functional groups rather than other overlapping substructures that are irrelevant for the mechanism of interaction. This was done by generating coefficients that represent the relevance of a substructure to a DDI.[9] These methods can be applied in early drug discovery due to being dependent only on chemical structure as an input. Their use-case may therefore be to aid in selecting drug candidates in early discovery by prioritizing drugs that have a limited chance of interaction with an established co-medication based on structural similarity.

Structure-based approaches cannot describe the clinical significance of the DDIs because they do not factor in dose. In addition, they do not utilize the extensive PK data generated later in development that could otherwise improve predictive accuracy. Therefore, ML models that utilize knowledge graphs have been developed for late-stage drug development DDI prediction. These knowledge graphs contain rich PK information for a given drug, which is utilized as features for DDI prediction. Such approaches have resulted in high classification accuracy and even the prediction of rare side effects caused by a DDI.[10] Other models have demonstrated that combining structural and knowledge graph feature sets can enable improved performance than using either alone.[11] The application of ML models to dose adjustment can be enhanced by the use of regression-based models that predict fold-changes in PK values. Development of such models has begun, with one model capable of predicting 94.8% of AUC fold-changes within two-fold of the observed AUC fold-change.[12] This information could then be used by drug design, drug metabolism, and PK groups to determine whether the DDI risk should be further investigated, or later in development whether the co-administration of drugs should be avoided.

COMPARISON OF DATA REQUIREMENTS

PBPK models require in vitro experimental data (e.g., inhibition constants, plasma protein binding, and induction parameters) and predicted or measured clinical in vivo data. Their production efficiency for new drug candidates is therefore limited by the requirement to run these experiments before the model can be refined and appropriate for use.

In contrast, domain knowledge is used to select features for an ML model but may not be subject to the same data requirements as PBPK/Pop-PK models. ML models can be trained on existing drug candidate using their known DDIs as label data, and relevant features available earlier on in the drug development cycle (Figure 1). The optimized and validated ML models can then be applied to new drug candidates without the need for later-stage drug development data. For example, non-experimental data, such as chemical descriptors, can be used as features available early in discovery. Transforming these basic chemical features into more advanced descriptors, such as similarity profiles, may lead to even stronger results. For example, feature vectors called structural similarity profiles have been derived from simplified molecular-input line-entry system (SMILES),[8] which represent chemical structure as a string. Scores for each combination of drugs were generated using the Tanimoto coefficient with extended-connectivity fingerprints of the drugs. The structural similarity profile feature vector for a drug was then built from an array of all the similarity scores for that drug with other drugs in the database. This model yielded scores of ~90% on different classification scoring metrics, outperforming the use of chemical descriptors as a feature set.[8] Other models have been built on more expansive features sets, for example, drug pair DDI probability can be calculated based on similarity to known DDI-inducing pairs considering compound chemical structure, known protein–protein interactions, and PK pathways.[13] The model was built on a gold-standard DDI database, including information on whether drugs were metabolized by the same CYP enzyme and returns the DDI-mediating CYP. Known interactions among drugs, enzymes, and transporters can also be represented as interaction networks. Such data have been leveraged for DDI prediction, because a DDI is more likely to occur if a putative perpetrator is structurally similar to drugs which bind to the enzymes/transporters interacting with a putative victim.[14]

Natural Language Processing (NLP) has also been utilized to extract DDI information from existing literature. An approach in 2013 analyzed biomedical literature (DrugBank entries and MedLine abstracts) to build a model which classifies DDI with an F1-score of 0.65 (the harmonic mean of the precision and recall of the model),[15] a score which was improved to 0.85 using the same dataset by implementation of deep neural networks in 2018.[16] This rapid improvement in performance highlights how advances in neural network architecture from the graph ML space can yield improved predictive power when applied to DDI prediction. The

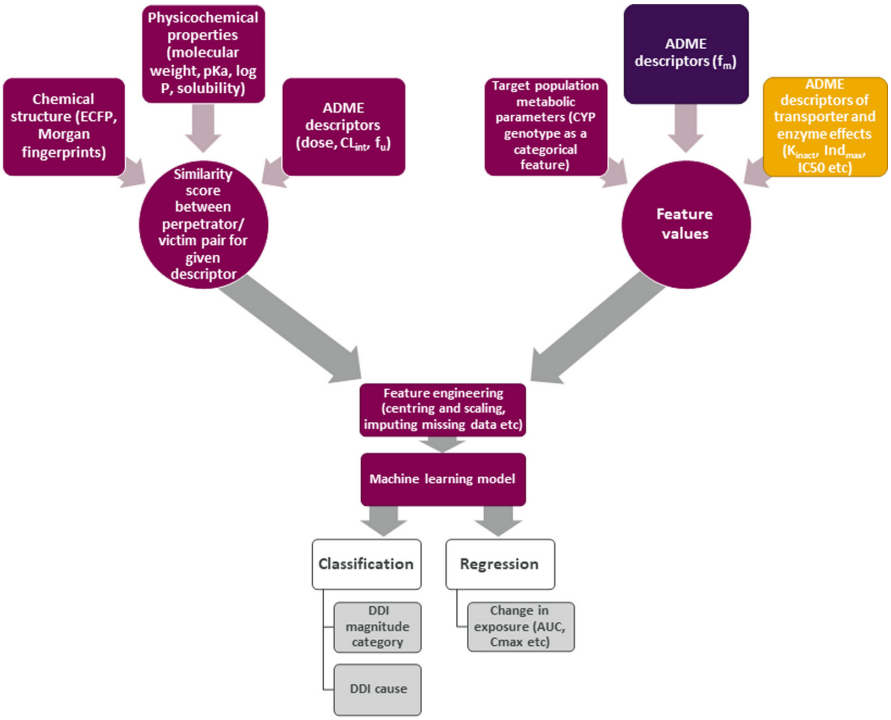

FIGURE 1 A summary of features and labels for current machine learning models for DDI. The inputs (features) for prediction shown have been utilized in the literature for machine learning-based prediction of DDIs. They are grouped into colors based on whether they apply to the victim (purple), perpetrator (orange) or both (red). Chemical feature-based models would enable DDI prediction at the point of design, whereas models using ADME descriptors and physicochemical properties would require drug in vitro data. Most machine learning models developed for PK DDIs thus far have been classification-based, rather than regression, and so the example labels for regression are only speculative. These classification models can predict the cause of a DDI or the magnitude of a PK DDI using thresholds defined by the FDA.[5] ADME, absorption, distribution, metabolism, and excretion; AUC, area under the drug concentration curve; C_{max}, maximum drug concentration; CL_{int}, intrinsic clearance; DDI, drug–drug interaction; ECFP, extended connectivity fingerprints; FDA, US Food and Drug Administration; f_m, fraction metabolized; f_u, fraction unbound in plasma; IC_{50},half maximal inhibitory concentration; Ind_{max}, induction maximum; K_{inact}, maximum rate of enzyme inactivation.

ability to extract DDI information from the literature using NLP could enable less established DDIs to be detected, and increase ML training dataset sizes without the additional cost of conducting new DDI studies.

The optimization of a diverse set of features, from chemical similarity to text-based literature, provides new use-cases for ML models across the entire drug development pipeline, whereas PBPK/Pop-PK models are limited to when extensive in vitro and in vivo data is available (especially Pop-PK which requires clinical data). Additionally, PBPK/Pop-PK models only use experimental data pertaining to the specific drugs involved, whereas ML models can utilize experimental data from other DDIs to make predict new DDIs. However, ML models can show poor performance when predicting the effects of drug combination for novel drugs.[17] A key consideration for ML models when investigating new drug combinations is whether these interactions fall beyond the applicability domain and the challenge of generalizing outside the current known chemical space. Regardless, ML models would still be of great practical value early to triage down DDIs with appropriate safety, which would later be consolidated by more case-focused mechanistic models.

ML models are vulnerable to the effects of biases and correlations in their training data. Correlations between features can lead to performance issues, so feature selection methods have long been developed to address this.[18] Data bias is another problem that can affect ML models. For example, classification models, which use samples that are labeled with a specific class, can show a prediction bias toward a specific class if it is over-represented in the dataset compared with other classes.[19] To address this, artificial sampling techniques have been developed to increase the number of samples of the minority class, such as synthetic minority oversampling technique.[20] This has been applied in the neighboring field of drug-target interaction modeling leading to large improvements in performance upon balancing.[21] Although this technique can address class imbalance in binary classification tasks, it shows poorer performance in multiclass problems.[22] In the context of PK DDI modeling, this may mean that class imbalance can be resolved for binary tasks, such as prediction of whether a

DDI will occur or not, but may remain a problem for multiclass tasks, such as classifying the DDI type (for example, classifying the magnitude of change in exposure). In these cases, the ML task can be turned into a regression approach if appropriate continuous labels can be assigned. Although mechanistic models avoid this problem as they are not classifiers, they may suffer from other biases (e.g., if the derived equations are trained and verified on only a specific subset of physiological conditions). Overall, ML models may be used to make predictions when experimental data may not have been produced yet, but care must be taken to ensure input datasets have minimal bias to achieve maximal performance. However, when experimental data are available for a specific drug pair, PBPK models should be preferred to ML models due to their increased interpretability.

To improve the interpretability of ML models used in DDI prediction, methods that explain how a model prediction was made could be utilized. For example, "Shapley additive explanations" describe the feature importance of individual predictions made by a model (including more complex deep learning methods) in a manner that is consistent with explanations made by humans.[23] Therefore, the effect of features such as dose and important ADME parameters could be understood for each DDI prediction made, enabling pharmacologists to identify and troubleshoot poor predictions.

The predictions of models trained on preclinical data depends on the accuracy of input experimental data and outcome labels. Uncertainty around parameter estimates can be incorporated using Bayesian methods to describe parameters as being drawn from a distribution (rather than point estimates) and subsequently propagating this uncertainty to model outputs. In ML, explainable artificial intelligence tools exist to describe uncertainty surrounding model predictions, such as probabilistic methods (which describe the posterior probability of a model output) or ensemble methods (which find the variability between multiple model predictions).[24] An example from the field of protein-binding prediction showed that using both point estimates and their standard deviations for model training improves performance when there is high experimental variability, especially for measurements close to categorization thresholds.[25] These approaches could be extended to incorporate uncertainty around human PK parameters, which are extrapolated from preclinical data.

COMPARISON OF PERFORMANCE

Because ML models are founded on statistical patterns between inputs and outputs, they have an inductive capability.[26] This means they can identify patterns in observed data and use these patterns to build models that are predictive within the range of the training data. On the other hand, mechanistic models have deductive capability,[26] so they can use the mechanistic relationships between variables to make predictions on data that may be outside of the range of data used to build the model, yet within the particular DDI case domain. This means that mechanistic models are much more suited to extrapolation of new data ranges when compared with ML models, because the statistical relationships are more likely to change outside of the training data range. Therefore, in the context of PK DDIs, mechanistic models may be more useful for modeling instances involving abnormal parameters, such as specific disease populations, because the statistical relationships observed in the average population may not be applicable to these niche populations. ML models could be generated for these specific subgroups, for example, by encoding the distinguishing features of the subgroups as additional features, but there may be insufficient data to build these more specific models (because ML models require large datasets),[26] whereas this is not a limitation for mechanistic models due to their lower sample size requirements.

The discussion so far has assumed the validity of the relationships described in mechanistic models. However, PBPK models utilize the extrapolation of in vitro experimental data to make predictions in vivo.[2] Because in vivo conditions are different to in vitro, models built using in vitro data can fail to capture in vivo conditions. In contrast, ML models built using drug-inherent features (such as chemical structure) and outcome labels from human data can be applied to novel compounds in the absence of clinical data – the relationship between structure and DDI risk is not readily interpretable but because the implicit relationship was learned on human data the implicit relationship between chemical structure and DDI risk still holds true. Because the reliance of PBPK models on in vitro data could limit the performance of PBPK models, it would be beneficial to utilize experimental data from models more representative of physiological conditions in humans. For example, organ-on-a-chip technology was used to produce a duodenum intestine-chip system that more accurately models drug PK affects, such as changes in CYP activity compared with established in vitro methods (e.g., Caco-2 cultures).[27] Of course, data from these systems could also be incorporated into ML models, with the potential to improve performance. The overall strengths and weaknesses for ML and PBPK/Pop-PK modeling approaches are summarized in Table 1.

COMPARISON OF CURRENT AND POTENTIAL APPLICATIONS

Integration of these methods can help accelerate drug development and support decision making across the

TABLE 1 Comparison of PBPK/Pop-PK models and machine learning models used in pharmacokinetic predictions

Characteristic	PBPK	Pop-PK	Machine learning
Methodology	A type of PK modeling that models organs/tissues as compartments as well as the blood flow between them, utilizing a "bottom-up" approach to modeling built on in vitro data and in vivo data first. The model takes drug-specific and system-specific specific parameters as input to make predictions for PK parameters. Plasma concentration curves and in vitro inhibition and induction parameters are used to predict the effects of perpetrators.	Also compartment-based, but utilizes a "top-down" approach where a descriptive model is built that fits the observed data first. Retrospective estimation of PK parameters is performed after population studies, used to make statistical adjustments to predictive models. DDI effects are described rather than predicted, using observed data from the victim drug in the DDI.	Mathematical algorithms that describe the statistical relationships and patterns between features and labels, and use these relationships to make predictions about new samples. These algorithms can either act as classifiers or regressors.
Data requirements for improved performance	In vivo mass balance data. Incorporation of sensitivity analysis to factor in possible discrepancies between in vivo and in vitro parameters, such as the perpetrator K_i.[28]	Longitudinal/time-varying investigation of DDI as a categorical covariate[6]	Large sample sizes for observed DDIs, especially for deep learning where n should be >1000 (ref)
Examples of software used	SimCYP, Monolix, R	NONMEM	Python (scikit-learn, TensorFlow, PyTorch), R (CARET)
Strengths	Deductive capability, enabling extrapolation of predictions to data ranges outside of training data. Can be implemented on smaller datasets.	Descriptive capability, enabling estimation of the sources of variability in PKs of a drug in a target population, such as concomitant drug use.	Inductive capability, enabling detection of unseen patterns and relationships in training data. Less human intervention required when updating the model compared to PBPK/Pop-PK. Can be implemented for predicting a variety of different tasks very quickly, as long as the required feature data is available. Preclinical stage: Can estimate PK DDI risk at the point of design using chemical structure-based feature sets. Postmarket monitoring stage: Could be used for cases where data limitations limit PBPK/Pop-PK accuracy (e.g., poor elimination characterization)
Limitations	Requires a wide variety of experimental data. Requires more time and human intervention to develop because the relationships must be explicitly described by the modeler themselves. May perform poorly when used with less characterized drugs/enzymes.	Only descriptive – less appropriate for extrapolations compared to PBPK.	Initially requires large sample sizes and computational resources for good performance, which may not always be available. Can be difficult to interpret (especially nonlinear models and neural networks).

Abbreviations: DDI, drug-drug interaction; PBPK, physiologically-based pharmacokinetic; PK, pharmacokinetic; Pop-PK, population pharmacokinetic.

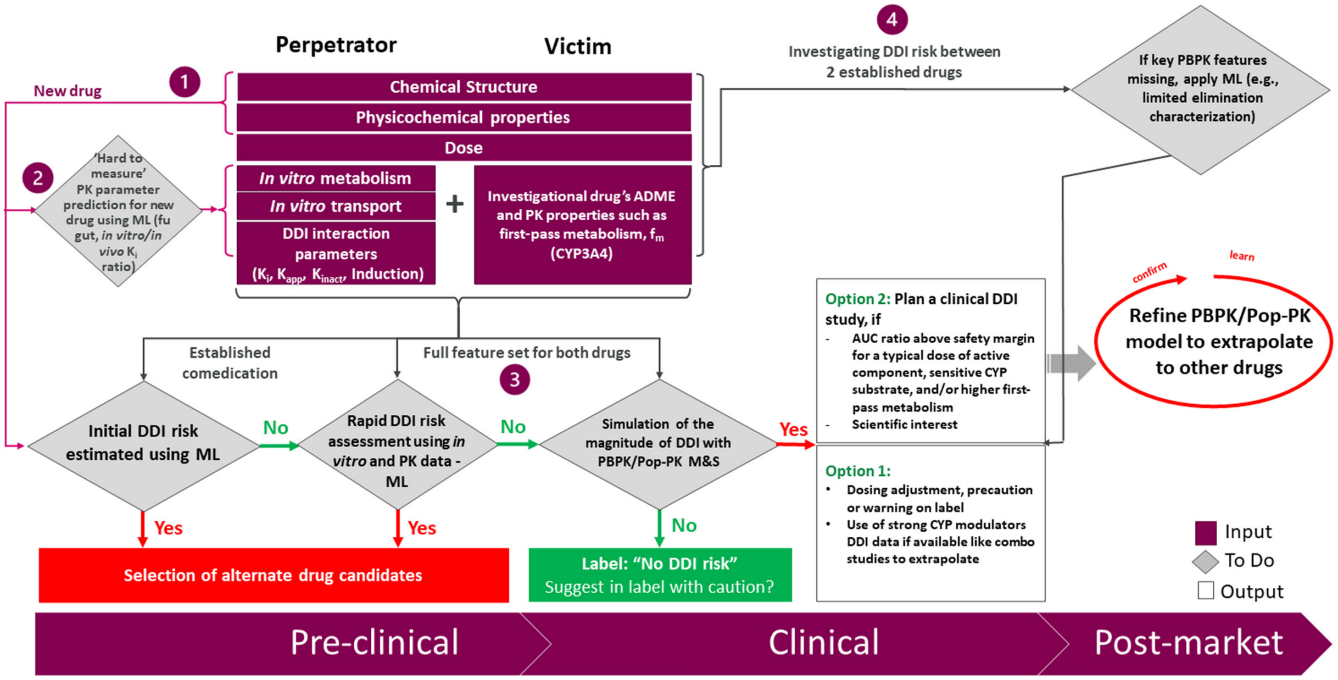

FIGURE 2 Potential workflow for DDI assessment. ML has potential to be used for PK DDI risk assessment across the drug development timeline. (1) In preclinical stages, chemical and physicochemical features could be used to classify DDI risk at the point of design of a new drug when used in combination with an established co-medication. (2) ML can be used to predict PK/ADME parameters required for PBPK/Pop-PK modeling in the cases where such parameters are hard to determine (e.g., f_u gut) using chemical and physicochemical features (as well as any other PK features that are available for a given drug). (3) After appropriate experiments are performed, additional PK parameters and dosing data could be used to more precisely estimate DDI risk using regression-based ML models, but before the more time consuming PBPK and Pop-PK models are utilized. This could aid in the decision to filter out drugs with high DDI risk. (4) In the postmarket stage, ML can be used to predict DDI risk between established drugs in cases where there is limited knowledge of parameters required for PBPK modeling. ADME, absorption, distribution, metabolism, and excretion; AUC, area under the drug concentration curve; DDI, drug–drug interaction; f_u, fraction unbound in plasma; IC_{50},half maximal inhibitory concentration; Ind_{max}, induction maximum; K_{inact}, maximum rate of enzyme inactivation; M&S, modeling and simulation; ML, machine learning; PBPK, physiologically-based pharmacokinetic; PK, pharmacokinetic; Pop-PK, population pharmacokinetic.

drug-discovery pipeline. PBPK and Pop-PK models are already extensively used in the DDI-risk assessment process, but ML is not. An overview showing how the addition of ML could be used in addition to PBPK and Pop-PK models for DDI assessment is shown in Figure 2. The ability of ML models to make DDI predictions based on shared chemical structure would allow potential risks to be identified before the compound is synthesized, reducing the number of design-make-test-analyze cycles and prioritizing progression of drugs with no chemical basis for a potential DDI. Models incorporating different possible doses could be applied later on when initial dose estimates are available, enabling prediction of the clinical effects of potential DDIs. This would then enable regression-based estimation of PK DDI effects and dose adjustment.

More generally, predictions made during drug discovery would allow project teams to benchmark against the competitive landscape. As a compound progresses through lead optimization and data are generated to profile the compound, the use of mechanistic models in addition to ML models would provide a more comprehensive early risk assessment for target patient populations. Accuracy improvements should also be explored for later-stage PK DDI risk-assessment by testing ML models that incorporate in vivo and clinical feature data. ML could also be used to detect DDIs in the postmarket monitoring stage if certain PK data is poorly established (e.g., incomplete characterization of elimination of the victim or perpetrators that target the enzyme/transporter involved).

COMBINING ML AND PBPK/POP-PK

Finally, whereas we have so far considered clinical pharmacology and ML as two distinct and parallel approaches, the two methods may also be combined to complement each other. The experimentally derived parameters used in PBPK models may be difficult to measure. For example, in vitro estimates of K_i can differ by 10-fold compared to the in vivo K_i of inhibitors.[28] Therefore, ML models

are being developed to predict such parameters with improved accuracy. For example, a neural network approach enabled prediction of K_i, K_d, and half-maximal inhibitory concentration (IC_{50}) of protein-ligand complexes with high accuracy using chemical structures.[29] Therefore, development of ML models that predict individual ADME parameters may aid in building PBPK-based prediction of DDIs involving reversible inhibitors.

Conversely, PBPK models could be used to generate additional parameters that could be used as input features for the ML model. The algorithms could then detect these previously unseen patterns and information from the mechanistic model output and thus be used to make better predictions.[26] To our knowledge, this type of modeling is yet to be explored in the DDI space. Further research into hybrid models, along with the application of traditional ML models into appropriate parts of the drug discovery pipeline could result in a more efficient drug discovery process.

FUNDING INFORMATION

No funding was received for this work.

CONFLICT OF INTEREST

V.P.R., M.M., A.M., F.M., and R.H.A. are full-time employees of and hold shares of AstraZeneca. B.W. was an employee of AstraZeneca at the time this research work carried out. All other authors declared no competing interests for this work.

ORCID

Anton Martinsson https://orcid.org/0000-0003-3963-6105
Filip Miljković https://orcid.org/0000-0001-5365-505X
Venkatesh Pilla Reddy https://orcid.org/0000-0002-7786-4371

REFERENCES

1. Midão L, Giardini A, Menditto E, Kardas P, Costa E. Polypharmacy prevalence among older adults based on the survey of health, ageing and retirement in Europe. *Arch Gerontol Geriatr*. 2018;78:213-220.
2. Yeo KR, Jamei M, Rostami-Hodjegan A. Predicting drug–drug interactions: application of physiologically based pharmacokinetic models under a systems biology approach. *Expert Rev Clin Pharmacol*. 2013;6(2):143-157.
3. Zhuang X, Lu C. PBPK modeling and simulation in drug research and development. *Acta Pharm Sin B*. 2016;6(5):430-440.
4. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;18(6):463-477.
5. Research CfDEa. *Clinical Drug Interaction Studies — Cytochrome P450 Enzyme- and Transporter-Mediated Drug Interactions Guidance for Industry*. Federal Information & News Dispatch LLC. 2020. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-drug-interaction-studies-cytochrome-p450-enzyme-and-transporter-mediated-drug-interactions
6. Bonate PL, Ahamadi M, Budha N, et al. Methods and strategies for assessing uncontrolled drug-drug interactions in population pharmacokinetic analyses: results from the International Society of Pharmacometrics (ISOP) working group. *J Pharmacokinet Pharmacodyn*. 2016;43(2):123-135.
7. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920-1930.
8. Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug–drug and drug–food interactions. *Proc Natl Acad Sci USA*. 2018;115(18):E4304-E4311.
9. Huang K, Xiao C, Hoang T, Glass L, Sun J. CASTER: predicting drug interactions with chemical substructure representation. *Proc AAAI Confer Artif Intell*. 2020;34(1):702-709.
10. Yao J, Sun W, Jian Z, Wu Q, Wang X. Effective knowledge graph embeddings based on multidirectional semantics relations for polypharmacy side effects prediction. *Bioinformatics*. 2022;38(8):2315-2322.
11. Chen Y, Ma T, Yang X, Wang J, Song B, Zeng X. MUFFIN: multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics*. 2021;37(17):2651-2658.
12. Jang HY, Song J, Kim JH, et al. Machine learning-based quantitative prediction of drug exposure in drug–drug interactions using drug label information. *Npj Digital Med*. 2022;5(1):88.
13. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan R. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol*. 2012;8(1):592.
14. Takeda T, Hao M, Cheng T, Bryant SH, Wang Y. Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge. *J Cheminform*. 2017;9(1):16.
15. Faisal Mahbub Chowdhury AL. A multi-phase kernel based approach for drug–drug interaction detection and classification that exploits linguistic information. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013); 2013:351–355.
16. Xia Sun JF, Ma L, Dong K, Du X. Deep convolution neural networks for drug–drug interaction extraction. In: Griol D, Zheng H, Schmidt H, eds. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE; 2018:1662-1668.
17. Preuer K, Lewis RPI, Hochreiter S, Bender A, Bulusu KC, Klambauer G. DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics*. 2018;34(9):1538-1546.
18. Yu L, Liu H. Feature selection for high-dimensional data: a fast correlation-based filter solution. In: Fawcett T, Mishra N, eds. *Proceedings of the twentieth international conference on international conference on machine learning*. AAAI Press; 2003:856-863.
19. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data*. 2019;6(1):27.
20. Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321-357.
21. Redkar S, Mondal S, Joseph A, Hareesha KS. A machine learning approach for drug-target interaction prediction using

wrapper feature selection and class balancing. *Mol Inform*. 2020;39(5):e1900062.
22. Singh RRR. Review on class imbalance learning: binary and multiclass. *Int J Comput Appl*. 2015;131(16):4-8.
23. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inform Process Syst*. 2017;30:4765-4774.
24. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Machine Intell*. 2020;2(10):573-584.
25. Mervin LH, Trapotsi M-A, Afzal AM, Barrett IP, Bender A, Engkvist O. Probabilistic random Forest improves bioactivity predictions close to the classification threshold by taking into account experimental uncertainty. *J Chem*. 2021;13(1):62.
26. Baker RE, Pena JM, Jayamohan J, Jerusalem A. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol Lett*. 2018;14(5):20170660.
27. Kasendra M, Luc R, Yin J, et al. Duodenum intestine-Chip for preclinical drug assessment in a human relevant model. *Elife*. 2020;9:e50135.
28. Kilford PJ, Chen KF, Crewe K, et al. Prediction of CYP-mediated DDIs involving inhibition: approaches to address the requirements for system qualification of the Simcyp Simulator. *CPT: Pharmacometrics Syst Pharmacol*. 2022;11:822-832.
29. Karlov DS, Sosnin S, Fedorov MV, Popov P. graphDelta: MPNN scoring function for the affinity prediction of protein–ligand complexes. *ACS Omega*. 2020;5(10):5150-5159.

How to cite this article: Gill J, Moullet M, Martinsson A, et al. Comparing the applications of machine learning, PBPK, and population pharmacokinetic models in pharmacokinetic drug–drug interaction prediction. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:1560-1568. doi:10.1002/psp4.12870