OXFORD
SOT | Society of Toxicology
academic.oup.com/toxsci

TOXICOLOGICAL SCIENCES, 189(1), 2022, 7–19

https://doi.org/10.1093/toxsci/kfac075
Advance Access Publication Date: 21 July 2022
Contemporary Review

CONTEMPORARY REVIEW

Machine Learning and Artificial Intelligence in Toxicological Sciences

Zhoumeng Lin [*,†,1] and Wei-Chun Chou [*,†]

[*]Department of Environmental and Global Health, College of Public Health and Health Professions, University of Florida, Gainesville, Florida 32610, USA; and [†]Center for Environmental and Human Toxicology, University of Florida, Gainesville, Florida 32608, USA

[1]To whom correspondence should be addressed at Department of Environmental and Global Health, College of Public Health and Health Professions, University of Florida, 1225 Center Drive, Gainesville, FL 32610, USA. E-mail: linzhoumeng@ufl.edu.

ABSTRACT

Machine learning and artificial intelligence approaches have revolutionized multiple disciplines, including toxicology. This review summarizes representative recent applications of machine learning and artificial intelligence approaches in different areas of toxicology, including physiologically based pharmacokinetic (PBPK) modeling, quantitative structure-activity relationship modeling for toxicity prediction, adverse outcome pathway analysis, high-throughput screening, toxicogenomics, big data, and toxicological databases. By leveraging machine learning and artificial intelligence approaches, now it is possible to develop PBPK models for hundreds of chemicals efficiently, to create *in silico* models to predict toxicity for a large number of chemicals with similar accuracies compared with *in vivo* animal experiments, and to analyze a large amount of different types of data (toxicogenomics, high-content image data, etc.) to generate new insights into toxicity mechanisms rapidly, which was impossible by manual approaches in the past. To continue advancing the field of toxicological sciences, several challenges should be considered: (1) not all machine learning models are equally useful for a particular type of toxicology data, and thus it is important to test different methods to determine the optimal approach; (2) current toxicity prediction is mainly on bioactivity classification (yes/no), so additional studies are needed to predict the intensity of effect or dose-response relationship; (3) as more data become available, it is crucial to perform rigorous data quality check and develop infrastructure to store, share, analyze, evaluate, and manage big data; and (4) it is important to convert machine learning models to user-friendly interfaces to facilitate their applications by both computational and bench scientists.

Key words: artificial intelligence; computational toxicology; machine learning; physiologically based pharmacokinetic (PBPK) modeling; quantitative structure-activity relationship (QSAR).

Toxicology is a disciplinary of science that studies the adverse effects and the underlying mechanisms of toxicity caused by chemicals, substances, or situations on humans, animals, and the environments, and the prevention and amelioration of such harmful effects, as well as the application of toxicology knowledge to safety evaluation and risk assessment of xenobiotics (Klaassen, 2018; NIEHS, 2022). Toxicology includes a variety of subject areas based on different classifications, including chemical toxicology (toxicity of different chemical classes, such as pesticides, metals, etc.), organ systems toxicology (toxicity on different target organs), nonorgan-directed toxicity (carcinogenesis, genetic toxicology, and developmental toxicology), toxicokinetics (eg, physiologically based pharmacokinetic [PBPK] modeling), environmental toxicology, as well as toxicology applications in regulatory risk assessment, ecotoxicology, food toxicology, clinical toxicology, and occupational toxicology.

Artificial intelligence is a rapidly developing subdiscipline of computer science with the goal of designing and creating machines or computational models that can perform a variety of cognitive tasks at a level comparable or even exceed human intelligence (Davidovic *et al.*, 2021). The term artificial intelligence can have different meanings in different fields. In the present contemporary review, it refers to the applications of various machine learning methods in the prediction and evaluation of chemical toxicokinetic (ie, absorption, distribution, metabolism, and excretion [ADME]) and toxicity properties. Machine learning is a subarea of artificial intelligence, and it refers to mathematical or computer algorithms designed to teach or train a computational model to solve a problem or perform complex tasks based on some input parameters (Russell and Norvig, 2020). Machine learning is generally categorized into 3 types: supervised learning, unsupervised learning, and reinforcement. Commonly used machine learning methods in the field of toxicology and a brief description of each method are listed in Table 1 (Baskin, 2018; Lin *et al.*, 2022).

In recent years, machine learning and artificial intelligence approaches are increasingly applied in different subject areas of toxicology, including neurotoxicity (Aschner *et al.*, 2022), cardiovascular toxicology (Glass *et al.*, 2022), nanotoxicology (Ji *et al.*, 2022; Singh *et al.*, 2020), toxicokinetics (Bhhatarai *et al.*, 2019; Chou and Lin, forthcoming), dermal toxicity (Hu *et al.*, 2022), carcinogenesis (Li *et al.*, 2021), etc. This contemporary review aims to analyze this emerging area of applying machine learning and artificial intelligence approaches to study toxicology research questions and provide an overview of the current state of the science in this area. The progress and challenges on how to integrate machine learning and artificial intelligence approaches with traditional toxicology approaches, such as PBPK modeling, quantitative structure-activity relationship (QSAR) modeling, adverse outcome pathway (AOP) analysis, toxicogenomics, and high-content image-based screening data, will be summarized, followed by our future perspectives. The timeline of the applications of machine learning, artificial intelligence, PBPK, and QSAR modeling approaches in the fields of pharmacology and toxicology is presented in Figure 1.

PHYSIOLOGICALLY BASED PHARMACOKINETIC MODELS

PBPK modeling is a computational simulation process that describes the ADME of a xenobiotic and its metabolite(s) in the body based on interrelationships among key anatomical, physiological, biochemical, and physicochemical determinants using mathematical equations (Fisher *et al.*, 2020). PBPK models are an important tool in human health risk assessment, especially in dose-response analysis, exposure assessment, *in vitro* to *in vivo* extrapolation (IVIVE), and interspecies extrapolation of toxicity and dosimetry data. In the field of toxicology, a number of PBPK models for different chemicals have been developed, and many of them have been used to support chemical risk assessment (Fisher *et al.*, 2020; Reddy *et al.*, 2005; Tan *et al.*, 2018). To build a PBPK model, a traditional approach is to experimentally measure relevant parameters, such as tissue:plasma partition coefficients and metabolic rates, and then estimate values of parameters that do not have experimental values by fitting to *in vivo* pharmacokinetic dataset(s). This process is labor-intensive, time-consuming, expensive, and unethical from animal welfare perspective as the *in vivo* pharmacokinetic datasets have to be collected from animals *in vivo*. Also, this traditional approach cannot keep up with increasing demand of PBPK models for thousands of chemicals whose risks remain to be evaluated.

Machine learning approaches have been applied to predict PBPK parameters based on compounds' physicochemical properties to generate PBPK models for a large number of compounds efficiently. A list of presentative recent PBPK studies using machine learning approaches is provided in an accompanying manuscript (Chou and Lin, forthcoming). For example, Kamiya *et al.* (2021) developed an *in silico* model based on a gradient boosting framework (LightGBM) machine learning approach to predict 3 key PBPK parameters, including absorption rate constant, volume of distribution, and hepatic intrinsic clearance based on around 14–26 physicochemical properties of 246 compounds. The results showed that PBPK-predicted concentration values of the 246 compounds in plasma, liver, and kidney of rats using the *in silico* estimated parameter values were well correlated with those based on traditionally determined parameter values with a correlation coefficient of $r \geq 0.83$. Another research group tested multiple machine learning algorithms (eg, lasso regression, support vector machine, random forest, and neural network multiple layer perceptron) to determine the optimal model for the prediction of 2 essential toxicokinetic parameters: fraction of the chemical unbound in plasma and intrinsic clearance, based on structural properties from a dataset of 1487 environmental chemicals; the final models (based on support vector machine and random forest) can be used to predict these toxicokinetic parameters for other chemicals of which experimental data are not available (Pradeep *et al.*, 2020). These studies demonstrate that it is feasible to use machine learning approaches to estimate PBPK parameters based on compounds' physicochemical properties and then to develop a generic PBPK model for a large number of compounds to facilitate dosimetry estimation for risk assessment and ranking.

Machine learning approaches can support the development of PBPK models. In turn, a PBPK model can be used to generate a large amount of simulated data to be analyzed with machine learning approaches to obtain new insight. A recent study reported a generic PBPK model for nanoparticles in tumor-bearing mice (Cheng *et al.*, 2020). This model was trained with 376 datasets for different types of nanoparticles. The final model was used to predict the delivery efficiency of different nanoparticles to tumors based on 4 dose metrics, including tumor delivery efficiency estimated at 24 h, 168 h, and the last sampling time point, as well as the maximum delivery efficiency. Various machine learning and deep learning algorithms (briefly described in Table 1), such as linear regression, *k*-nearest neighbors, random forest, bagged model, stochastic gradient boosting, support vector machine, and deep neural network were used to analyze the data to determine the best model that can predict tumor delivery efficiency of a nanoparticle based on its physicochemical properties, including Zeta potential, hydrodynamic diameter, shape, targeting strategy, core material, and type of nanoparticles (Lin *et al.*, 2022). The results showed that the deep neural network model adequately predicted the delivery efficiency of different nanoparticles to different tumors and it outperformed all other machine learning methods. This strategy of using machine learning methods to analyze a large amount of PBPK-simulated data can well be applied to small molecular environmental chemicals. It is anticipated that this approach will greatly expedite the application of PBPK in

Table 1. A List of Machine Learning Methods Commonly Used in Toxicological Research

Method	Brief Description
Supervised linear methods	
Multiple linear regression	Use multiple explanatory variables to predict the outcome of a response variable with a multivariate linear equation
Naïve Bayes classifier	Based on Bayes' theorem with strong assumptions of conditional independence among molecular descriptors (ie, explanatory variables)
Supervised nonlinear methods	
k-nearest neighbors	Classify a test chemical by looking for the training chemicals with the nearest distance to it
Support vector machine	Map molecular descriptor vectors into a higher dimensional feature space to build a maximal margin hyperplane to distinguish active (toxic) from inactive (nontoxic) chemicals
Decision trees	Each model is a series of rules organized in the format of a tree containing a single root node and any number of internal nodes and several leaf nodes. The path from the root to a leaf stands for a sequence of classification rules predicting a toxicity endpoint for a given chemical
Ensemble learning	Combine several base models into a more predictive one. Popular types of ensemble modeling include bagging, random spaces, boosting, and stacking.
Random forest	Combine the bagging with the random spaces approaches in application to decision trees base models
Artificial neural networks	
Backpropagation neural networks	All neurons are divided into 3 layers, with information flowing from the first layer of input neurons to the second layer of hidden neurons, and then to the third layer of output neurons
Bayesian-regularized neural networks	Apply Bayesian methods to perform regularization so that the model complexity is balanced against the accuracy of reproducing training data
Associative neural networks	Apply ensemble learning to backpropagation neural networks
Deep neural networks	Artificial neural networks with multiple hidden layers (also called deep learning)
Unsupervised methods	
Principle component analysis	Reduce the dimensionality of the data to only the first few principal components while preserving as much of the data's variation as possible
Kohonen's self-organizing maps	Map molecules from the original descriptor space onto a 2D grid of neurons. Similar molecules will be mapped to the same closely located neurons in the grid

This table is based on the book chapter by Baskin (2018). Please refer to Baskin (2018) for detailed description about each of the listed machine learning algorithms.

combination with machine learning for a large-scale chemical screening, risk ranking and prioritization.

Traditionally, PBPK models are described with ordinary differential equations (ODEs) and solved with ODE solvers; and population pharmacokinetic (PopPK) models are developed using a nonlinear mixed effects (NLME) approach. Recent advancements in machine learning and artificial intelligence have led to significant progress in applying these approaches to pharmacometrics or toxicometrics. A deep learning approach based on neural ordinary differential equations (neural-ODE) (Chen *et al.*, 2018) was created for automated construction of pharmacokinetic models directly based on clinical data (Lu *et al.*, 2021a,b). The performance of the neural-ODE model was compared with other machine learning approaches (ie, the lightGBM and long short-term memory [LSTM] neural network) and the traditional NLME modeling. The results showed that the neural-ODE, lightBGM, and NLME models had similar prediction performance when the training data and testing data were from the same treatment regimens, but the neural-ODE outperformed other algorithms when applying to new dosing regimens. It is anticipated that this deep learning-based neural-ODE approach may also be applied to PBPK models to facilitate its applications in both pharmacology and toxicology.

QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP MODELS

QSAR is a computational modeling and simulation method for studying relationships between structural properties of chemicals and biological activities. The biological activities include ADME properties, as well as toxicity of chemical substances. QSAR approaches have been extensively applied in the areas of drug discovery and development, as well as toxicology. Emerging machine learning and artificial intelligence approaches are now commonly employed to build robust QSAR models to predict bioactivities of a large number of chemicals. Table 2 lists representative recent studies that used machine learning and artificial intelligence approaches to train QSAR models. These models are an ideal too to perform read-across in toxicology (ie, to predict the bioactivities of new chemicals based on structurally related or similar analogues without doing additional *in vitro* or *in vivo* experimentation).

Development of a QSAR model typically involves 4 main steps: (1) collecting a training dataset (ie, chemicals with experimentally-derived physical and/or biological properties), (2) encoding chemicals with molecular descriptors (ie, the features of each molecule), (3) training the model to predict chemical properties based on their molecular descriptors using mathematical algorithms (from simple multiple linear regression to state-of-the-art machine learning algorithms), and (4) evaluating the model performance using a validation dataset (Cheng and Ng, 2019; OECD, 2014).

Per- and polyfluorinated alkyl substances (PFAS) are a large chemical family with >5000 members that are widely used in industrial and consumer products. PFAS are ubiquitous environmental contaminants and represent a major global public health issue. Due to a large number of PFAS, it is difficult and impractical to evaluate the toxicity of each of them individually using *in vitro* and/or *in vivo* assays. To address this challenge, a machine learning-based QSAR model was built and successfully

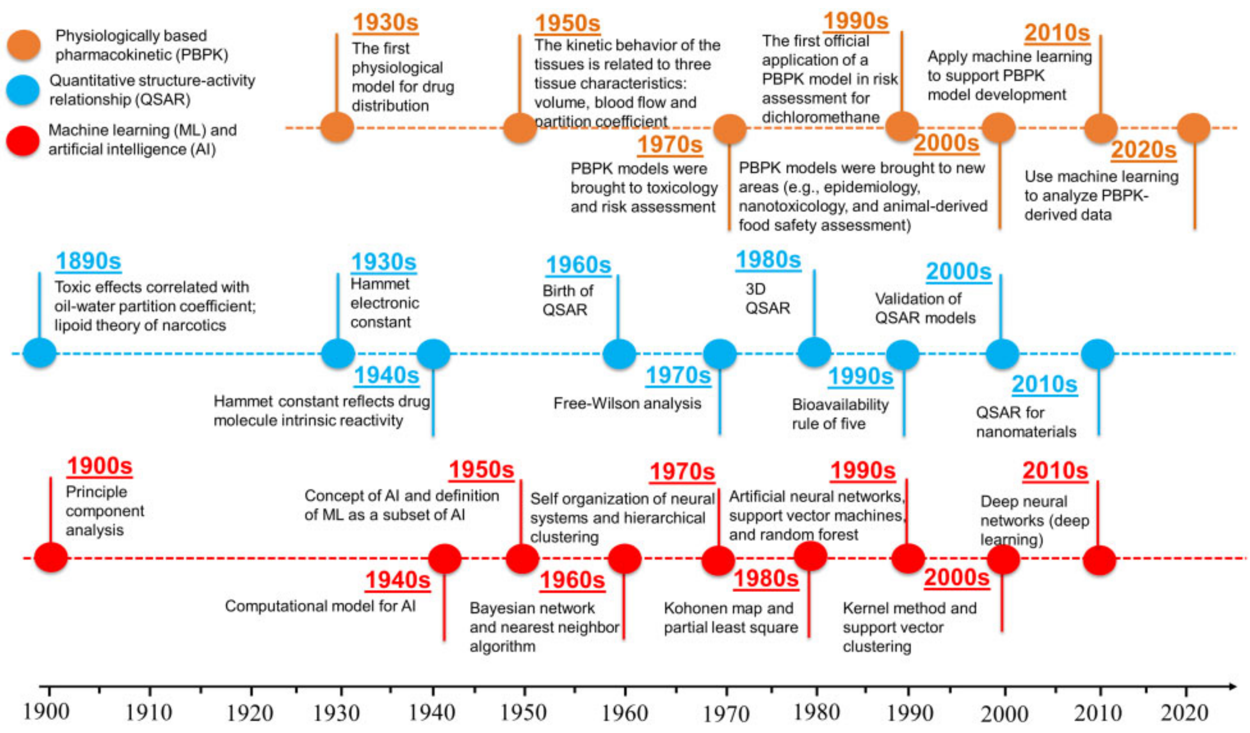

Figure 1. A timeline of the applications of machine learning (ML), artificial intelligence (AL), physiologically based pharmacokinetic (PBPK), and quantitative structure-activity relationship (QSAR) modeling approaches in the fields of pharmacology and toxicology. This figure was created based on Figure 3 in Zhu (2020), Figure 1 in Lin and Fisher (2020), and Figure 1 in Singh *et al.* (2020). Please refer to these references for the original references for the milestones listed in this figure.

classified bioactivity for 3486 PFAS (Cheng and Ng, 2019). The authors constructed a PFAS-specific database that contains bioactivity information on 1010 PFAS for 26 assays to serve as the training dataset. These bioassays were all binary classification assays (active or inactive), and involved different target receptors or enzymes, such neuropeptide S receptor, CYP2C9, aldehyde dehydrogenase 1. Five different machine learning models (ie, logistic regression, random forest, multitask neural network, graph convolutional model, and weave model) were evaluated on different assays, and the best model was selected for each assay. The models were evaluated with a dataset from Organisation for Economic Co-operation and Development (OECD) containing bioactivity information on 3486 PFAS. The results showed that the average of the area under the curve (AUC) of the receiver operating characteristic (ROC) curve for each bioassay was >0.9. This study provides a valuable model to classify bioactivity of a large number of PFAS based on chemical structures.

In 2014, the National Center for Advancing Translational Sciences (NCATS) launched Tox21 Data Challenge to develop and compare different computational models for toxicity prediction based on chemical structure data. The goal was that some of the robust machine learning-based models could be used as decision-making tools for governmental agencies in determining which chemicals may be of great potential concern to human health. The challenge organizers provided a training set consisting of 11 764 chemicals, a leaderboard set consisting of 296 chemicals, both with structural and bioassay data, as well as a test set consisting of 647 chemicals with only structural data. The bioassay data included 12 toxic endpoints, mainly related to nuclear receptor effects, such as activation of the estrogen receptor, and stress response effects, such as the heat shock response effect. More than 10 research teams from all over the world participated in the challenge. Mayr *et al.* (2016) developed a DeepTox pipeline that embedded with different machine learning models (Table 1), mainly deep learning model, and also with complementary models, such as support vector machines, random forests, and elastic nets. The results showed that the DeepTox pipeline had a consistent high performance compared with all competing methods from other research teams, and won a total of 9 of the 15 challenges and did not rank lower than fifth place in any of the subchallenges. It is also worth to note that within DeepTox pipeline, the deep learning model had a superior performance than other complementary methods for toxicity prediction in 10 out of 15 evaluation test sets.

Similar to NCATS' Tox21 Data Challenge, recently the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Acute Toxicity Workgroup organized an international collaboration to develop machine learning-based *in silico* models to predict acute oral toxicity (eg, LD_{50}) based on a database for 11 992 chemicals (Mansouri *et al.*, 2021). Thirty-five research groups submitted 139 predictive models. The final consensus models were submitted to regulatory agencies for evaluation of its utility and applicability to potentially replace *in vivo* rat acute oral toxicity studies. The final consensus models and the prediction results are publicly available (Mansouri *et al.*, 2021).

Carcinogenicity testing plays an important role in identifying carcinogens in drug development and environmental chemical risk assessment. Traditionally, the carcinogenic potency is evaluated with a 2-year carcinogenicity study in rodents, but this process is very time-consuming and resource-intensive. There has been a great need to develop alternative approaches for reliable and efficient assessment of carcinogenicity. Multiple QSAR models have been developed to assess carcinogenicity for

Table 2. Representative Studies Integrating Machine Learning Approaches With Quantitative Structure-Activity Relationship Modeling

Best Machine learning Method	Training Dataset	Endpoint	Reference
Deep learning (ie, DeepTox)	11 764 chemicals from Tox21	12 bioassays	Mayr *et al.* (2016)
Ensemble extreme gradient boosting	1003 chemicals	Carcinogenicity	Zhang *et al.* (2017)
Random forest	Over 866 000 chemical properties/hazards	Acute oral and dermal toxicity, eye and skin irritation, mutagenicity, and skin sensitization	Luechtefeld *et al.* (2018)
Ensemble support vector machine	400 chemicals	Aquatic acute toxicity	Ai *et al.* (2019)
Multitask neural networks and graph convolutional networks	1012 PFAS	Bioactivity on 26 bioassays	Cheng and Ng (2019)
Extra trees	Over 1000 chemicals from different databases	Various toxicities	Pu *et al.* (2019)
Ensemble model	7385 chemicals	Acute toxicity in rats	Russo *et al.* (2019)
Support vector machine	482 chemicals	Acute toxicity in fathead minnow	Chen *et al.* (2020)
Deep learning (ie, CapsCarcino)	1003 chemicals from CPDB	Carcinogenicity	Wang *et al.* (2020)
Kernel-weighted local polynomial approach	Hundreds of chemicals depending on the species	Acute aquatic toxicity	Gajewicz-Skretna *et al.* (2021)
Meta ensembling of multitask deep learning models (ie, QuantitativeTox)	Hundreds to thousands of compounds depending on the endpoint	LD_{50} and LC_{50}	Karim *et al.* (2021)
Deep learning-based model-level representations (ie, DeepCarc)	692 chemicals	Carcinogenicity	Li *et al.* (2021)
Extra trees	Over 18 600 drug-bacteria interactions	Gut bacterial growth	McCoubrey *et al.* (2021)
Support vector machine	676 pesticides	Acute contact toxicity on honey bees	Xu *et al.* (2021)
A consensus model based on 4 algorithms	1244 chemicals	Prenatal developmental toxicity	Ciallella *et al.* (2022)
Deep learning	31 chemicals with known or suspected clinical skin toxicity	Skin toxicity	Hu *et al.* (2022)
Random forest	1476 food contact chemicals	Carcinogenicity	Wang *et al.* (2022)

CPDB, Carcinogenic Potency Database. LC_{50} and LD_{50} refer to the compound concentrations that kill half the members of the tested animal population, respectively.

particular chemical classes (eg, aromatic amines, polycyclic aromatic hydrocarbon) in rats (Wang *et al.*, 2020; Zhang *et al.*, 2017). However, chemical carcinogenicity assessment is required to be conducted in at least 2 rodent species. Recently, Li *et al.* (2021) developed a DeepCarc model to predict carcinogenicity for small molecules using deep learning-based model-level representations. The DeepCarc model was developed with a dataset of 692 chemicals and evaluated with a test set consisting of 171 chemicals. The data were obtained from the National Center for Toxicological Research liver cancer database and involved both rats and mice. The authors also compared performance of the DeepCarc model with other deep learning models that were based on molecule-level representations, including Text Convolutional neural network from DeepChem, Convolutional Neural Network Fingerprint, Edge Attention-based Multi-relational Graph Convolutional Networks, and Chemistry Chainer-Neural Fingerprint. The results showed that model predictions from DeepCarc had an accuracy of 0.754, a sensitivity of 0.910, and a specificity of 0.467 in the test dataset. Also, DeepCarc had a superior performance in accuracy and sensitivity than the molecular-based deep learning models. This DeepCarc model provides an early screening nonanimal-based tool to assess potential carcinogenicity of new chemicals and is useful for prioritizing chemicals on their potential carcinogenic risk.

In carcinogenicity assessment using computational models, one common issue is insufficient coverage of mechanisms and chemicals in the applicability domain of individual models. To address this challenge, a machine learning-based weight-of-evidence model was developed to prioritize chemical carcinogenesis by integrating results from multiple computational methods with complementary mechanisms, including structural alert models, QSAR models, *in silico* toxicogenomics models into a weight-of-evidence score (Wang *et al.*, 2022). The model was developed based on a training dataset with 597 chemicals and a test dataset with 198 chemicals from the International Agency for Research on Center (IARC) chemical list. A random forest algorithm was used to develop the weight-of-evidence classifiers. The results showed that the machine learning-based weight-of-evidence model had 8% and 19.7% improvement compared with the best single method in the area under the receiver operating characteristic curve (AUROC) value and chemical coverage, respectively. The weight-of-evidence model was then applied to assess the weight-of-evidence scores of 1623 food contact chemicals and prioritize these chemicals based on the weight-of-evidence scores. The model identified 44 chemicals to be of high carcinogenicity concern based on a predefined weight-of-evidence threshold of 0.7. Among these 44 chemicals, 34 had carcinogenic data in public databases (either IARC or ECHA). A high ratio of chemicals with consistent results between model predictions and known carcinogenic potential from public databases suggest the effectiveness of the developed machine learning-based weight-of-evidence model for prioritizing chemicals of high carcinogenicity concern.

ADVERSE OUTCOME PATHWAY ANALYSIS

An AOP is a conceptual construct that describes existing knowledge on the connection between a direct molecular initiating event and an adverse outcome at a biological level of organization that is relevant to human health risk assessment (Ankley *et al.*, 2010). A typical AOP includes a molecular initiating event (eg, interaction between a chemical and a specific biomolecule at the molecular level), key events that characterize the progression of toxicity after the molecular initiating event, and adverse outcomes that may occur at individual or population levels. In the past 10 years, a number of AOPs have been characterized and summarized in the AOP Knowledge Base and/or AOP Wiki (OECD, 2022).

To determine whether xenobiotics are involved in an AOP, an efficient approach is to perform high-throughput screening (HTS) assays that are designed to measure key events of AOPs. One of the most prominent HTS initiatives in toxicology is U.S. Environmental Protection Agency (EPA)'s Toxicity Forecaster (ToxCast) program, which later progressed to become Tox21 program among multiple agencies, including U.S. EPA, Food and Drug Administration (FDA), NCATS, and National Toxicology Program (NTP). Tox21 program has screened thousands of chemicals in over 70 high-throughput assays covering more than 125 important biological processes in the body and generating >120 million data points (Tox21, 2020). Among all studied AOPs, one of the most commonly studied AOPs is related to nuclear estrogen receptor α and β (Ciallella *et al.*, 2021; Huang *et al.*, 2014; Lin and Lin, 2020). Estrogen receptors play important roles in many biological functions, such as cell differentiation, fertility, and morphogenesis. Multiple xenobiotics have been shown to bind to and active estrogen receptors, with the potential to result in endocrine disruption and adverse effects on reproductive organs.

Note that traditional approaches to evaluate endocrine disruptors that activate estrogenic signaling requires labor- and resource-intensive *in vitro* or *in vivo* experiments. In a recent study, Ciallella *et al.* (2021) developed a knowledge-based deep neural network model to analyze publicly available HTS data to identify compounds with nuclear estrogen receptor α and β binding potentials. In this model, the input layer of the network contained information on 1024 functional connectivity chemical fingerprints plus 3 known ERα/ERβ toxicophores (ie, steroid and diethylstilbestrol scaffolds and the phenol group). The output layer of the network was the target activity, which was *in vivo* rodent uterotrophic bioactivity. There were 5 hidden layers that connected the input and the output layers. The 5 hidden layers were organized and ordered using an AOP framework, with each layer represented a higher level of biological organization than the last. The 5 layers included 57 neurons in total, each of which represented one *in vitro* high-throughput assay included in the training dataset. After training, the resulting model successfully predicted critical relationships among ERα/ERβ target bioassays based on chemical fingerprints. The model used an AOP framework to mimic the signaling pathway initiated by ERα and was able to identify compounds that mimic endogenous estrogens. This virtual pathway model, starting from a compound's chemistry initiating ERα activation and ending with rodent uterotrophic bioactivity, can efficiently prioritize new estrogen mimetics. This artificial intelligence-based model provides a promising strategy to integrate AOP and high-throughput data to characterize hazards and prioritize potential toxic compounds for further risk assessment.

Although traditional descriptive or qualitative AOP is useful in chemical risk assessment, it does not provide quantitative relationships from chemical exposure to effect timing and magnitude. When there are sufficient data on quantitative relationships between chemical exposure and key events (or molecular initiating events or adverse outcomes), a mathematical model may be developed to connect chemical exposure to key events in a quantitative AOP (qAOP). For example, Zgheib *et al.* (2019) used 3 approaches to build qAOP models to quantitatively describe a simplified oxidative stress induced chronic kidney disease AOP based on *in vitro* data from human proximal tubule (RPTEC/TERT1) cells treated with potassium bromate. These 3 approaches included: empirical dose-response modeling, Bayesian network calibration, and systems biology modeling. The authors concluded that the Bayesian network approach was more precise than the dose-response models and simpler than the systems biology models. In light of the potential regulatory applications of qAOP in chemical risk assessment, an increasing number of qAOP models have been proposed as computational toxicity predictive tools. Readers are referred to these recent review articles (Perkins *et al.*, 2019; Sinitsyn *et al.*, 2022; Spinu *et al.*, 2020) on the detailed methodology of qAOP development and applications.

Among all key events of an AOP, the molecular initiating event links a chemical's structural properties to an interaction at a biological target, thus providing an opportunity to build QSAR models to predict a chemical's molecular initiating event based on its structural properties (Allen *et al.*, 2014). In a series of studies, Allen *et al.* developed a tool to predict a chemical's molecular initiating event based on 2D structural alerts of the chemical (Allen *et al.*, 2016, 2018). This tool was developed based on data from ChEMBL that contains more than a million annotated compounds with over 12 million bioactivities covering more than 10 000 biological targets. The final tool contained 4810 different structural alerts for 39 pharmacologically important targets. The performance of the final model's predictions of molecular initiating events was strong, with 82% sensitivity, 93% specificity, and 93% overall quality (Allen *et al.*, 2018).

TOXICOGENOMICS

Toxicogenomics is a subdiscipline of toxicology that applies genomic technologies (eg, gene expression profiling, proteomics, metabolomics, and related methods) to study adverse effects of chemicals or xenobiotic substances at the gene and/or protein levels within particular cells or tissue(s) of an organism. Toxicogenomics has emerged to be an important tool in the identification of potential molecular mechanisms of toxicity at the gene, protein, or metabolite level in cells or tissues of organisms in response to exposure to environmental chemicals, as well as serving as biomarkers for predictive toxicology. Recent advance in computational technologies has enable integration of toxicogenomics with computational models (eg, machine learning and PBPK models) to correlate molecular endpoints derived from toxicogenomics data with *in vivo* regulatory-relevant phenotypic toxicity or toxicokinetic endpoints (Chen *et al.*, 2022a; Liu *et al.*, 2019).

In a recent study, researchers collected *in vitro* assay-derived time-series toxicogenomic data on the expression of a library of 38 key proteins (covering all known recognized DNA damage repair pathways) after exposure to a wide concentration range of 20 selected genotoxicity-positive and genotoxicity-negative chemicals (Rahman *et al.*, 2022). Machine learning-based feature selection method (ie, maximum relevance and minimum redundancy [MRMR]) and classification method (support vector machine [SVM]) was employed to identify an optimal number of biomarkers with minimum redundancy for prediction of phenotypic toxicity endpoints (*in vivo* carcinogenicity and Ames genotoxicity) in rats. The authors found that a small number of properly selected molecular biomarker-ensemble involved in conserved DNA damage and repair pathways among eukaryotes were able to predict both *in vivo* carcinogenicity and Ames genotoxicity endpoints with good accuracies ($\geq$70% for both endpoints with the top 5 biomarkers). The identified top 5 biomarkers are associated with known DNA damage and repair pathways. For example, the identified top 5 biomarkers for the *in vivo* carcinogenicity prediction were mainly related to double strand break repair and DNA recombination. This study provides a proof-of-concept that machine learning methods can be applied to analyze toxicogenomic data to bridge molecular level biomarker data to regulatory-relevant *in vivo* phenotypic and toxicity endpoints.

Toxicogenomics data could be derived from *in vitro* or *in vivo* assays. Although *in vivo* toxicogenomics data are desirable, it is impractical and unethical to collect toxicogenomics data for thousands of chemicals from animal assays on different dose groups and treatment durations. A recent study applied a deep generative adversarial network (GAN) approach to develop an artificial intelligence-based Tox-GAN framework that is capable of generating *in vivo* gene activities and expression profiles in rats for multiple doses and treatment durations based on chemical structures (Chen *et al.*, 2022b). This model was trained with data from a large-scale publicly available toxicogenomics database that contains transcriptomic data derived from *in vivo* and *in vivo* exposure to 170 compounds at multiple dose levels and time points (Igarashi *et al.*, 2015). The Tox-GAN-derived toxicogenomics data had >87% agreement in Gene Ontology compared with the experimentally derived gene expression data. This framework serves as a promising alternative tool to generate high-quality *in vivo* toxicogenomic data without animal experimentation.

HIGH-CONTENT IMAGE-BASED SCREENING DATA

Artificial intelligence-based methods have been applied to study mechanisms of toxicity, such as oxidative stress and DNA damage. Oxidative stress is a common mechanism of different toxic effects induced by various environmental stressors (eg, heavy metals, ionizing radiations, antiblastic drugs) (Pizzino *et al.*, 2017). Generation of reactive oxygen species, such as hydrogen peroxide, hydroxyl radical, superoxide anion and singlet oxygen is one of the common causes of DNA damage. There are different types of DNA modifications (eg, single-strand breaks, double-strand breaks, bulky adduct formation), and different assays to evaluate the severity of DNA damage. One of the commonly used assays is the comet assay that can evaluate the level of DNA fragmentation, which corresponds to the amount of damaged DNA. Although comet assay has been extensively applied for several decades, one shortcoming of this assay is lack of automation. In this regard, recent studies have applied artificial intelligence methods to evaluate DNA damage based on segmented comet assay images. For example, in a study by Atila *et al.* (2020), the researchers developed a convolutional neural network (CNN) model based on comet assay image data (the original data contained 796 images and the augmented data consisted of 9995 images). The CNN model architecture included an input layer, an output layer, and 9 hidden layers in between. The model was able to classify comet images into 4 classes (healthy, poorly defective, defective, and very defective) with an overall accuracy rate of 96.1%.

BIG DATA IN TOXICOLOGY AND TOXICOLOGICAL DATABASES

The term "big data" can be defined as datasets, structured or unstructured, that include a large variety of types of data and are generated in a high speed with a volume that is so large that they usually require high-performance computers and advanced computational approaches to analyze (Ciallella and Zhu, 2019). In the field of toxicology, examples of big data include high throughput/high content screening data (eg, ToxCast/Tox21 data), data generated with omics technologies and gene arrays (eg, transcriptomics, metabolomics, proteomics, and microbiome), toxicity data in large public databases (eg, Table 3), and epidemiological data, as well as environmental monitoring or human biomonitoring data (eg, the National Health and Nutrition Examination Survey [NHANES]). One of the prerequisites in the application of artificial intelligence approaches to study biomedical problems is the requirement of big data (ie, the dataset should be large enough to enable to develop a reliable model without overfitting). The availability of various types of big data sources in toxicology makes it possible to apply artificial intelligence approaches to predictive toxicology.

Combining machine learning approaches and toxicological big data enables development of read-across structure activity relationship (RASAR) that may outperform animal test reproducibility (Luechtefeld *et al.*, 2018). Based on a big database containing more than 866,000 chemical properties/hazards, 2 RASAR models (ie, simple RASAR and data fusion RASAR) were trained with an unsupervised learning step and a supervised learning step. The simple RASAR model combined an unsupervised aggregation function based on *k*-nearest neighbor algorithm to generate a 2D vector for each chemical, and then a supervised learning step based on logistic regression was applied to the vectors generated by the unsupervised learning step. The data fusion RASAR extended the simple RASAR by building similarity-based features for every chemical and properties and created large feature vectors, which were then applied to train a random forest as the supervised learning model. The results showed that the simple RASAR model achieved a sensitivity of >80% with specificities of 51%–69% on the animal reproducibility test results, and the data fusion RASAR further improved the sensitivity to the 80%–95% range. Note that in general the probability that an animal test based on OECD guidelines that would achieve the same result in a repeat test is around 78%–96% (Luechtefeld *et al.*, 2018). These results suggest that big data and machine learning-based advanced QSAR or RASAR models may achieve similar or even outperform animal test reproducibility (Luechtefeld *et al.*, 2018).

Machine learning methods have also been used to study adverse effects of chemicals on gut microbiome. Based on a dataset consisting of the effects of 1197 drugs on the *in vitro* growth of 40 representative strains of gut bacteria, McCoubrey *et al.* (2021) develop a machine learning model to predict whether the drugs impair the growth of the 40 gut bacterial strains based on chemical structural features. A total of 13 different machine learning models were tried, including extra trees, random forest, *k*-nearest neighbors, multilayer perceptron, decision trees,

Table 3. A List of Databases Relevant to Computational Toxicology

Database	Data Size[a]	Data Type	Reference
ACToR	Over 800 000 compounds and 500 000 assays	*In vitro* and *in vivo* toxicity	Judson *et al.* (2008)
Biosolids list	726 chemical pollutants	Concentration data in biosolids	Richman *et al.* (2022)
CEBS	Over 11 000 compounds and 8000 studies	Gene expression data	Lea *et al.* (2017)
ChEMBL	1.1 million bioassays, 1.8 million compounds, over 15 million activities	Literature data on binding, function, and toxicity of drugs and drug-like chemicals	Gaulton *et al.* (2012)
Connectivity map	Around 1300 compounds and 7000 genes	Gene expression data	Subramanian *et al.* (2017)
CTD	Over 14 000 compounds, 42 000 genes, 6000 diseases	Relationships among compounds, genes, and diseases	Davis *et al.* (2021)
DrugMatrix	Around 600 drug molecules and 10 000 genes	Gene expression data	Ganter *et al.* (2005)
GEO	Over 4300 subdata sets	Microarray, next-generation sequencing, and other forms of high-throughput functional genomics data	Barrett *et al.* (2013)
eNanoMapper	Over 700 types of nanomaterials	Diverse data types on nanomaterial physicochemical properties and safety	Jeliazkova *et al.* (2015)
MoleculeNet	Over 700 000 compounds	Quantum mechanics, physical chemistry, biophysics, and physiology	Wu *et al.* (2018)
Open TG-GATEs	170 compounds	Gene expression data and metadata	Igarashi *et al.* (2015)
PubChem	Over 111 million compounds, 1.39 million bioassays, and 293 million bioactivity data points	Toxicology, genomics, pharmacology, and literature data	Kim *et al.* (2021)
Pubvinas	11 types of nanomaterials with 705 unique nanomaterials	Up to 6 physicochemical properties and/or bioactivities	Yan *et al.* (2020)
REACH	21,405 unique substances with information from 89,905 dossiers	Data submitted in European Union chemical legislation	Luechtefeld *et al.* (2016)
RepDose	364 compounds investigated in 1017 studies, resulting in 6,002 specific effects	Repeat-dose study data in dogs, mice, and rats	Bitsch *et al.* (2006)
SEURAT	Over 5500 cosmetic-type compounds in the current COSMOS database web portal	Animal toxicity data	Vinken *et al.* (2012)
ToxicoDB	231 chemicals	Toxicogenomic data	Nair *et al.* (2020)
ToxNET	Over 50 000 environmental chemicals from 16 resources	*In vitro* and *in vivo* toxicity data	Fonger *et al.* (2000)

[a]On the basis of live web counts or most recent literature publications as of March 2022. ACToR, Aggregated Computational Toxicology Resource; CTD, Comparative Toxicogenomics Database; CEBS, Chemical Effects in Biological Systems; GEO, Gene Expression Omnibus; Open TG-GATEs, a large-scale toxicogenomic database; REACH, Registration, Evaluation, Authorization, and Restriction of Chemicals; SEURAT, Safety Evaluation Ultimately Replacing Animal Testing; ToxNET, Toxicology Data Network. This table was adapted from Ciallella and Zhu (2019) with permission from the publisher.

support vector machines, stochastic gradient descent, perceptron, passive aggressive classification, gradient boosting, etc. The results showed that the extra trees model had the best performance based on all evaluation metrics (AUROC: 0.850, recall: 0.595; precision: 0.785; f1: 0.666), followed by the random forest model. This model can be used by pharmaceutical companies or regulatory agencies to predict whether a new drug may impact gut microbiome of patients.

As toxicology enters the big data era, more and more toxicology-related databases are available to perform a large-scale computational data to obtain new toxicology findings using machine learning and artificial intelligence approaches. Some representative databases are listed in Table 3. Most of these databases store physicochemical and toxicological data on small molecular environmental chemicals, and studies on how to analyze data from these databases with machine learning and artificial approaches are published. It is worth to highlight that nanomaterials are emerging environmental toxicants, and big databases on nanomaterial toxicological properties have begun to be developed. Yan *et al.* recently constructed a web-based nanomaterial database through big data curation and modeling friendly nanostructure annotations (Yan *et al.*, 2020). This database contains 705 unique nanomaterials covering 11 material types with 6 physicochemical properties and/or bioactivities for each nanomaterial, resulting in >10 endpoints in the database. Note that the nanostructure annotation contains 2142 nanodescriptors for all nanomaterials that are available for download from the web portal for subsequent machine learning research purposes. In the Europe, eNanoMapper has been created as a computational infrastructure for toxicological data storage, sharing, analysis, and management, as well as the creation of computational toxicological models for nanomaterials (Jeliazkova *et al.*, 2021). This database was designed based on the FAIR (findable, accessible, interoperable and reusable) guiding principles. It includes a wide variety of data types, including physicochemical, (eco)toxicological and exposure-related parameters in line with current regulatory requirements for the safety assessment of nanomaterials, as well as information derived from nonstandardized new approach methodologies, such as omics data. This database contains millions of data points from thousands of studies. This large database provides an ideal data source to apply machine learning approaches to build robust computational nanotoxicology models.

CHALLENGES AND FUTURE PERSPECTIVES

Machine learning and artificial intelligence approaches and the availability of many large toxicological databases present a great opportunity to advance the science of toxicology, especially in the paradigm shift from traditional animal-based risk assessment framework to the 21st century risk assessment framework that is primarily based on *in vitro* high throughput assays coupled with *in silico* modeling for IVIVE. This opportunity also comes along with multiple challenges.

First, with the advance of computer and mathematical sciences, there are more and more machine learning algorithms available to analyze toxicological data. Different algorithms have different requirements on the data size and types (eg, continuous vs categorical). Some algorithms may work better for certain data types, but others may not. Toxicology is an interdisciplinary science and has a variety of data types. There is no consensus on which machine learning algorithm that works the best for a certain data type or dataset. In order to develop the best machine learning model, usually researchers have to try different machine learning algorithms and compare their performances (Cheng and Ng, 2019; Lin *et al.*, 2022; McCoubrey *et al.*, 2021; Wang *et al.*, 2020). Once the best machine learning model is identified, it can then be used for model predictions and subsequent analyses.

Second, machine learning and artificial intelligence algorithms are commonly viewed as black boxes that are lack of mechanistical explainablity (Guha, 2008; Sjöberg *et al.*, 1995), which brings a certain challenge to application of machine learning models in toxicology. In order to overcome this limitation and make interpretable predictions, knowledge-based machine learning approaches should be developed. For example, with the use of the AOP structure and a set of *in vitro* HTS bioassays, a knowledge-based deep neural network model allows for a mechanism-based prediction of a chemical's estrogen receptor binding potential over traditional black-box models, which represents a significant advancement in computational toxicology (Ciallella *et al.*, 2021). However, more interpretable machine learning models with supporting mechanistic data remain to be developed.

Third, traditional machine learning approaches are limited in extracting critical features and are thus difficult to predict with a high accuracy. As more high-throughput data become available, these data often involve a large number of chemicals with multiple fingerprint descriptors. Considering each of the descriptors might lead to overfitting in many machine learning models and thus hindering the performance on model validation, but these limitations could be overcome by more advanced deep neural network models. With the efforts to control overfitting by automatically feature selection algorithms, the deep neural networks approach presents more effective predictability than traditional machine learning methods. In our recent study (Lin *et al.*, 2022), the deep learning model outperformed all traditional machine learning methods in the prediction of delivery efficiency of nanomedicines in tumor-bearing mice. With advancement in deep learning models, they have a multitude of benefits that have been shown to improve predictive power in the application in different areas of toxicology and pharmacology, including toxicogenomics (O'Donovan *et al.*, 2020) and HTS data (Pham *et al.*, 2021).

Fourth, current machine learning-based computational toxicology models are mostly based on bioactivity classification, ie, yes or no for bioactivity/toxicity, which cannot predict the intensity of toxic effect, dose-response relationship, or time-dependence (Table 2). There are only some models based on quantitative endpoints (eg, median lethal dose [LD_{50}]) (Feinstein *et al.*, 2021; Gadaleta *et al.*, 2019; Karim *et al.*, 2021). A fundamental tenet in the field of toxicology is "the dose makes the poison." In modern toxicology, toxicity varies depending on multiple factors, including exposure dose, time, target cell, species, and *in vitro* versus *in vivo*. More advanced machine learning models that can predict relative toxicity of environmental chemicals quantitatively based on different variables (eg, dose, time, and species) remain to be developed.

Fifth, although big data enable to develop robust machine learning models, there comes a risk of being overwhelmed by the flood of data, confounding by low quality data, and losing sight of the objective of the hazard or risk assessment to be undertaken (Richarz, 2019). With the increasing volume and generation speed of data, it is important to develop adequate infrastructure to store, share, analyze, evaluate, and manage data. It is recommended that data be generated following standard test guidelines, such as those recommended by OECD. Before choosing data to develop machine learning models, data quality, completeness, reliability, and relevance should be rigorously checked and if possible, modelers should choose high-quality, complete, reliable, and relevant data to develop machine learning models.

Sixth, although many machine learning studies have either developed novel computational models to predict toxicity or provide important insights in toxicology, these models are limited to some mathematical equations or computer codes that often do not share with the readers. This is, in part, because these computer codes are "intimidating" to nonmodelers. This limits the reproducibility of existing machine learning studies in toxicology. This issue is similar to many earlier PBPK modeling studies in which researchers did not share the model code. However, in the field of PBPK modeling, this issue is mostly resolved as now more and more PBPK modelers realize the importance of sharing model code and actually publish their model codes. Likewise, in the field of machine learning and artificial intelligence in toxicology, it is recommended that the entire machine learning codes that were used to train and test the model be published as part of the manuscript to facilitate reproducibility of study findings and to enable other researchers to develop better models based on published models.

Finally, although the studies discussed in this article show promising applications of machine learning and artificial intelligence approaches in different areas of toxicology, many of the cited applications are still relatively new and have not been actually used in industry or governmental agencies to support public health decision-making. Similar to other areas of biomedical sciences, it will take time for new methodology and applications to be standardized, validated, and then eventually adopted by the industry and regulatory agencies. Note that machine learning and artificial intelligence-based software products have been accepted as a medical device by U.S. FDA (FDA, 2021) and artificial intelligence approaches have started to be used in different stages of drug discovery and development processes (Paul *et al.*, 2021). It is anticipated that machine learning and artificial intelligence approaches will be increasingly applied in chemical toxicity and risk assessment by the industry and regulatory agencies in the future.

FUNDING

The authors would like to acknowledge funding support from United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) for the Food Animal Residue Avoidance Databank (FARAD) Program (2021-41480-35271); the United States National Institutes of Health (NIH) National Institute of Biomedical Imaging and Bioengineering (NIBIB) Research Grant Program (R01EB031022); and the New Faculty Start-up Funds from the University of Florida.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

Ai, H., Wu, X., Zhang, L., Qi, M., Zhao, Y., Zhao, Q., Zhao, J., and Liu, H. (2019). QSAR modelling study of the bioconcentration factor and toxicity of organic compounds to aquatic organisms using machine learning and ensemble methods. *Ecotoxicol. Environ. Saf.* **179**, 71–78.

Allen, T. E., Goodman, J. M., Gutsell, S., and Russell, P. J. (2014). Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. *Chem. Res. Toxicol.* **27**, 2100–2112.

Allen, T. E., Liggi, S., Goodman, J. M., Gutsell, S., and Russell, P. J. (2016). Using molecular initiating events to generate 2D structure-activity relationships for toxicity screening. *Chem. Res. Toxicol.* **29**, 1611–1627.

Allen, T. E. H., Goodman, J. M., Gutsell, S., and Russell, P. J. (2018). Using 2D structural alerts to define chemical categories for molecular initiating events. *Toxicol. Sci.* **165**, 213–223.

Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., *et al.* (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* **29**, 730–741.

Aschner, M., Mesnage, R., Docea, A. O., Paoliello, M. M. B., Tsatsakis, A., Giannakakis, G., Papadakis, G. Z., Vinceti, S. R., Santamaria, A., Skalny, A. V., *et al.* (2022). Leveraging artificial intelligence to advance the understanding of chemical neurotoxicity. *Neurotoxicology* **89**, 9–11.

Atila, Ü., Baydilli, Y. Y., Sehirli, E., and Turan, M. K. (2020). Classification of DNA damages on segmented comet assay images using convolutional neural network. *Comput. Methods Programs Biomed.* **186**, 105192.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., *et al.* (2013). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–995.

Baskin, I. I. 2018. Chapter 5—Machine learning methods in computational toxicology. In *Computational Toxicology: Methods and Protocols, Methods in Molecular Biology* (N. Orazio, Ed.), pp. 119–139 Springer Science+Business Media, LLC, Part of Springer Nature, Berlin/Heidelberg, Germany.

Bhhatarai, B., Walters, W. P., Hop, C., Lanza, G., and Ekins, S. (2019). Opportunities and challenges using artificial intelligence in ADME/Tox. *Nat. Mater.* **18**, 418–422.

Bitsch, A., Jacobi, S., Melber, C., Wahnschaffe, U., Simetska, N., and Mangelsdorf, I. (2006). REPDOSE: A database on repeated dose toxicity studies of commercial chemicals—A multifunctional tool. *Regul. Toxicol. Pharmacol.* **46**, 202–210.

Chen, Q., Chou, W. C., and Lin, Z. (2022a). Integration of toxicogenomics and physiologically based pharmacokinetic modeling in human health risk assessment of perfluorooctane sulfonate. *Environ. Sci. Technol.* **56**, 3623–3633.

Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud D. (2018). Neural ordinary differential equations. *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, Canada, pp. 6571–6583.

Chen, X., Dang, L., Yang, H., Huang, X., and Yu, X. (2020). Machine learning-based prediction of toxicity of organic compounds towards fathead minnow. *RSC Adv.* **10**, 36174–36180.

Chen, X., Roberts, R., Tong, W., and Liu, Z. (2022b). Tox-GAN: An AI approach alternative to animal studies-a case study with toxicogenomics. *Toxicol. Sci.* **186**, 242–259.

Cheng, W., and Ng, C. A. (2019). Using machine learning to classify bioactivity for 3486 per- and polyfluoroalkyl substances (PFASs) from the OECD list. *Environ. Sci. Technol.* **53**, 13970–13980.

Cheng, Y. H., He, C., Riviere, J. E., Monteiro-Riviere, N. A., and Lin, Z. (2020). Meta-analysis of nanoparticle delivery to tumors using a physiologically based pharmacokinetic modeling and simulation approach. *ACS Nano* **14**, 3075–3095.

Chou, W. C., and Lin, Z. Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling. *Toxicol. Sci.* Forthcoming.

Ciallella, H. L., Russo, D. P., Aleksunes, L. M., Grimm, F. A., and Zhu, H. (2021). Revealing adverse outcome pathways from public high-throughput screening data to evaluate new toxicants by a knowledge-based deep neural network approach. *Environ. Sci. Technol.* **55**, 10875–10887.

Ciallella, H. L., Russo, D. P., Sharma, S., Li, Y., Sloter, E., Sweet, L., Huang, H., and Zhu, H. (2022). Predicting prenatal developmental toxicity based on the combination of chemical structures and biological data. *Environ. Sci. Technol.* **56**, 5984–5998.

Ciallella, H. L., and Zhu, H. (2019). Advancing computational toxicology in the big data era by artificial intelligence: Data-driven and mechanism-driven modeling for chemical toxicity. *Chem. Res. Toxicol.* **32**, 536–547.

Davidovic, L. M., Laketic, D., Cumic, J., Jordanova, E., and Pantic, I. (2021). Application of artificial intelligence for detection of chemico-biological interactions associated with oxidative stress and DNA damage. *Chem. Biol. Interact.* **345**, 109533.

Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., Wiegers, J., Wiegers, T. C., and Mattingly, C. J. (2021). Comparative toxicogenomics database (CTD): Update 2021. *Nucleic Acids Res*. **49**, D1138–D1143.

FDA. (2021). Artificial intelligence and machine learning in software as a medical device. U.S. Food and Drug Administration (FDA). Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device. Accessed June 6, 2022.

Feinstein, J., Sivaraman, G., Picel, K., Peters, B., Vazquez-Mayagoitia, A., Ramanathan, A., MacDonell, M., Foster, I., and Yan, E. (2021). Uncertainty-informed deep transfer learning of perfluoroalkyl and polyfluoroalkyl substance toxicity. *J. Chem. Inf. Model*. **61**, 5793–5803.

Fisher, J. W., Gearhart, J., and Lin, Z. (2020). *Physiologically Based Pharmacokinetic (PBPK) Modeling—Methods and Applications in Toxicology and Risk Assessment*, 1st ed, pp. 1–346. Elsevier, Amsterdam, Netherlands.

Fonger, G. C., Stroup, D., Thomas, P. L., and Wexler, P. (2000). TOXNET: A computerized collection of toxicological and environmental health information. *Toxicol. Ind. Health*. **16**, 4–6.

Gadaleta, D., Vukovic, K., Toma, C., Lavado, G. J., Karmaus, A. L., Mansouri, K., Kleinstreuer, N. C., Benfenati, E., and Roncaglioni, A. (2019). SAR and QSAR modeling of a large collection of LD50 rat acute oral toxicity data. *J. Cheminform*. **11**, 58.

Gajewicz-Skretna, A., Furuhama, A., Yamamoto, H., and Suzuki, N. (2021). Generating accurate in silico predictions of acute aquatic toxicity for a range of organic chemicals: Towards similarity-based machine learning methods. *Chemosphere* **280**, 130681.

Ganter, B., Tugendreich, S., Pearson, C. I., Ayanoglu, E., Baumhueter, S., Bostian, K. A., Brady, L., Browne, L. J., Calvin, J. T., Day, G. J., *et al.* (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol*. **119**, 219–244.

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., *et al.* (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. **40**, D1100–1107.

Glass, C., Lafata, K. J., Jeck, W., Horstmeyer, R., Cooke, C., Everitt, J., Glass, M., Dov, D., and Seidman, M. A. (2022). The role of machine learning in cardiovascular pathology. *Can. J. Cardiol*. **38**, 234–245.

Guha, R. (2008). On the interpretation and interpretability of quantitative structure-activity relationship models. *J. Comput. Aided Mol. Des*. **22**, 857–871.

Hu, F., Santagostino, S. F., Danilenko, D. M., Tseng, M., Brumm, J., Zehnder, P., and Wu, K. C. (2022). Assessment of skin toxicity in an in vitro reconstituted human epidermis model using deep learning. *Am. J. Pathol*. **192**, 687–700.

Huang, R., Sakamuru, S., Martin, M. T., Reif, D. M., Judson, R. S., Houck, K. A., Casey, W., Hsieh, J. H., Shockley, K. R., Ceger, P., *et al.* (2014). Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. *Sci. Rep*. **4**, 5664.

Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., and Yamada, H. (2015). Open TG-GATEs: A large-scale toxicogenomics database. *Nucleic Acids Res*. **43**, D921–927.

Jeliazkova, N., Apostolova, M. D., Andreoli, C., Barone, F., Barrick, A., Battistelli, C., Bossa, C., Botea-Petcu, A., Chatel, A., De Angelis, I., *et al.* (2021). Towards FAIR nanosafety data. *Nat. Nanotechnol*. **16**, 644–654.

Jeliazkova, N., Chomenidis, C., Doganis, P., Fadeel, B., Grafstrom, R., Hardy, B., Hastings, J., Hegi, M., Jeliazkov, V., Kochev, N., *et al.* (2015). The eNanoMapper database for nanomaterial safety information. *Beilstein J. Nanotechnol*. **6**, 1609–1634.

Ji, Z., Guo, W., Wood, E. L., Liu, J., Sakkiah, S., Xu, X., Patterson, T. A., and Hong, H. (2022). Machine learning models for predicting cytotoxicity of nanomaterials. *Chem. Res. Toxicol*. **35**, 125–139.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T. R., Spencer, R., and Wolf, M. (2008). ACToR—Aggregated computational toxicology resource. *Toxicol. Appl. Pharmacol*. **233**, 7–13.

Kamiya, Y., Handa, K., Miura, T., Yanagi, M., Shigeta, K., Hina, S., Shimizu, M., Kitajima, M., Shono, F., Funatsu, K., *et al.* (2021). In silico prediction of input parameters for simplified physiologically based pharmacokinetic models for estimating plasma, liver, and kidney exposures in rats after oral doses of 246 disparate chemicals. *Chem. Res. Toxicol*. **34**, 507–513.

Karim, A., Riahi, V., Mishra, A., Newton, M. A. H., Dehzangi, A., Balle, T., and Sattar, A. (2021). Quantitative toxicity prediction via meta ensembling of multitask deep learning models. *ACS Omega*. **6**, 12306–12317.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., *et al.* (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res*. **49**, D1388–D1395.

Klaassen, C. D. 2018. *Casarett & Doull's Toxicology: The Basic Science of Poisons*, 9th ed. McGraw Hill, New York, NY, pp. 1–1648.

Lea, I. A., Gong, H., Paleja, A., Rashid, A., and Fostel, J. (2017). CEBS: A comprehensive annotated database of toxicological data. *Nucleic Acids Res*. **45**, D964–D971.

Li, T., Tong, W., Roberts, R., Liu, Z., and Thakkar, S. (2021). DeepCarc: Deep learning-powered carcinogenicity prediction using model-level representation. *Front. Artif. Intell*. **4**, 757780.

Lin, Y. J., and Lin, Z. (2020). In vitro-in silico-based probabilistic risk assessment of combined exposure to bisphenol A and its analogues by integrating ToxCast high-throughput in vitro assays with in vitro to in vivo extrapolation (IVIVE) via physiologically based pharmacokinetic (PBPK) modeling. *J. Hazard. Mater*. **399**, 122856.

Lin, Z., Chou, W. C., Cheng, Y. H., He, C., Monteiro-Riviere, N. A., and Riviere, J. E. (2022). Predicting nanoparticle delivery to tumors using machine learning and artificial intelligence approaches. *Int. J. Nanomed*. **17**, 1365–1379.

Lin, Z., and Fisher, J. W. 2020. Chapter 1—A history and recent efforts of selected physiologically based pharmacokinetic modeling topics. In *Physiologically Based Pharmacokinetic (PBPK) Modeling—Methods and Applications in Toxicology and Risk Assessment*, 1st ed. (J. W. Fisher, J. M. Gearhart, and Z. Lin, Eds.), pp. 1–26. Elsevier, Amsterdam, Netherlands.

Liu, Z., Huang, R., Roberts, R., and Tong, W. (2019). Toxicogenomics: A 2020 vision. *Trends Pharmacol. Sci*. **40**, 92–103.

Lu, J., Bender, B., Jin, J. Y., and Guan, Y. (2021a). Deep learning prediction of patient response time course from early data via neural-pharmacokinetic/pharmacodynamic modelling. *Nat. Mach. Intell*. **3**, 696–704.

Lu, J., Deng, K., Zhang, X., Liu, G., and Guan, Y. (2021b). Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *iScience* **24**, 102804.

Luechtefeld, T., Maertens, A., Russo, D. P., Rovida, C., Zhu, H., and Hartung, T. (2016). Global analysis of publicly available safety data for 9,801 substances registered under REACH from 2008–2014. *ALTEX* **33**, 95–109.

Luechtefeld, T., Marsh, D., Rowlands, C., and Hartung, T. (2018). Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility. *Toxicol. Sci.* **165**, 198–212.

Mansouri, K., Karmaus, A. L., Fitzpatrick, J., Patlewicz, G., Pradeep, P., Alberga, D., Alepee, N., Allen, T. E. H., Allen, D., Alves, V. M., *et al.* (2021). CATMoS: Collaborative acute toxicity modeling suite. *Environ. Health Perspect.* **129**, 47013.

Mayr, A., Klambauer, G., Unterthiner, T., and Hochreiter, S. (2016). DeepTox: Toxicity prediction using deep learning. *Front. Environ. Sci.* **3**, 80.

McCoubrey, L. E., Elbadawi, M., Orlu, M., Gaisford, S., and Basit, A. W. (2021). Machine learning uncovers adverse drug effects on intestinal bacteria. *Pharmaceutics* **13**, 1026.

Nair, S. K., Eeles, C., Ho, C., Beri, G., Yoo, E., Tkachuk, D., Tang, A., Nijrabi, P., Smirnov, P., Seo, H., *et al.* (2020). ToxicoDB: An integrated database to mine and visualize large-scale toxicogenomic datasets. *Nucleic Acids Res.* **48**, W455–W462.

NIEHS. (2022). Environmental health topics—Toxicology. National Institute of Environmental Health Sciences (NIEHS), Durham, NC. Available at: https://www.niehs.nih.gov/health/topics/science/toxicology/index.cfm. Accessed March 23, 2022.

O'Donovan, S. D., Driessens, K., Lopatta, D., Wimmenauer, F., Lukas, A., Neeven, J., Stumm, T., Smirnov, E., Lenz, M., Ertaylan, G., *et al.* (2020). Use of deep learning methods to translate drug-induced gene expression changes from rat to human primary hepatocytes. *PLoS One* **15**, e0236392.

OECD. (2014). *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*, OECD Series on Testing and Assessment, No. 69. OECD Publishing, Paris. https://doi.org/10.1787/9789264085442-en.

OECD. (2022). Adverse outcome pathways, molecular screening and toxicogenomics. Organisation for Economic Co-operation and Development (OECD). Available at: https://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm. Accessed March 4, 2022.

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., and Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discov. Today.* **26**, 80–93.

Perkins, E. J., Ashauer, R., Burgoon, L., Conolly, R., Landesmann, B., Mackay, C., Murphy, C. A., Pollesch, N., Wheeler, J. R., Zupanic, A., *et al.* (2019). Building and applying quantitative adverse outcome pathway models for chemical hazard and risk assessment. *Environ. Toxicol. Chem.* **38**, 1850–1865.

Pham, T. H., Qiu, Y., Zeng, J., Xie, L., and Zhang, P. (2021). A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nat. Mach. Intell.* **3**, 247–257.

Pizzino, G., Irrera, N., Cucinotta, M., Pallio, G., Mannino, F., Arcoraci, V., Squadrito, F., Altavilla, D., and Bitto, A. (2017). Oxidative stress: Harms and benefits for human health. *Oxid. Med. Cell. Longev.* **2017**, 8416763.

Pradeep, P., Patlewicz, G., Pearce, R., Wambaugh, J., Wetmore, B., and Judson, R. (2020). Using chemical structure information to develop predictive models for in vitro toxicokinetic parameters to inform high-throughput risk-assessment. *Comput. Toxicol.* **16**, 100136.

Pu, L., Naderi, M., Liu, T., Wu, H. C., Mukhopadhyay, S., and Brylinski, M. (2019). eToxPred: A machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacol. Toxicol.* **20**, 2.

Rahman, S. M., Lan, J., Kaeli, D., Dy, J., Alshawabkeh, A., and Gu, A. Z. (2022). Machine learning-based biomarkers identification from toxicogenomics - Bridging to regulatory relevant phenotypic endpoints. *J. Hazard. Mater.* **423**, 127141.

Reddy, M., Yang, R. S., Andersen, M. E., and Clewell, H. J. 2005. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*, 1st ed., pp. 1–420. Wiley, Hoboken, NJ.

Richarz, AN. 2019. Chapter 1: Big data in predictive toxicology: Challenges, opportunities and perspectives. In *Big Data in Predictive Toxicology* (D. Neagu, A.-N. Richarz, Ed.), pp. 1–37. Royal Society of Chemistry, London, United Kingdom.

Richman, T., Arnold, E., and Williams, A. J. (2022). Curation of a list of chemicals in biosolids from EPA National Sewage Sludge Surveys & Biennial Review Reports. *Sci. Data.* **9**, 180.

Russell, S., and Norvig, P. 2020. *Artificial Intelligence: A Modern Approach*, 4th ed., pp. 1–1069. Pearson Education, Inc., London, United Kingdom.

Russo, D. P., Strickland, J., Karmaus, A. L., Wang, W., Shende, S., Hartung, T., Aleksunes, L. M., and Zhu, H. (2019). Nonanimal models for acute toxicity evaluations: Applying data-driven profiling and read-across. *Environ. Health Perspect.* **127**, 47001.

Singh, A. V., Ansari, M. H. D., Rosenkranz, D., Maharjan, R. S., Kriegel, F. L., Gandhi, K., Kanase, A., Singh, R., Laux, P., and Luch, A. (2020). Artificial intelligence and machine learning in computational nanotoxicology: Unlocking and empowering nanomedicine. *Adv. Healthcare Mater.* **9**, e1901862.

Sinitsyn, D., Garcia-Reyero, N., and Watanabe, K. H. (2022). From qualitative to quantitative AOP: A case study of neurodegeneration. *Front. Toxicol.* **4**, 838729.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P. Y., Hjalmarsson, H., and Juditsky, A. (1995). Nonlinear black-box modeling in system identification: A unified overview. *Automatica* **31**, 1691–1724.

Spinu, N., Cronin, M. T. D., Enoch, S. J., Madden, J. C., and Worth, A. P. (2020). Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. *Arch. Toxicol.* **94**, 1497–1510.

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., *et al.* (2017). A next generation connectivity map: l 1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e1417.

Tan, Y. M., Worley, R. R., Leonard, J. A., and Fisher, J. W. (2018). Challenges associated with applying physiologically based pharmacokinetic modeling for public health decision-making. *Toxicol. Sci.* **162**, 341–348.

Tox21. (2020). United States Federal Government Tox21 collaboration advancing toxicology to improve environmental health and pharmaceutical safety. Tox21 Fact Sheet. Available at: https://tox21.gov/wp-content/uploads/2019/04/Tox21_FactSheet_Apr2019.pdf. Accessed April 19, 2022.

Vinken, M., Pauwels, M., Ates, G., Vivier, M., Vanhaecke, T., and Rogiers, V. (2012). Screening of repeated dose toxicity data present in SCC(NF)P/SCCS safety evaluations of cosmetic ingredients. *Arch. Toxicol.* **86**, 405–412.

Wang, C. C., Liang, Y. C., Wang, S. S., Lin, P., and Tung, C. W. (2022). A machine learning-driven approach for prioritizing food contact chemicals of carcinogenic concern based on complementary in silico methods. *Food Chem. Toxicol.* **160**, 112802.

Wang, Y. W., Huang, L., Jiang, S. W., Li, K., Zou, J., and Yang, S. Y. (2020). CapsCarcino: A novel sparse data deep learning tool for predicting carcinogens. *Food Chem. Toxicol.* **135**, 110921.

Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K., and Pande, V. (2018). MoleculeNet: A benchmark for molecular machine learning. *Chem. Sci.* **9**, 513–530.

Xu, X., Zhao, P., Wang, Z., Zhang, X., Wu, Z., Li, W., Tang, Y., and Liu, G. (2021). In silico prediction of chemical acute contact toxicity on honey bees via machine learning methods. *Toxicol. In Vitro* **72**, 105089.

Yan, X., Sedykh, A., Wang, W., Yan, B., and Zhu, H. (2020). Construction of a web-based nanomaterial database by big data curation and modeling friendly nanostructure annotations. *Nat. Commun.* **11**, 2519.

Zgheib, E., Gao, W., Limonciel, A., Aladjov, H., Yang, H., Tebby, C., Gayraud, G., Jennings, P., Sachana, M., Beltman, J. B., *et al.* (2019). Application of three approaches for quantitative AOP development to renal toxicity. *Comput. Toxicol.* **11**, 1–13.

Zhang, L., Ai, H., Chen, W., Yin, Z., Hu, H., Zhu, J., Zhao, J., Zhao, Q., and Liu, H. (2017). CarcinoPred-EL: Novel models for predicting the carcinogenicity of chemicals using molecular fingerprints and ensemble learning methods. *Sci. Rep.* **7**, 2118.

Zhu, H. (2020). Big data and artificial intelligence modeling for drug discovery. *Annu. Rev. Pharmacol. Toxicol.* **60**, 573–589.