Environmental Science & Technology

Open Access

This article is licensed under CC-BY 4.0

pubs.acs.org/est

Critical Review

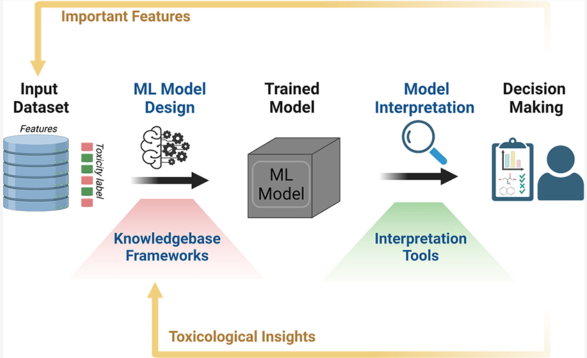
Advancing Computational Toxicology by Interpretable Machine Learning

Xuelian Jia, Tong Wang, and Hao Zhu*

Cite This: Environ. Sci. Technol. 2023, 57, 17690−17706

Read Online

ACCESS | Metrics & More | Article Recommendations



ABSTRACT: Chemical toxicity evaluations for drugs, consumer products, and environmental chemicals have a critical impact on human health. Traditional animal models to evaluate chemical toxicity are expensive, time-consuming, and often fail to detect toxicants in humans. Computational toxicology is a promising alternative approach that utilizes machine learning (ML) and deep learning (DL) techniques to predict the toxicity potentials of chemicals. Although the applications of ML- and DL-based computational models in chemical toxicity predictions are attractive, many toxicity models are "black boxes" in nature and difficult to interpret by toxicologists, which hampers the chemical risk assessments using these models. The recent progress of interpretable ML (IML) in the computer science field meets this urgent need to unveil the underlying toxicity mechanisms and elucidate the domain knowledge of toxicity models. In this review, we focused on the applications of IML in computational toxicology, including toxicity feature data, model interpretation methods, use of knowledge base frameworks in IML development, and recent applications. The challenges and future directions of IML modeling in toxicology are also discussed. We hope this review can encourage efforts in developing interpretable models with new IML algorithms that can assist new chemical assessments by illustrating toxicity mechanisms in humans.

KEYWORDS: *Machine learning, Interpretable modeling, Computational toxicology, Risk assessment, Systems toxicology, Adverse outcome pathway*

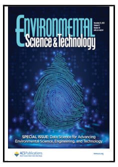
1. INTRODUCTION

Chemical risk assessments and safety testing are important for the early identification of hazardous chemicals in multiple industry sectors.[1−3] For example, in the drug development procedure, toxicity evaluation in the early stage can reduce the attrition rate and late failure, which significantly reduce the cost of developing a new drug.[4,5] Traditional toxicity evaluations for pharmaceuticals, xenobiotics, and environmental chemicals often involve the use of toxicological tests conducted in animal models, which are expensive, time-consuming, and raise concerns about animal welfare. The rapidly increasing number of chemicals in medical, industrial, and agricultural fields has made it impractical to use animal models for evaluating tens of thousands of new chemicals.[6,7] As an alternative strategy, computational toxicology using machine learning (ML) techniques has shown promise for chemical toxicity evaluations because it can quickly predict the toxicity of a large number of new compounds in the risk assessment process and prioritize potentially hazardous compounds for experimental testing.[8] In the National Research Council (NRC) 2007 report *Toxicity Testing in the 21st Century: A Vison and a Strategy*, the development of computational techniques for risk assessment was emphasized.[9,10] In 2016, the Frank R. Lautenberg Chemical Safety for the 21st Century Act (LCSA) was approved to advance chemical risk assessment. The LCSA called for computational approaches and strategies for safety evaluation to reduce or replace the use of vertebrate animals while providing evidence to support regulatory decisions.[11] In chemical industries (e.g., consumer products), the use of computational models for chemical toxicity assessments is also important for decision-making during product development.[12]

In the past decade, the development of new experimental protocols, especially high-throughput screening (HTS) assays, and the progress of combinatorial chemistry has generated toxicity data for millions of compounds.[8,9] With the development of advanced ML and deep learning (DL) algorithms,

Special Issue: Data Science for Advancing Environmental Science, Engineering, and Technology

Received: January 25, 2023
Revised: May 5, 2023
Accepted: May 5, 2023
Published: May 24, 2023



ACS Publications

https://doi.org/10.1021/acs.est.3c00653

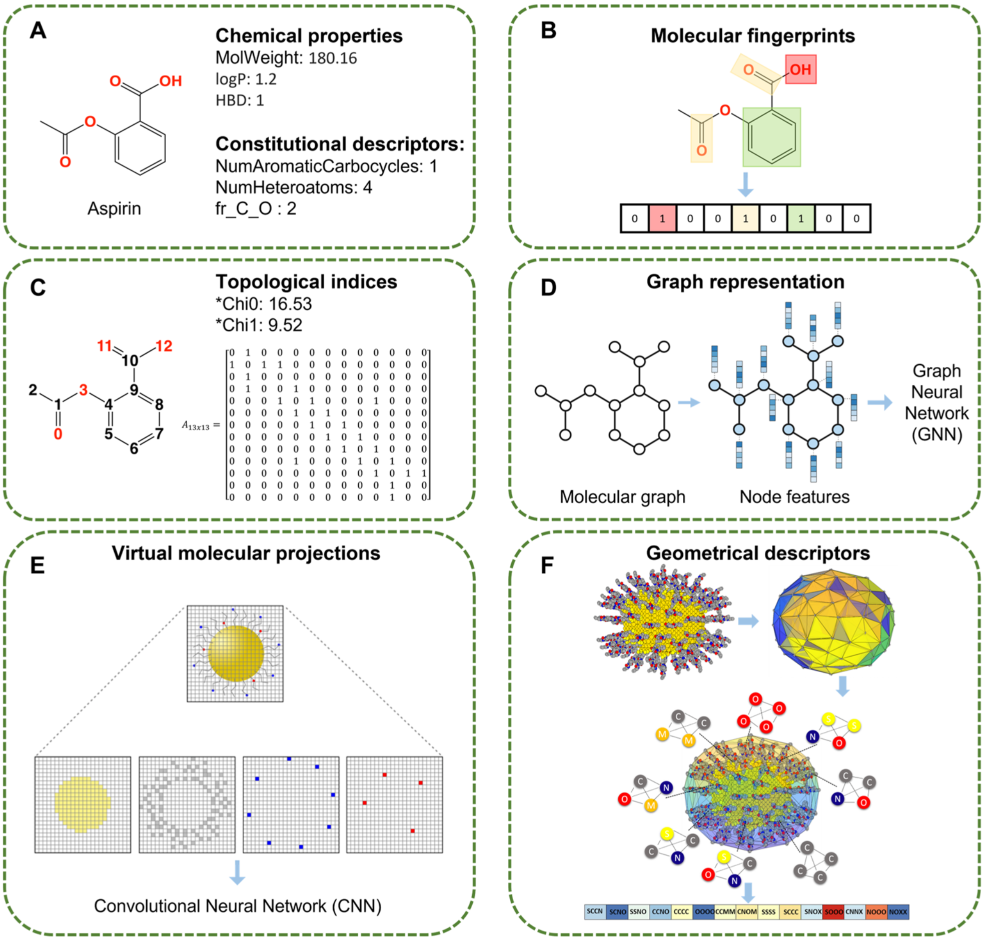

A
Chemical properties
MolWeight: 180.16
logP: 1.2
HBD: 1
Constitutional descriptors:
NumAromaticCarbocycles: 1
NumHeteroatoms: 4
fr_C_O : 2
O
OH
O
O
Aspirin
B
Molecular fingerprints
O
OH
O
O
0 1 0 0 1 0 1 0 0
C
Topological indices
*Chi0: 16.53
*Chi1: 9.52
11
12
10
9
8
7
6
5
4
3
2
1
0
$A_{13x13} = \begin{bmatrix} 0&1&0&0&0&0&0&0&0&0&0&0&0 \\ 1&0&1&1&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&1&0&0&0&1&0&0&0 \\ 0&0&0&0&1&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&0&0&0&1&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&1&1 \\ 0&0&0&0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0&0 \end{bmatrix}$
D
Graph representation
Graph Neural Network (GNN)
Molecular graph
Node features
E
Virtual molecular projections
Convolutional Neural Network (CNN)
F
Geometrical descriptors
C
M
O
N
S
SCCN SCNO SSNO CCNO CCCC OOOO CCMM CNOM SSSS SCCC SNOX SOOO CNNX NOOO NOXX

Figure 1. Examples of different types of chemical descriptors. (A) Chemical properties and constitutional descriptors. (B) Molecular fingerprints as a vector of bits that denotes the presence "1" or absence "0" of a specific structural feature. (C) Topological indices as global features that derive information from the adjacency matrix of the molecular graph of a chemical (*calculated on the basis of eqs 1, 9, 10, and 1, 11, and 12 of ref 56). (D) Graph representation of a molecule including node features (e.g., atom type and aromaticity), adjacency matrix, and edge features, which can be used as inputs to a Graph Neural Network (GNN). (E) Virtual molecular projections of nanoparticles (Reprinted with permission from ref 51. Copyright 2020 American Chemical Society). (F) Geometrical surface descriptors of nanoparticles (Reprinted with permission from ref 52. Copyright 2019 Royal Society of Chemistry).

computational modeling can use massive toxicity data for more accurate chemical toxicity predictions. Some DL models have been shown to match or even outperform other ML algorithms on prediction accuracy.[13−16] However, a common limitation of complex ML models, especially DL models using neural network architecture, is their "black box" nature, which means their inner working mechanisms cannot be easily understood by users.[17] There is an increasing demand for developing strategies to help toxicologists understand the model and how the predictions are made. The development of interpretable ML (IML) is an effective approach to mitigate the lack of interpretability underlying a trained model to reveal underlying toxicity mechanisms and to augment decision-making.

ML models are being ubiquitously used in daily lives (e.g., hiring, advertising, and music recommendations)[18−20] and health-related fields (e.g., healthcare, risk assessment, and drug discovery).[21−23] However, without an understanding of working mechanisms, black box models can lead to mistrust of the results.[24] In health-related fields, black box models can negatively affect human health, racial bias, and safety.[25−28] For example, Obermeyer et al. uncovered a racial bias issue in a widely used model for predicting health needs.[26] Some pollution models incorrectly predicted highly polluted air as nonhazardous to humans, due to the unknown working mechanism of the models.[27] Such negative consequences caused by black-box ML models can be avoided by developing IML with increased transparency and interpretability.[29,30] Lundberg et al. reported the use of IML for the prevention of hypoxemia during surgery, which can provide explanations of the risk factors in real-time and increase the anesthesiologists' anticipation of hypoxemia events by 15%.[31] Recent applications of IML in the healthcare field showed IML's potential in detecting bias and ensuring the interpretability of the model while maintaining accuracy.[32−35]

IML can perform robust model validations to avoid making wrong decisions learned from biased training data and make the underlying decision-making understandable.[36−39] Ideally, besides the predictions made by a model, the knowledge about chemical toxicants in the training data can help toxicologists better evaluate the trained model and make decisions on new compounds, i.e., determining which environmental chemicals and drugs are of the greatest potential concern to human health.[40] Because different scientific communities use ML for different prediction tasks, there is no universal definition of IML.[36] Regarding the chemical toxicity assessments, the desired IML models need to fulfill the following criteria:

https://doi.org/10.1021/acs.est.3c00653
Environ. Sci. Technol. 2023, 57, 17690−17706

Table 1. Publicly Available Big Data Repositories for Toxicology Modeling[a]

database	data type	description
PubChem[69,70]	chemical molecules, biological activities	Over 110 million compounds and over 1.5 million bioassays related to toxicity, genomics, and literature data
ChEMBL[85]	chemical molecules with druglike properties, bioactivity, and genomic data	Contains a total of >15 million bioactivity measurements for 1.8 million compounds and pharmacokinetic measurements for 85 drugs
ToxCast[63,64]	toxicity-related *in vitro* assays	Phase I: 309 compounds (mostly pesticides) tested in ∼500 HTS assays in human primary cells, cell lines, and rat primary hepatocytes
		Phase II: additional 776 compounds, including failed pharmaceuticals tested in ∼700 HTS assays
Tox 21[61,62]	toxicity-related *in vitro* assays	Phase I: ∼2800 compounds tested in ∼75 nuclear receptor and stress response pathway assays
		Phase II: expand to ∼10 000 chemicals including all marketed pharmaceuticals
Integrated Chemical Environment (ICE)[86−88]	curated *in vivo* and *in vitro* toxicity data	Provides high-quality curated *in vivo* and *in vitro* test data, reference chemical lists, and computational tools for chemical characterization and toxicity prediction. Curated HTS (cHTS) data in ICE are curated to bolster data confidence and are annotated to mechanistic targets
Gene Expression Omnibus (GEO)[82,83]	genomics data	Public repository that archives and freely distributes genomics data submitted by the research community and stores approximately a billion gene expression measurements from over 100 organisms
Open TG-GATEs[77]	*in vivo* and *in vitro* toxicogenomic data	Microarray gene expression data of 170 compounds tested in primary rat and human hepatocytes and *in vivo* rat liver and kidney with multiple treatment durations at three different doses
DrugMatrix[78]	*in vivo* and *in vitro* toxicogenomic data	Microarray gene expression data of ∼600 different compounds in rat hepatocytes and up to eight rat tissues with different durations of exposure
LINCS L1000[81]	*in vitro* toxicogenomic data	High-throughput measurements of 978 landmark gene set were used to infer the whole genome-wide expression (11 350 inferred genes), with ∼20,000 compounds tested at a variety of time points, doses, and in nine core human cell lines
Immunological Genome Project (ImmGen)[89]	genes and their networks in the immune system	Largest publicly available compendium of genome-wide transcriptional expression profiles for ∼250 distinct immunological cell states in mice

[a] LINCS, NIH Library of Integrated Network-Based Cellular Signatures; Open TG-GATES, Open Toxicogenomics Project-Genomics-Assisted Toxicity Evaluation System.

https://doi.org/10.1021/acs.est.3c00653

- Models should be constructed by explicit/understandable architectures.
- Users can understand how the model reached a specific prediction.
- Models can provide toxicity knowledge insights to support decision-making.

Recent efforts to develop IML and facilitate its application in toxicology are discussed in this review. We first overview feature data of chemicals that can be used for toxicity model development. Then, examples of computational algorithms and their specific interpretation methods are presented, followed by strategies for explaining black box models. The use of biological and toxicological knowledge in guiding the design of IML models is then discussed. Finally, we conclude with potential challenges in the practical applications of IML in toxicology.

2. FEATURE DATA IN COMPUTATIONAL TOXICOLOGY MODELING

ML is the technique to build predictive models by learning from input feature data using computational algorithms.[41,42] A typical procedure to develop a predictive model includes data collection and curation, model building, and model validation. Different types of data present different features and different levels of interpretability for users.[36] Unlike raw data (such as pixel units in an image),[29] most of the data used in toxicity modeling reflect the properties/activities of chemicals, which possess the ability to be interpreted during the modeling process and/or after the model has been developed. Therefore, the training data consisting of different feature types in toxicological modeling is the base of IML.

2.1. Structural/Chemical Properties. For toxicity modeling, the most intuitive data to be used is the chemical structure information. To make it machine-readable during modeling, chemical structures need to be transformed into vectors of numerical or binary values.[36] Quantitative structure−activity relationship (QSAR), a statistical approach that correlates a compound's chemical structural or physicochemical properties to its activities, has been used traditionally for chemical toxicity modeling.[43] The molecule structures were normally transferred into molecular descriptors at the beginning of a modeling procedure. The calculated molecular descriptors can represent local or global salient characteristics of the structures (Figure 1). Major classes of descriptors include (a) physicochemical descriptors (such as molecular weight, lipophilicity, etc.; Figure 1A) representing properties determining the absorption and distribution of chemicals in the body; (b) fingerprints, which are binary bits representing the presence "1" or absence "0" of substructures and molecular features of interest (Figure 1B); (c) constitutional descriptors representing the counts for corresponding atoms, bonds, and functional groups (Figure 1A); (d) geometrical descriptors capturing the three-dimensional structure features, such as the molecular size and shape (Figure 1E); and (e) atom distributions and topological indices representing the connectivity of atoms in the molecules (Figure 1C).[44,45] The structure information on chemicals can be stored in various formats, including linear representations such as SMILES (Simplified Molecular Input Line Entry Specification) and InChI (the IUPAC International Chemical Identifier) and connection table-based file formats such as SDF (Structure Data Format).[46,47] These chemical structure data can be accessed from chemical data-sharing repositories, such as PubChem and ChEMBL (Table 1), and can be further processed by cheminformatics tools (e.g., RDKit, http://www.rdkit.org/, accessed January 2023) to generate descriptors for toxicological modeling.

Besides structural/chemical descriptors, chemical molecules can be treated as graphs, where graph embedding techniques are applied to generate feature vectors for modeling. During the graph embedding process, chemical molecules are first represented as undirected graphs with atoms as nodes and edges as bonds (Figure 1D), and then, embedding techniques are introduced to obtain a spatial graph matrix for each atom.[48−50] These embeddings are used as features to train a model. The chemical molecules can also be transformed to image-like data. For example, in a nanotoxicity study, nanoparticle structures were transformed into "virtual molecular projections" (Figure 1E), which are multidimensional digital data representing the components of a nanoparticle structure without losing critical structure information.[51] The atomic coordinates of a virtual nanoparticle are projected onto a 2D space on the basis of the atom type and coordinates in 3D space. These projections were then used as inputs to predict the properties and activities of nanoparticles using an image processing convolutional neural network (CNN). Structure annotation techniques, such as Delaunay tessellation, which decomposes the surface of nanostructures into tetrahedra, have been developed to generate nanodescriptors that simulate surface chemistry and properties of complex nanoparticle structures (Figure 1F).[52,53] Overall, structure-based modeling, such as QSAR, is reliable in predicting some pharmacokinetic properties and *in vitro* assay responses with simple mechanisms for new compounds.[54,55] However, for complex toxicity endpoints (e.g., carcinogenicity and hepatotoxicity), the use of only structural information and chemical properties for modeling (i.e., QSAR) is error-prone, particularly when compounds with similar structures or chemical properties exhibit dissimilar toxicities.[5]

2.2. Pathway-Based Toxicity Data. In the past decade, the development of automatic experimental screening technology has significantly enhanced the efficiency of *in vitro* biomolecular or cell-based assays, thereby resulting in the HTS technique capable of screening thousands to millions of compounds.[57,58] The adverse outcome pathway (AOP) is a conceptual framework that links chemical-induced responses at the molecular, cellular, and organ levels to adverse outcomes at the organism level. Mechanism-based assay outcomes can be used within an AOP pathway to systematically assess whether a compound is likely to induce the target adverse outcome.[7,59] The chemical responses obtained from target-specific, mechanism-oriented *in vitro* assays in HTS projects like ToxCast and Tox21[60−63] keep growing and have contributed to the current toxicity big data (Table 1). Using a ToxCast/Tox21 assay, compounds were tested in multiple concentrations to generate concentration−response curves defining compound activity.[62,64] Then, statistical analysis was performed to define mechanistic outcomes for tested compounds, such as receptor binding, inhibition, and activation representing key events of a toxicity pathway. The outcomes can be used as biological descriptors of chemicals, which can be further combined with molecular descriptors to improve the ML models.[65−67] Moreover, quantitative outcomes from the concentration−response curves of active chemicals, such as half-maximal response concentration (AC_{50}) or lowest effective concentration (LEC), can be used for the extrapolation of *in*

Table 2. Knowledge Resources for Toxicological Modeling

resource type	name	description
AOP database	AOP-wiki (https://aopwiki.org, accessed January 2023)	One component of a larger OECD-sponsored AOP knowledge base (AOP-KB) effort, central web-based tool for disseminating and reviewing AOP knowledge; currently features more than 400 AOPs.
	AOP4EUpest[186]	A resource for annotated pesticides-biological events involved in AOPs.
gene (sets) annotation	Gene Ontology[178,179]	A comprehensive resource for computable knowledge of gene products comprised of gene ontology terms for many kinds of biological functions, involved pathways, and relationships between them.
	Molecular Signatures Database (MSigDB)[187,188]	Resource of ∼32 000 annotated gene sets for use with Gene Set Enrichment Analysis, including human and mouse collections.
pathway database	Kyoto Encyclopedia of Genes and Genomes (KEGG)[189,190]	KEGG has a collection of databases dealing with genomic information, biological pathway, diseases, drugs, and chemical substances. PATHWAY database contains pathway maps for the molecular systems in both normal and perturbed states.
	REACTOME[180]	A curated knowledge base of biological pathways in a hierarchical structure. It provides molecular details of biological processes as an ordered network of molecular transformations.
	Wikipathways[191]	A database of biological pathways collected and curated by the research community.
	Pathway commons[192]	Database that integrates data from public databases and contains over 5700 pathways and 2 million interactions.
toxicogenomics knowledge base	Comparative Toxicogenomics Database (CTD)[84]	Includes more than 30.5 million toxicogenomic connections relating chemical−gene/protein interaction, chemical−disease, and gene−disease relationships; gene ontology annotations; and pathways modules.
	Chemical Effects in Biological Systems (CEBS)[177,193]	Combines molecular expression data from transcriptomics, proteomics, metabonomics, and conventional toxicology with metabolic and toxicological pathway and gene regulatory network information relevant to environmental toxicology and human disease.
network analysis and visualization	Gephi[194]	An open-source visualization and exploration platform for all kinds of networks, complex systems, and graphs.
	Cytoscape[195]	Software platform for visualizing molecular interaction networks and biological pathways to integrate these networks with annotations, gene expression profiles, and other state data.

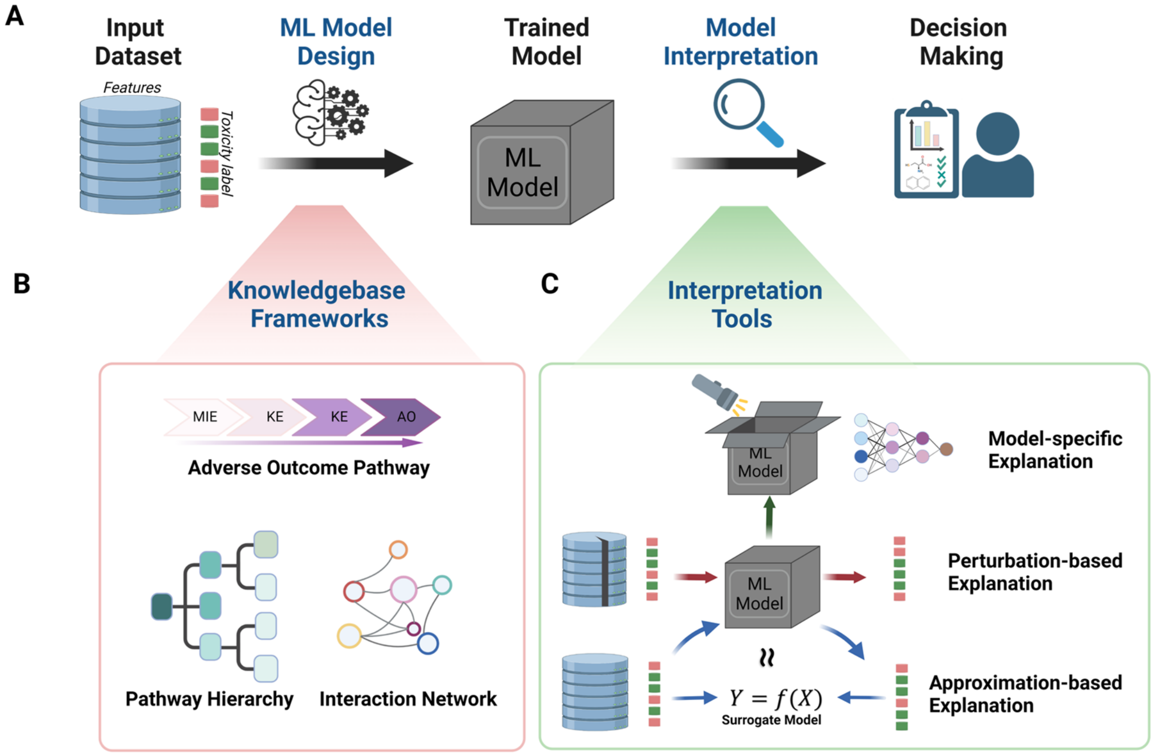


Figure 2. Strategies for the development of IML models for chemical toxicity. (A) Workflow for toxicity modeling using ML approaches where strategies can be applied before and after model training to improve intrinsic and posthoc interpretability, respectively. (B) Toxicological knowledgebase frameworks can be used to design models that are intrinsically interpretable. MIE, molecular initiating event; KE: key event; AO, adverse outcome. (C) With a trained model, interpretation tools can be applied to explain the trained model, examine important features, and support decision making.

vivo equivalent dose and prediction of toxicity potentials.[68] In parallel with the progress of various HTS projects, several data-sharing projects were also developed in the past decade (Table 1). For example, PubChem is a public repository for over 110 million chemicals and their associated bioactivities.[69,70] The tremendous amount of PubChem bioassay data that are updated daily constitutes a publicly accessible big data resource for compounds with a variety of target response information, which can also be used in toxicity ML and IML modeling studies.

2.3. Toxicogenomic Data. Cellular or organismal responses to chemical compounds are being measured at

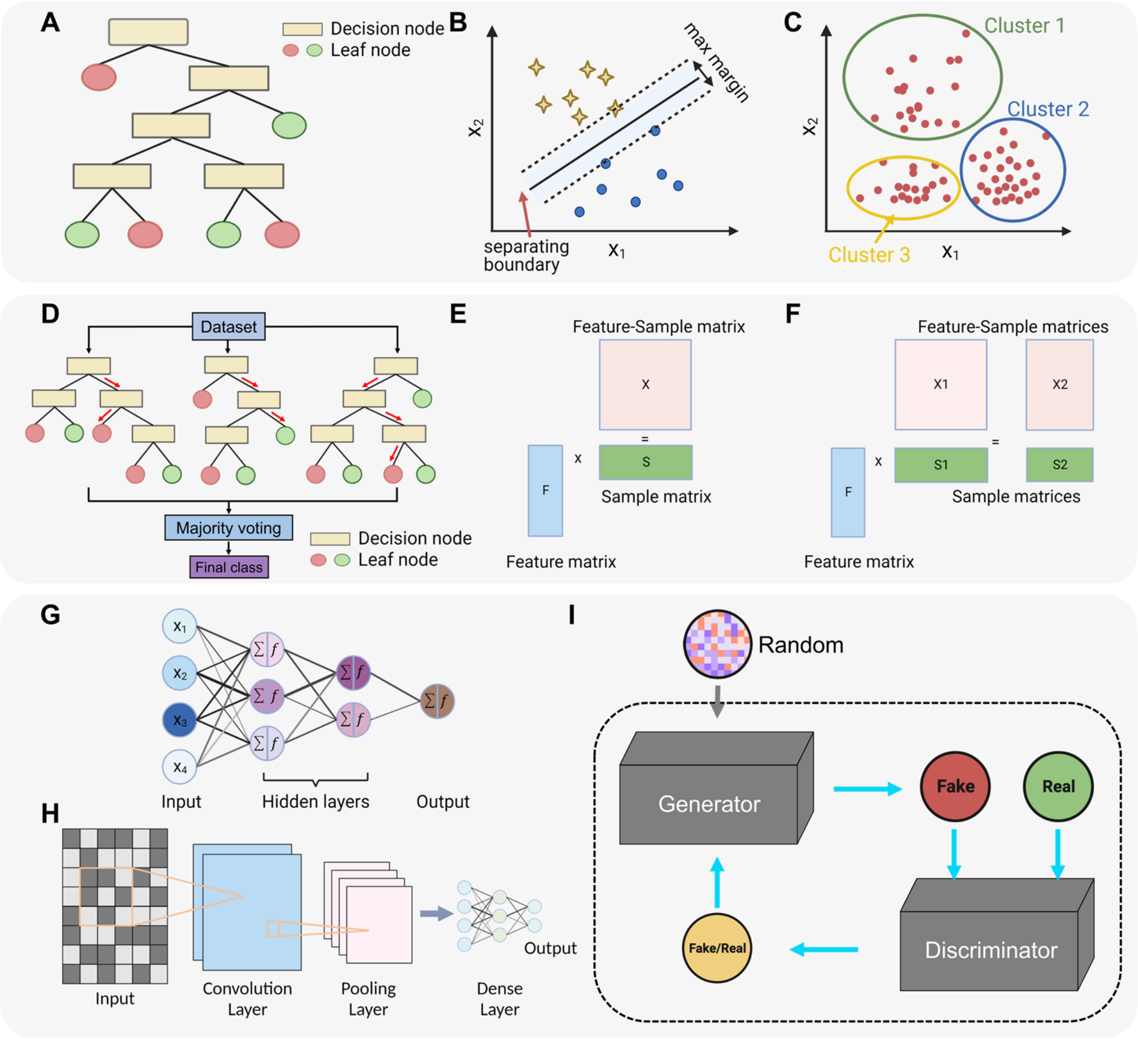


Figure 3. Examples of ML approaches: (A) decision tree, (B) support vector machine, (C) clustering, (D) random forest, (E) matrix factorization, (F) group factor analysis, (G) deep neural network, (H) convolutional neural network, (I) generative adversarial network.

different levels. Genome-wide transcriptomic data enables the assessment of alterations in gene expression profiles induced by chemicals. The rapid increase of genomic-sequence data and associated gene annotations (e.g., gene ontology) also accelerate the application of gene-expression modeling to understand the toxicity mechanism of toxicants.[71,72] Toxicogenomic data generated using these techniques provide extra valuable information in the chemical toxicity modeling process.[73−76] Table 1 also includes several toxicogenomic data repositories that store gene expression data of animals, human primary cells, and cell lines with/without exposure to drugs, industrial and environmental chemicals, etc. For example, Open TG-GATEs and DrugMatrix conducted short-term repeat-dose rat studies to obtain gene expression profiles coupling with histopathology measurements to enable a better understanding of chemical effects in rats.[77−79] Meanwhile, results from rat and human hepatocytes for the same set of chemicals allowed the identification of similarities and relationships between the *in vitro* and *in vivo* systems.[80] The L1000 project, as the next generation Connectivity Map, has developed a low-cost and high-throughput transcriptomic assay, which uses measurements of the 978 “landmark” genes to infer the expression levels of 81% of nonmeasured transcripts.[81] It generates transcriptomic profiles in multiple human-derived cell lines for around 20 000 chemicals. Toxicogenomic data generated from the above projects are deposited in the National Institutes of Health’s Gene Expression Omnibus (GEO) database. GEO is an international public repository that archives and freely distributes microarray, next generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community.[82,83] In addition to gene expression data, some databases also provide associations between chemicals, protein/gene targets, and disease that can aid in the mechanistic modeling of chemical-induced adverse outcomes. For example, Comparative Toxicogenomics Database (CTD) is a publicly available database of manually curated toxicogenomic information extracted from literature (Table 2).[84] It provides information for chemical−gene/protein interactions, chemical−disease associations, and gene−disease relationships, which can be integrated with pathway and functional data to facilitate the development of hypotheses about how environmental exposures influence human health.

https://doi.org/10.1021/acs.est.3c00653

3. ML APPROACHES AND MODEL-SPECIFIC INTERPRETATION METHODS FOR IML

The interpretability of ML models can be classified as intrinsic interpretability and posthoc interpretability[40,90,91] and can be achieved before and after model training (Figure 2A). Intrinsic interpretability is achieved by constructing self-explanatory models (e.g., using toxicological knowledge base frameworks) (Figure 2B), which incorporate interpretability directly into the model structures.[90] Understanding the inner logic of a ML algorithm is important for troubleshooting during model training. Posthoc interpretability is achieved after obtaining a trained model (Figure 2C). The goal of posthoc method is to understand the model predictions on the basis of the training data.[92] This section overviews examples of important ML algorithms and algorithm-specific techniques for interpreting the derived ML models.

3.1. Classic ML Approaches. ML includes three basic branches: supervised learning that aims to learn a mapping from features to labels (toxicity) on the basis of labeled training samples, unsupervised learning that aims to find patterns from unlabeled samples, and semisupervised learning that combines labeled samples with unlabeled samples during training.[93−95] Most classic supervised learning algorithms are well studied and interpretable for humans, such as linear regression, decision rule, and decision tree.[91] Linear regression models predict the target label as a weighted sum of feature data. The linearity of the learned relationship is easy to understand. For example, as a model of lipophilicity, logP is predicted using chemical structure or properties as regressors (e.g., functional groups, molecular volumes, and molecular weights).[96,97] Decision rules are also interpretable models that follow a general IF−THEN structure: IF the conditions are met, THEN the model makes a certain prediction. The conditions are built from interpretable features where pairs of conditions can be combined with AND/OR.[36,91] Decision trees are graphs to represent multiple true/false questions in a tree structure, where internal nodes represent tests on features to split the samples, edges represent the split decisions, and leaf nodes represent corresponding class labels (Figure 3A).[93] Predictions for new chemicals are made by following a decision path from the root to the leaf of a developed decision tree model. In a modeling study of oral toxicity, a decision tree was constructed with 33 questions on the basis of structure, biochemistry, and physiological chemistry information.[98] Each answer leads to another question and eventually ends with a final classification into one of three classes reflecting low, moderate, or serious toxicity. Support vector machine (SVM) is an algorithm that aims to find the hyperplane that best separates samples, such as chemicals, by their labels when the samples are placed in a high-dimensional feature space (Figure 3B). SVM can be trained to learn either linear or nonlinear relationships between features and labels. When dealing with nonlinear relationships, SVM projects the original data into a higher-dimension space where it can be separated by a linear hyperplane using the "kernel trick," thereby making it less interpretable.[99,100]

Most of the above algorithms are interpretable since their structures and inner working mechanisms are transparent.[90] Besides that, they are also interpreted on the modular level after model training, i.e., understanding the effects of training data and model parameters on predictions. For example, the weights of a linear model can be described as reflecting the strengths of relationships between features and target toxicity.[17] The positive weight of a feature means this feature contributes by increasing the model's output and vice versa.[36] In a decision tree model, the importance of a feature can be computed by going through the splits where the feature was used and measuring the increase of accuracy compared with the parent node.[90,91] By extracting the coefficient weights that define the hyperplane in a linear SVM model, toxicologists can interpret the features (molecular patterns) that were assigned a higher absolute weight as having stronger impacts on toxicity prediction.[101,102] The interpretation of nonlinear SVM is relatively complex and should be based on the specific kernel used for transforming the data.[101,103,104]

It should be noted that when the size and complexity of a ML model increase, the model will become less interpretable.[36] For example, a decision tree can be hard to interpret if it has a large width and depth. As an ensemble learning approach, random forest (RF) combines many decision trees to generate one prediction (Figure 3D) and is generally less interpretable.[105] To resolve this issue, approaches estimating feature importance and contributions for RF models have been developed and used in toxicity QSAR modeling, which can identify chemical substructures or features expected to be responsible for toxicity.[106−109] For example, Yu et al. proposed a tree-based RF feature importance and feature interaction network analysis to interpret the developed RF models for immune response and the lung burden of nanoparticles.[109] In this study, multiple-indicator feature importance analysis (e.g., predicted label change, node purity increase, etc.) was used to identify important features, and feature interaction networks were built to explore the interactions among multiple features. The modeling approach selection should depend on the complexity of the problem, specifically the relationship between input features and target toxicity labels. Linear regression or linear SVM models will not be applicable when the relationship between input features and toxicity is not straightforward. During the interpretation of feature importance, correlated features may cause issues where weights are split between them and feature importance is underestimated. A possible solution is to apply feature selection to remove redundant or irrelevant features, which can reduce the model's complexity and increase interpretability.[103,110−112]

3.2. DL Approaches. The advancement of computational infrastructure stimulated the applications of advanced ML algorithms to address the challenge of the explosive growth of toxicity data. DL is a part of the ML family based on artificial neural networks (ANN) with representation learning. DL algorithms are being applied widely in fields, including voice and image recognition, language processing, and clinic studies, because of their good predictive performance in modeling complex systems.[113−115] The structure of an ANN mimics the interlinked neurons in the brain, where a set of input nodes connects to a second set of nodes called the "hidden" layer and then eventually to an output layer.[116] A weight is associated with each of these connections between nodes, and there may be more than one hidden layer to construct a "deep" neural network (DNN) (Figure 3G). Other DL algorithms, such as CNN and adversarial learning, were designed for specific tasks. Inspired by the biological organization of the animal visual cortex, CNNs were constructed to learn spatial patterns or feature representation from input raw data (e.g., pixels from an image) (Figure 3H) which makes them ideal for image and speech applications.[117,118] In a generative adversarial network

(GAN), two DL models are trained during contesting with each other, the generative network generates synthetic data, and the discriminative network distinguishes synthetic data from true data distributions (Figure 3I), which is ideal for the generation of new data with similar statistics as the training set.[119] A recent study developed a GAN-based modeling framework (Tox-GAN) that learned from existing animal transcriptomic profiles to generate new transcriptomic profiles on the basis of chemical structures, doses, and treatment durations.[73] Tox-GAN can generate transcriptomic profiles without animal testing, which facilitates an understanding of toxicity mechanisms of new compounds and enhances the biomarker development in predictive toxicology.

Since the internal structures and underlying working mechanisms are less interpretable, DNN models are black-box in nature compared with classic ML models. To make predictions with a DNN model, input data pass through many layers of multiplications with the learned weights and through transformations by activation functions that can be nonlinear.[91] This process may involve millions of weight parameters depending on the architecture of the DNN, thereby making it difficult to understand the meanings of inner neurons and weights and how the predictions are made. To resolve this issue, methods for interpreting DNN models have been developed and can be classified as three major categories. Backpropagation-based methods calculate the gradient of an output with respect to the inputs to derive the contributions of features.[90,120−123] Connection weight-based methods track the magnitudes and directions of weights between neurons to identify individual and interacting effects of input variables on the outputs. It enables the estimations of feature importance when summing all connection weights.[124] Investigation of neuron representations looks at the hidden neuron representations to provide explanations. For example, in CNN, visualization of the inner neurons' output can show the encoded meanings of the original image.[90,125,126] In a DNN model for toxicity prediction using chemical structure data, Mayr et al. visualized the fragments represented in hidden neurons of different layers and found high correlations between neuron representations and toxicophores.[15] This study shows that new chemical knowledge can be found from the hidden neurons of DNN. Backpropagation-based methods have also been used in DL toxicity models to extract important substructures for toxicity prediction and further identify potential toxicophores.[127,128] Some software tools have been developed to facilitate DL interpretations. For example, Lucid (https://github.com/tensorflow/lucid, accessed January 2023) is an open-source toolbox containing methods for visualizing and interpreting neurons in ANN. iNNvestigate is a comprehensive library for implementing multiple interpretation methods for ANN models.[129]

3.3. Unsupervised Approaches. Besides the supervised learning algorithms discussed above, unsupervised techniques, such as clustering and matrix factorization algorithms, have been applied to study the feature variable relationships and reveal novel patterns. Clustering is the task of grouping a set of objects so that objects in the same cluster are more similar to each other than those in other clusters (Figure 3C). Clustering methods have been applied to cluster gene expression profiles,[130] biological assays,[131,132] and chemicals[67,133] in groups to help mechanistic interpretations and predictions of chemical toxicities. On the basis of the hypothesis that similar chemical share similar toxicological profiles, the read-across strategy was developed to predict toxicity for new compounds using similar compounds with known toxicity results, which is easy to interpret and implement.[134] Traditional read-across studies are only based on chemical structure similarity.[135−137] Software tools like ToxMatch and OECD QSAR Toolbox use chemical structure-based similarity to perform chemical grouping and read-across.[138,139] Similar to QSAR, only using structural information is error-prone when compounds with similar structures exhibit dissimilar toxicities.[55] To address this issue, a hybrid read-across strategy was developed by combining chemical structure similarity and biosimilarity, which is calculated on the basis of biological profiles (e.g., HTS assays, omics data).[55,140] The hybrid read-across could improve the discriminational power to distinguish compounds with similar chemical structures and could reveal the potential toxicity mechanisms by examining the bioprofiles.[73,131,141]

Matrix factorization (MF) is a collaborative filtering algorithm commonly used in recommendation systems.[142] It works by decomposing a high-dimensional matrix into a product of two low-dimensional matrices to capture key patterns in the data (Figure 3E). For example, high-dimensional biological data can be stored in a matrix with the feature values in rows and individual samples in columns. MF can be applied to characterize both features and samples by vectors of latent factors inferred from the original matrix.[72,143] In a study of toxicogenomic modeling, an extended MF technique, group factor analysis (GFA) (Figure 3F), was applied to model the relationships between the drug−gene matrix and drug-toxicity matrix. The identified shared components could capture cross-expression and toxicity relationships, which represent molecular mechanisms of toxicity.[72]

4. MODEL-AGNOSTIC INTERPRETATION METHODS FOR IML

Besides interpreting ML models on the basis of specific algorithms, some universal interpretation strategies can be applied after model training and treat a model as a black box without inspecting internal model parameters (i.e., model-agnostic) (Figure 2C).[24,36,90]

4.1. Perturbation-Based Explanation. Perturbation-based strategy modifies or removes parts of feature data to measure the corresponding change of the model output (Figure 2C). This method provides explanations in the form of feature contributions. Commonly used perturbation-based methods include sensitivity analysis and feature effect plots. Sensitivity analysis (SA) studies the correlation between the uncertainty in the model outputs and the uncertainty in the inputs.[144,145] SA can be performed with perturbation made to remove/permute one or more features at a time. One simple approach is altering one-feature-at-a-time (OAT) to see changes of the outputs.[146] However, the OAT approach does not fully explore the input space since it does not detect the interactions between input features. The variance-based method quantifies the contributions of input features to the variance of the model predictions by treating the input and output uncertainties as probability distributions.[145,147] As a measure of sensitivity, the total effect index gives the total variance in output Y caused by a feature X and its interactions with any other input features, which allows full explorations of the input space accounting for interactions. In a Bayesian network model to predict chemical modes of action (MoA) for aquatic toxicology, SA is applied to examine individual and multiple features' abilities to maximize MoA probabilities.[148] Only highly influential features were used to predict MoA, which reduced the model complexity and aided model interpretations. Feature effect plots is a powerful interpretation tool that visualizes the effects of features on the model's outputs. The partial dependence plot (PDP) visualizes the average partial dependence between the predicted label and one or two features while keeping all other features fixed.[91] It can show whether the relationship between the prediction and a feature is linear, monotonic, or more complex. For example, in modeling chemicals' P-glycoprotein (P-gp) transport, Svetnik et al. selected 49 descriptors with high feature importance, which were related to functional groups, to do PDP visualization.[149] Trends in PDPs can indicate whether a functional group tends to raise or lower P-gp activity and infer potential structure−activity relationships. An individual conditional expectation (ICE) plot extends PDP and visualizes the dependence of predictions on one or two features for each instance separately, thereby resulting in one curve per instance.[150−152] ICEs can reveal individual differences and identify subgroups and interactions between model inputs. Goldstein et al. demonstrated the use of ICE plots to analyze how different subjects respond to depression treatments in a clinical trial.[150] A black box model was built to predict treatment response scores using 37 personal features of the subjects and one binary treatment-type descriptor (cognitive therapy as 0 and paroxetine as 1). An ICE plot of two features, marital status and treatment type, showed that cognitive therapy is generally predicted to do better with married subjects, and paroxetine is predicted to do better with unmarried subjects. PDP and ICE plots are easy to interpret; however, they may miss important features since the partial dependence of the examined features (up to two) is computed on the basis of the assumption that they are not correlated with other features.[91] Several software tools have been developed to facilitate perturbation-based interpretations. For example, the *iml* R package implements many model-agnostic methods, including PDP, ICE, and feature importance.[153] The *ICEbox* and *pdp* R packages can be used for making ICE and PDP plots, respectively.[150,154] In addition, several software tools have been developed to perform sensitivity analysis, including the *sensitivity* R package (https://cran.r-project.org/web/packages/sensitivity/index.html, accessed March 2023), the *SAlib* python library,[155] and the *SAFE* (Sensitivity Analysis For Everybody) MATLAB package.[156]

4.2. Approximation-Based Explanation. Approximation-based methods involve learning an interpretable model (i.e., a surrogate model) to approximate the output of a black box model (Figure 2C). Training a surrogate model does not require information about the inner structure of the black box model, but access to the input data and model output is sufficient.[91] With an input data set and its corresponding output from a black box model, an interpretable surrogate model can be trained. The performance of a surrogate model can be measured in approximating predictions of the black box model and interpreting the surrogate model. Examples of surrogate models include linear models for characterizing linear relationships and decision trees and decision rules for characterizing nonlinear relationships.[157,158] A limitation of surrogate models is that complex black box models cannot be well approximated. To address this, an approach known as local interpretable model-agnostic explanations (LIME) was developed to focus on a small subset of instances.[159] LIME starts with instances of interest to generate a new data set consisting of perturbed features and the corresponding output of a black box model. Then, LIME trains an interpretable linear model, which is a good approximation of predictions for the instances of interest. The predictions of the black box model can be explained by examining the parameters of the linear model. For example, in a DL modeling study for *in vitro* toxicity predictions, Ramsundar et al. applied LIME to extract potential toxicophores responsible for relevant toxicity.[160] Sometimes, a linear surrogate model could lead to poor performance when the local relationship is nonlinear. Another local approximation-based approach, *anchors*, has been developed to characterize nonlinear relationships using decision rules.[161] Anchor explanations are effective in capturing nonlinear behaviors and can highlight the part of feature data that is sufficient for making a prediction. The *anchors* method was implemented in the python package *anchor*,[161] and integrated in *alibi*,[162] a Python library for ML model inspection and interpretation.

5. TOXICOLOGICAL KNOWLEDGE IN GUIDING THE DESIGN OF IML MODELS

Although posthoc interpretations are useful in understanding important features affecting toxicity predictions, they can be unreliable and misleading when the model is not appropriately designed. If self-explanatory models incorporate interpretability directly to the model structures, they can provide explanations to what the model computes.[27,90] For toxicity evaluations, an IML model can be developed to follow the toxicology knowledge (Figure 2B). In the past decade, systemic understanding of chemical toxicity in organisms using knowledge from an AOP framework and systems toxicology has become a trend.[163−165] The knowledge base frameworks represent a sequence and/or network of ordered events leading to adverse outcomes that show the interactions between toxicity-related components that can guide the design of intrinsic IML models in toxicity predictions.

5.1. The AOP Framework. An AOP is a structured representation of linked events between a molecular initiating event (MIE) (e.g., molecular interaction between a chemical and a receptor) and an adverse outcome in organisms (Figure 2B).[59,166] The MIE triggers a cascade of key events that occur at different biological levels relevant to adverse outcomes. The AOP-Wiki (aopwiki.org) is the primary repository for international AOP development efforts coordinated by the Organisation for Economic Co-operation and Development (OECD) (Table 2). Currently, the AOP-Wiki features more than 400 AOPs, which include those for various toxicity endpoints, such as acute inhalation toxicity,[167] reproductive and developmental toxicity,[168] and cholestatic and steatosis liver injury.[169] The AOP development efforts, together with publicly available toxicity big data, pave the way for computational AOP modeling that is more interpretable than traditional ML models. Moreover, the AOP structure organizations can be applied in designing IML models for efficient toxicity predictions. By integrating the chemical structure data and results of mechanism-based assays characterizing key events in AOP, the pathway models can systematically assess the potential of a compound to induce the target adverse outcome.[170]

An individual AOP may focus on a specific pathway, where mechanistically linked events proceed to a toxicity effect in a unidirectional and linear way.[163,171] In a mechanistic model for hepatotoxicity predictions, the toxicity potential of a chemical is assessed on the basis of whether it (1) possesses certain structural alerts and (2) activates the antioxidant response element (ARE) pathway, an oxidative stress-related key event.[172] This is a decision rule-based model, where chemicals that satisfy both two conditions are predicted as toxic, and chemicals that dissatisfy both two conditions are predicted as nontoxic. Chemicals that possess the identified structural alerts are suspected to be metabolized into reactive intermediates (i.e., MIE), which can trigger oxidative stress in the liver, thereby forming a plausible pathway that leads to hepatotoxicity. The limitation of this model is that it only can evaluate a small portion of hepatotoxicants since other toxicity mechanisms besides oxidative stress can also lead to hepatotoxicity. After the inclusion of additional assays representing key events in other AOPs, some false negative predictions could be corrected.

Several AOPs sharing at least one common component can form an AOP network.[163,171,173] They can focus on a single adverse outcome but describe different MIEs leading to this adverse outcome or share the same MIE but diverge to different AOPs.[173] Judson et al. reported a network model for chemical perturbations of the estrogen receptor (ER).[174] This model integrated three associated pathways, ER agonist, ER antagonist, and pseudoreceptor pathways, and utilized data from 18 HTS bioassays to identify ER agonists and antagonists. In a recent study, a knowledge base DNN model was developed to mimic the toxicity pathway for ER agonists using a virtual AOP framework.[175] In the DNN architecture, 57 HTS bioassays were organized among five network layers on the basis of the biological processes in the ER pathway—receptor binding as MIE, receptor dimerization, DNA binding, transcriptional activation, and cell proliferation as key events—and eventually led to the adverse outcome of *in vivo* rodent uterotrophic bioactivity. This model could efficiently and accurately evaluate rodent uterotrophic bioactivity for new compounds [area under the receiver operating characteristic (ROC) curve (AUC) = 0.864−0.927], which outperformed the QSAR model that only used chemical structure data as inputs (AUC = 0.594). Moreover, the model could virtually simulate the perturbations in the toxicity pathway for each predicted toxic compound. The design of DNN models to mimic AOP networks is promising for developing future interpretable models of complicated toxicity endpoints.

5.2. Systems Toxicology.

AOPs are usually constructed by literature compilations and focus on the states of a series of systems (cells, tissues, organs, organisms), whereas systems biology studies the molecular details (e.g., genes, proteins, metabolites) of these biological systems using -omics technologies.[163,176] As part of systems biology, systems toxicology describes the toxicological evaluation of biological systems, which involves perturbing systems by toxicants and stressors, monitoring molecular expressions and conventional toxicological parameters, and iteratively integrating response data.[71,177] Toxicology programs, such as ToxCast/Tox21 and L1000, have made progress in integrating data from diverse technologies and endpoints in different levels into systems biology approaches for toxicity evaluations.[62,81] Knowledge databases that reflect the functional characterization of components and interactions among diverse components provide informatic tools to support systems toxicology (Table 2). For example, the Gene Ontology database annotates each gene product regarding molecular function, biological process, and cell component, and relationships among these annotations form a loose hierarchical network (Figure 2B).[178,179] Reactome is a pathway database where relations of signaling and metabolic molecules are organized into a hierarchical network of biological pathways and processes (Figure 2B).[180] These structural knowledge bases enable the ML modeling of a biological system from the molecular level to larger pathways and cellular and even organism-level systems.

A recent development of a visible neural network (VNN) reported the integration of ontological and pathway hierarchy in the design of interpretable DNN models.[181−185] In VNNs, the connectivity of neurons in different layers is set to mirror the biological hierarchy. Genes or proteins as inputs only connect to specific neurons representing their associated pathways, and these pathway neurons subsequently connect to their parent pathways, thereby making it a sparse DNN with reduced complexity and intrinsic interpretability. Kuenzi et al. developed a VNN model named DrugCell to simulate the response of human cancer cells to therapeutic compounds.[182] DrugCell was designed with two parts: a VNN modeling the hierarchical organization of Gene Ontology terms and a conventional ANN embedding the chemical fingerprints. DrugCell could correctly predict drug responses (Spearman correlation rho = 0.8 between predicted and observed response value), which significantly outperformed the elastic net regression model ($p < 0.0001$). Furthermore, DrugCell could provide insights into the underlying mechanisms of action by inspecting the simulated pathway neuron states. Elmarakeby et al. developed a VNN model named P-NET that integrates Reactome hierarchical knowledge to predict cancer state on the basis of patients' genomic profiles.[184] The trained P-NET model outperformed classic ML models, including SVM, logistic regression, and decision trees, with AUC = 0.93 and accuracy = 0.83. Additionally, P-NET demonstrated significantly better performance than the traditional dense DNN model in sample sizes up to 500 ($p < 0.05$). In a recent study, Hao et al. reported a VNN model named DTox (deep learning for toxicology) for predict a chemical's outcomes in 15 toxicity assays.[183] DTox connects target protein profiles (input features) to toxicity assay outcomes (outputs) by hidden neurons mapping to the Reactome pathways. DTox can achieve the same level of performance as classic ML approaches and further explain toxicity mechanisms by identifying VNN paths. Since the gene expression change of the components constituting a pathway may reflect whether the pathway is perturbed, the identified VNN paths were further validated by differential expression analysis using LINCS transcriptomic profiles.

6. IMPLICATIONS AND PERSPECTIVES

In toxicology, ML-based computational modeling is a promising alternative method to traditional animal models for predicting chemical toxicity potentials. In the current big data era, chemical toxicity data continue to grow at a rapid pace, and advanced ML and DL approaches are urgently needed to deal with these data. Interpretability is critical for the application of ML in risk assessments of chemicals that may impact human and environmental health. In the above sections, we have presented strategies for applying IML in computational toxicology, including the use of interpretable feature data, interpretation methods, and the development of intrinsic IML models using knowledgebase frameworks in toxicology.

Data standardization and curation are critical in computational modeling approaches, where care should be taken to avoid introducing technical artifacts and to ensure the quality of modeling sets, as well as the resulting model performance and interpretation.[196] Many modeling studies use well-established molecular representations as features, such as properties, binary fingerprints, and geometrical descriptors, which can capture chemical and structural features defined in advance.[197,198] The generation of these molecular representations is standardized, and various structure curation protocols are available to facilitate the chemical descriptor generations.[199,200] However, as described above, toxicogenomic and assay data, which can capture intricate biological responses for chemicals of interest, are much more diverse, heterogeneous, and unstandardized than chemical structures. Experimental conditions and protocols for generating these data can vary widely among laboratories, thereby leading to poor data quality in some data resources that may impact model performance and interpretability.[7] Similarly, assay data from various studies and HTS programs may exhibit different data structures (e.g., classifications, dose/concentration dependent curves, or even raw data) and may have inconsistent results for the same chemicals.[8,201] When collecting training data from multiple resources, curation workflows, such as those employed by Integrated Chemical Environment and ChEMBL, should be implemented to ensure data quality and integrity.[86,202,203]

The interpretation strategy should be tailed to what needs to be learned from the model.[196,204] Traditionally, understanding which structure features of a toxicant contribute to its toxicity is useful to toxicologists for decision making and medicinal chemists to modify the molecule.[109,205] Mechanistic explanations of the model predictions are crucial for high-stakes decision-making, such as determining whether a chemical is safe for humans and the environment. However, there are still many challenges to be resolved, and further efforts are needed to advance IML in this field. For example, new mechanistic IML models are trained on heterogeneous data (e.g., chemical structure, gene expression, and bioactivities), which increases the complexity of modeling tasks and makes it challenging to identify critical features and explain underlying mechanisms. Another issue caused by such heterogeneous data is the existence of missing values in the feature profiles for target compounds, which is a common issue in big data modeling.[5,206] Methods to impute missing values (e.g., read-across) may introduce uncertainties in the training data and the following modeling procedure. Therefore, the development of novel and interpretable representations of chemicals will be critical in future research.[204] Intrinsic interpretable models that incorporate toxicological knowledge frameworks can overcome the challenges posed by big data by providing both mechanism explanations and accurate predictions.[204] In QSAR modeling, the chemical descriptor space is normally used to define the applicability domain of the model to assess whether the prediction for a target chemical is reliable or not.[207−209] However, it is difficult to define the applicability domain for mechanistic models, especially those using DL and heterogeneous data. In terms of IML explanations, a challenge lies in how to measure the levels of interpretability, compare the interpretability of different IML models, or determine the faithfulness of different interpretation methods to the same model.[40,90] There are no universal criteria for selecting ML approaches for toxicological modeling, nor is there a clear choice for the optimal interpretation methods. Confidence in the interpretation results will be enhanced when multiple approaches yield consistent conclusions.[196] As many interpretation strategies being developed for IML in the toxicology community, the use of these strategies can require significant computational knowledge for toxicologists. Further development and improvements of user-friendly software platforms can facilitate the design, validation, and acceptance of IML and its associated explanations.

AUTHOR INFORMATION

Corresponding Author

Hao Zhu − *Department of Chemistry and Biochemistry, Rowan University, Glassboro, New Jersey 08028, United States;* orcid.org/0000-0002-3559-6129; Phone: (856) 256-4500; Email: zhuh@rowan.edu

Authors

Xuelian Jia − *Department of Chemistry and Biochemistry, Rowan University, Glassboro, New Jersey 08028, United States;* orcid.org/0000-0001-9901-9104

Tong Wang − *Department of Chemistry and Biochemistry, Rowan University, Glassboro, New Jersey 08028, United States;* orcid.org/0000-0002-6719-7368

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.est.3c00653

Author Contributions

Xuelian Jia: Writing - Original Draft, Writing - Review & Editing, Visualization. Tong Wang: Writing - Review & Editing, Visualization. Hao Zhu: Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was partially supported by the National Institute of Environmental Health Sciences (Grants R01ES031080 and R35ES031709), National Institute of General Medical Sciences (Grant R01GM148743), and National Science Foundation (Grant 2211489).

REFERENCES

(1) Balbus, J. M. Ushering in the new toxicology: toxicogenomics and the public interest. *Environ. Health Perspect.* **2005**, *113* (7), 818−822.

(2) Nigsch, F.; Macaluso, N. J.; Mitchell, J. B.; Zmuidinavicius, D. Computational toxicology: an overview of the sources of data and of modelling methods. *Expert Opin. Drug Metab. Toxicol.* **2009**, *5* (1), 1−14.

(3) Richard, A. M. Future of toxicology−predictive toxicology: An expanded view of "chemical toxicity". *Chem. Res. Toxicol.* **2006**, *19* (10), 1257−1262.

(4) Muster, W.; Breidenbach, A.; Fischer, H.; Kirchner, S.; Muller, L.; Pahler, A. Computational toxicology in drug development. *Drug Discovery Today* **2008**, *13* (7−8), 303−310.

(5) Zhu, H. Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annu. Rev. Pharmacol. Toxicol.* **2020**, *60*, 573−589.

(6) Judson, R.; Richard, A.; Dix, D. J.; Houck, K.; Martin, M.; Kavlock, R.; Dellarco, V.; Henry, T.; Holderman, T.; Sayre, P.; Tan, S.; Carpenter, T.; Smith, E. The toxicity data landscape for

environmental chemicals. *Environ. Health Perspect.* **2009**, *117* (5), 685−695.

(7) Wu, X.; Zhou, Q.; Mu, L.; Hu, X. Machine learning in the identification, prediction and exploration of environmental toxicology: Challenges and perspectives. *J. Hazard. Mater.* **2022**, *438*, 129487.

(8) Ciallella, H. L.; Zhu, H. Advancing Computational Toxicology in the Big Data Era by Artificial Intelligence: Data-Driven and Mechanism-Driven Modeling for Chemical Toxicity. *Chem. Res. Toxicol.* **2019**, *32* (4), 536−547.

(9) Gibb, S. Toxicity testing in the 21st century: a vision and a strategy. *Reprod. Toxicol.* **2008**, *25* (1), 136−138.

(10) Krewski, D.; Acosta, D., Jr.; Andersen, M.; Anderson, H.; Bailar, J. C., 3rd; Boekelheide, K.; Brent, R.; Charnley, G.; Cheung, V. G.; Green, S., Jr.; Kelsey, K. T.; Kerkvliet, N. I.; Li, A. A.; McCray, L.; Meyer, O.; Patterson, R. D.; Pennie, W.; Scala, R. A.; Solomon, G. M.; Stephens, M.; Yager, J.; Zeise, L. Toxicity testing in the 21st century: a vision and a strategy. *J. Toxicol. Environ. Health B* **2010**, *13* (2−4), 51−138.

(11) *Frank R. Lautenberg Chemical Safety for the 21st Century Act § 15 U.S.C. 2601*, 2016; pp 114−182.

(12) Bassan, A.; Alves, V. M.; Amberg, A.; Anger, L. T.; Auerbach, S.; Beilke, L.; Bender, A.; Cronin, M. T. D.; Cross, K. P.; Hsieh, J. H.; Greene, N.; Kemper, R.; Kim, M. T.; Mumtaz, M.; Noeske, T.; Pavan, M.; Pletz, J.; Russo, D. P.; Sabnis, Y.; Schaefer, M.; Szabo, D. T.; Valentin, J. P.; Wichard, J.; Williams, D.; Woolley, D.; Zwickl, C.; Myatt, G. J. In silico approaches in organ toxicity hazard assessment: current status and future needs in predicting liver toxicity. *Comput. Toxicol.* **2021**, *20*, 100187.

(13) Xu, Y.; Pei, J.; Lai, L. Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction. *J. Chem. Inf. Model.* **2017**, *57* (11), 2672−2685.

(14) Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L. Deep Learning for Drug-Induced Liver Injury. *J. Chem. Inf. Model.* **2015**, *55* (10), 2085−2093.

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

(16) Mayr, A.; Klambauer, G.; Unterthiner, T.; Steijaert, M.; Wegner, J. K.; Ceulemans, H.; Clevert, D.-A.; Hochreiter, S. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chem. Sci.* **2018**, *9* (24), 5441−5451.

(17) Lipton, Z. C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* **2018**, *16* (3), 31−57.

(18) Koenigstein, N.; Dror, G.; Koren, Y. Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy. *Proceedings of the fifth ACM conference on Recommender systems* **2011**, 165−172.

(19) Linden, G.; Smith, B.; York, J. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet. Comput.* **2003**, *7* (1), 76−80.

(20) Dattner, B.; Chamorro-Premuzic, T.; Buchband, R.; Schettler, L. The legal and ethical implications of using AI in hiring. *Harvard Bus. Rev.* **2019**, *25*, 1−7.

(21) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discovery* **2019**, *18* (6), 463−477.

(22) Erickson, B. J.; Korfiatis, P.; Akkus, Z.; Kline, T. L. Machine learning for medical imaging. *Radiographics* **2017**, *37* (2), 505−515.

(23) Hosny, A.; Aerts, H. J. Artificial intelligence for global health. *Science* **2019**, *366* (6468), 955−956.

(24) Ribeiro, M. T.; Singh, S.; Guestrin, C. Model-agnostic interpretability of machine learning. *arXiv*, June 16, 2016, 1606.05386, ver. *1*. DOI: 10.48550/arXiv.1606.05386.

(25) Angwin, J.; Larson, J.; Mattu, S.; Kirchner, L. Machine bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed December 01, 2022).

(26) Obermeyer, Z.; Powers, B.; Vogeli, C.; Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **2019**, *366* (6464), 447−453.

(27) Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1* (5), 206−215.

(28) Lo Piano, S. Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanit. Soc. Sci. Commun.* **2020**, *7* (1), 1−7.

(29) Rudin, C.; Chen, C.; Chen, Z.; Huang, H.; Semenova, L.; Zhong, C. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Stat. Surv.* **2022**, *16*, 1−85.

(30) Thiebes, S.; Lins, S.; Sunyaev, A. Trustworthy artificial intelligence. *Electron. Mark.* **2021**, *31* (2), 447−464.

(31) Lundberg, S. M.; Nair, B.; Vavilala, M. S.; Horibe, M.; Eisses, M. J.; Adams, T.; Liston, D. E.; Low, D. K.; Newman, S. F.; Kim, J.; Lee, S. I. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2018**, *2* (10), 749−760.

(32) ElShawi, R.; Sherif, Y.; Al-Mallah, M.; Sakr, S. Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Comput. Intell.* **2021**, *37* (4), 1633−1650.

(33) Panigutti, C.; Perotti, A.; Pedreschi, D. Doctor XAI: an ontology-based approach to black-box sequential data classification explanations. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*; Association for Computing Machinery: New York, 2020; pp 629−639. DOI: 10.1145/3351095.3372855.

(34) Zafar, M. R.; Khan, N. M. DLIME: A deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems. *arXiv*, June 24, 2019, 1906.10263, ver. *1*. DOI: 10.48550/arXiv.1906.10263.

(35) Knapič, S.; Malhi, A.; Saluja, R.; Främling, K. Explainable artificial intelligence for human decision support system in the medical domain. *Mach. Learn. Knowl. Extr.* **2021**, *3* (3), 740−770.

(36) Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A survey of methods for explaining black box models. *ACM Comput. Surv.* **2019**, *51* (5), 1−42.

(37) Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **2019**, *267*, 1−38.

(38) Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv*, February 28, 2017, 1702.08608, ver. *1*. DOI: 10.48550/arXiv.1702.08608.

(39) Gilpin, L. H.; Bau, D.; Yuan, B. Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*; IEEE, **2018**; pp 80−89. .

(40) Murdoch, W. J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (44), 22071−22080.

(41) Jordan, M. I.; Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349* (6245), 255−260.

(42) Zhou, Z.-H. Introduction. In *Machine learning*; Springer Singapore: 2021; pp 1−24.

(43) Golbraikh, A.; Wang, X. S.; Zhu, H.; Tropsha, A. Predictive QSAR modeling: methods and applications in drug discovery and chemical risk assessment. In *Handbook of Computational Chemistry*; Springer Dordrecht: Netherlands, 2016; pp 1−48.

(44) Gini, G., QSAR methods. In *In silico methods for predicting drug toxicity*; Springer: New York, 2016; pp 1−20.

(45) Khan, A. U. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discovery Today* **2016**, *21* (8), 1291−1302.

(46) Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. InChI-the worldwide chemical structure identifier standard. *J. Cheminformatics* **2013**, *5* (1), 1−9.

(47) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comp. Sci.* **1988**, *28* (1), 31−36.

(48) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.* **2016**, *30* (8), 595−608.

(49) Ryu, S.; Lim, J.; Hong, S. H.; Kim, W. Y. Deeply learning molecular structure-property relationships using attention-and gate-augmented graph convolutional network. *arXiv*, May 28, 2018, 1805.10988, ver. *1*. DOI: 10.48550/arXiv.1805.10988.

(50) Goyal, P.; Ferrara, E. Graph embedding techniques, applications, and performance: A survey. *Knowl. Based Syst.* **2018**, *151*, 78−94.

(51) Russo, D. P.; Yan, X.; Shende, S.; Huang, H.; Yan, B.; Zhu, H. Virtual Molecular Projections and Convolutional Neural Networks for the End-to-End Modeling of Nanoparticle Activities and Properties. *Anal. Chem.* **2020**, *92* (20), 13971−13979.

(52) Yan, X.; Sedykh, A.; Wang, W.; Zhao, X.; Yan, B.; Zhu, H. In silico profiling nanoparticles: predictive nanomodeling using universal nanodescriptors and various machine learning approaches. *Nanoscale* **2019**, *11* (17), 8352−8362.

(53) Wang, T.; Russo, D. P.; Bitounis, D.; Demokritou, P.; Jia, X.; Huang, H.; Zhu, H. Integrating structure annotation and machine learning approaches to develop graphene toxicity models. *Carbon* **2023**, *204*, 484−494.

(54) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; et al. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57* (12), 4977−5010.

(55) Zhu, H.; Bouhifd, M.; Kleinstreuer, N.; Kroese, E. D.; Liu, Z.; Luechtefeld, T.; Pamies, D.; Shen, J.; Strauss, V.; Wu, S. t4 report: Supporting read-across using biological data. *ALTEX* **2016**, *33* (2), 167.

(56) Hall, L. H.; Kier, L. B. The molecular connectivity chi indexes and kappa shape indexes in structure-property modeling. *Reviews in computational chemistry* **2007**, 367−422.

(57) Zhu, H.; Zhang, J.; Kim, M. T.; Boison, A.; Sedykh, A.; Moran, K. Big data in chemical toxicity research: the use of high-throughput screening assays to identify potential toxicants. *Chem. Res. Toxicol.* **2014**, *27* (10), 1643−1651.

(58) Inglese, J.; Auld, D. S.; Jadhav, A.; Johnson, R. L.; Simeonov, A.; Yasgar, A.; Zheng, W.; Austin, C. P. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (31), 11473−11478.

(59) 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.; Serrano, J. A.; Tietge, J. E.; Villeneuve, D. L. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* **2010**, *29* (3), 730−741.

(60) Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* **2007**, *95* (1), 5−12.

(61) Attene-Ramos, M. S.; Miller, N.; Huang, R.; Michael, S.; Itkin, M.; Kavlock, R. J.; Austin, C. P.; Shinn, P.; Simeonov, A.; Tice, R. R.; et al. The Tox21 robotic platform for the assessment of environmental chemicals−from vision to reality. *Drug Discovery Today* **2013**, *18* (15−16), 716−723.

(62) Tice, R. R.; Austin, C. P.; Kavlock, R. J.; Bucher, J. R. Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect.* **2013**, *121* (7), 756−765.

(63) Judson, R. S.; Houck, K. A.; Kavlock, R. J.; Knudsen, T. B.; Martin, M. T.; Mortensen, H. M.; Reif, D. M.; Rotroff, D. M.; Shah, I.; Richard, A. M.; Dix, D. J. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environ. Health Perspect.* **2010**, *118* (4), 485−492.

(64) Kavlock, R.; Chandler, K.; Houck, K.; Hunter, S.; Judson, R.; Kleinstreuer, N.; Knudsen, T.; Martin, M.; Padilla, S.; Reif, D.; Richard, A.; Rotroff, D.; Sipes, N.; Dix, D. Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* **2012**, *25* (7), 1287−1302.

(65) Sedykh, A.; Zhu, H.; Tang, H.; Zhang, L.; Richard, A.; Rusyn, I.; Tropsha, A. Use of in vitro HTS-derived concentration−response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. *Environ. Health Perspect.* **2011**, *119* (3), 364−370.

(66) Wang, W.; Kim, M. T.; Sedykh, A.; Zhu, H. Developing enhanced blood−brain barrier permeability models: integrating external bio-assay data in QSAR modeling. *Pharm. Res.* **2015**, *32* (9), 3055−3065.

(67) Huang, R.; Xia, M.; Sakamuru, S.; Zhao, J.; Shahane, S. A.; Attene-Ramos, M.; Zhao, T.; Austin, C. P.; Simeonov, A. Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. *Nat. Commun.* **2016**, *7* (1), 1−10.

(68) Wetmore, B. A.; Wambaugh, J. F.; Ferguson, S. S.; Sochaski, M. A.; Rotroff, D. M.; Freeman, K.; Clewell, H. J., 3rd; Dix, D. J.; Andersen, M. E.; Houck, K. A.; Allen, B.; Judson, R. S.; Singh, R.; Kavlock, R. J.; Richard, A. M.; Thomas, R. S. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicol. Sci.* **2012**, *125* (1), 157−174.

(69) Wang, Y.; Bryant, S. H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B. A.; Thiessen, P. A.; He, S.; Zhang, J. PubChem BioAssay: 2017 update. *Nucleic Acids Res.* **2017**, *45* (D1), D955−D963.

(70) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44* (D1), D1202−D1213.

(71) Waters, M. D.; Fostel, J. M. Toxicogenomics and systems toxicology: aims and prospects. *Nat. Rev. Genet.* **2004**, *5* (12), 936−948.

(72) Khan, S. A.; Aittokallio, T.; Scherer, A.; Grafström, R.; Kohonen, P. Matrix and Tensor Factorization Methods for Toxicogenomic Modeling and Prediction. In *Advances in computational toxicology*; Challenges and Advances in Computational Chemistry and Physics, Vol. *30*; Springer Nature: Cham, Switzerland, **2019**; pp 57−74.

(73) Chen, X.; Roberts, R.; Tong, W.; Liu, Z. Tox-GAN: An Artificial Intelligence Approach Alternative to Animal Studies-A Case Study With Toxicogenomics. *Toxicol. Sci.* **2022**, *186* (2), 242−259.

(74) Liu, Z.; Fang, H.; Borlak, J.; Roberts, R.; Tong, W. In vitro to in vivo extrapolation for drug-induced liver injury using a pair ranking method. *ALTEX* **2017**, *34* (3), 399−408.

(75) Uehara, T.; Minowa, Y.; Morikawa, Y.; Kondo, C.; Maruyama, T.; Kato, I.; Nakatsu, N.; Igarashi, Y.; Ono, A.; Hayashi, H.; et al. Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. *Toxicol. Appl. Pharmacol.* **2011**, *255* (3), 297−306.

(76) Ellinger-Ziegelbauer, H.; Gmuender, H.; Bandenburg, A.; Ahr, H. J. Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat. Res. - Fundam. Mol. M.* **2008**, *637* (1−2), 23−39.

(77) Igarashi, Y.; Nakatsu, N.; Yamashita, T.; Ono, A.; Ohno, Y.; Urushidani, T.; Yamada, H. Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* **2015**, *43* (D1), D921−D927.

(78) 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.; Breckenridge, N.; Dunlea, S.; Eynon, B. P.; Furness, L. M.; Ferng, J.; Fielden, M. R.; Fujimoto, S. Y.; Gong, L.; Hu, C.; Idury, R.; Judo, M. S.; Kolaja, K. L.; Lee, M. D.; McSorley, C.; Minor, J. M.; Nair, R. V.; Natsoulis, G.; Nguyen, P.; Nicholson, S. M.; Pham, H.; Roter, A. H.; Sun, D.; Tan, S.; Thode, S.; Tolley, A. M.; Vladimirova, A.; Yang, J.; Zhou, Z.; Jarnagin, K. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* **2005**, *119* (3), 219−244.

(79) Ganter, B.; Snyder, R. D.; Halbert, D. N.; Lee, M. D. Toxicogenomics in drug discovery and development: mechanistic

analysis of compound/class-dependent effects using the DrugMatrix database. *Pharmacogenomics* **2006**, *7* (7), 1025−1044.

(80) Chen, M.; Zhang, M.; Borlak, J.; Tong, W. A decade of toxicogenomic research and its contribution to toxicological science. *Toxicol. Sci.* **2012**, *130* (2), 217−228.

(81) 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.; Lahr, D. L.; Hirschman, J. E.; Liu, Z.; Donahue, M.; Julian, B.; Khan, M.; Wadden, D.; Smith, I. C.; Lam, D.; Liberzon, A.; Toder, C.; Bagul, M.; Orzechowski, M.; Enache, O. M.; Piccioni, F.; Johnson, S. A.; Lyons, N. J.; Berger, A. H.; Shamji, A. F.; Brooks, A. N.; Vrcic, A.; Flynn, C.; Rosains, J.; Takeda, D. Y.; Hu, R.; Davison, D.; Lamb, J.; Ardlie, K.; Hogstrom, L.; Greenside, P.; Gray, N. S.; Clemons, P. A.; Silver, S.; Wu, X.; Zhao, W. N.; Read-Button, W.; Wu, X.; Haggarty, S. J.; Ronco, L. V.; Boehm, J. S.; Schreiber, S. L.; Doench, J. G.; Bittker, J. A.; Root, D. E.; Wong, B.; Golub, T. R. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171* (6), 1437−1452.

(82) Edgar, R.; Domrachev, M.; Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30* (1), 207−210.

(83) 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.; Yefanov, A.; Lee, H.; Zhang, N.; Robertson, C. L.; Serova, N.; Davis, S.; Soboleva, A. NCBI GEO: archive for functional genomics data sets−update. *Nucleic Acids Res.* **2012**, *41* (D1), D991−D995.

(84) Davis, A. P.; Wiegers, T. C.; Johnson, R. J.; Sciaky, D.; Wiegers, J.; Mattingly, C. J. Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Res.* **2023**, *51* (D1), D1257−D1262.

(85) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Felix, E.; Magarinos, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Maranon, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47* (D1), D930−D940.

(86) Bell, S. M.; Phillips, J.; Sedykh, A.; Tandon, A.; Sprankle, C.; Morefield, S. Q.; Shapiro, A.; Allen, D.; Shah, R.; Maull, E. A.; et al. An integrated chemical environment to support 21st-century toxicology. *Environ. Health Perspect.* **2017**, *125* (5), 054501.

(87) Bell, S.; Abedini, J.; Ceger, P.; Chang, X.; Cook, B.; Karmaus, A. L.; Lea, I.; Mansouri, K.; Phillips, J.; McAfee, E.; et al. An integrated chemical environment with tools for chemical safety testing. *Toxicol. In Vitro* **2020**, *67*, 104916.

(88) Abedini, J.; Cook, B.; Bell, S.; Chang, X.; Choksi, N.; Daniel, A. B.; Hines, D.; Karmaus, A. L.; Mansouri, K.; McAfee, E.; et al. Application of new approach methodologies: ICE tools to support chemical evaluations. *Comput. Toxicol.* **2021**, *20*, 100184.

(89) Heng, T. S.; Painter, M. W.; Elpek, K.; Lukacs-Kornek, V.; Mauermann, N.; Turley, S. J.; Koller, D.; Kim, F. S.; Wagers, A. J.; Asinovski, N.; et al. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **2008**, *9* (10), 1091−1094.

(90) Du, M.; Liu, N.; Hu, X. Techniques for interpretable machine learning. *Commun. ACM* **2019**, *63* (1), 68−77.

(91) Molnar, C. *Interpretable machine learning*. Lulu.com, 2020; p 1−329.

(92) Montavon, G.; Samek, W.; Müller, K.-R. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* **2018**, *73*, 1−15.

(93) Hastie, T.; Tibshirani, R.; Friedman, J. H.; Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*, Vol. 2. Springer: New York, 2009; pp 1−520.

(94) Murphy, K. P. *Machine learning: a probabilistic perspective*; MIT press: Cambridge, MA, 2012.

(95) Van Engelen, J. E.; Hoos, H. H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109* (2), 373−440.

(96) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. *J. Pharm. Sci.* **2009**, *98* (3), 861−893.

(97) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. Computer automated log P calculations based on an extended group contribution approach. *J. Chem. Inf. Comp. Sci.* **1994**, *34* (4), 752−781.

(98) Cramer, G. M.; Ford, R. A.; Hall, R. L. Estimation of toxic hazard−a decision tree approach. *Food Cosmet. Toxicol.* **1976**, *16* (3), 255−276.

(99) Scholkopf, B.; Smola, A. J. *Learning with kernels: support vector machines, regularization, optimization, and beyond*; MIT press: Cambridge, MA, 2018; pp 187−517.

(100) Cristianini, N.; Shawe-Taylor, J. *An introduction to support vector machines and other kernel-based learning methods*; Cambridge University Press: Cambridge, England, 2000; pp 97−135.

(101) Yang, P.; Henle, E. A.; Fern, X. Z.; Simon, C. M. Classifying the toxicity of pesticides to honey bees via support vector machines with random walk graph kernels. *J. Chem. Phys.* **2022**, *157* (3), 034102.

(102) Rosenbaum, L.; Hinselmann, G.; Jahn, A.; Zell, A. Interpreting linear support vector machine models with heat map molecule coloring. *J. Cheminformatics* **2011**, *3* (1), 11.

(103) Chen, Z.; Li, J.; Wei, L. A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue. *Artif. Intell. Med.* **2007**, *41* (2), 161−175.

(104) Barakat, N.; Bradley, A. P. Rule extraction from support vector machines: a review. *Neurocomputing* **2010**, *74* (1−3), 178−190.

(105) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45* (1), 5−32.

(106) Kuz'min, V. E.; Polishchuk, P. G.; Artemenko, A. G.; Andronati, S. A. Interpretation of QSAR Models Based on Random Forest Methods. *Mol. Inform* **2011**, *30* (6−7), 593−603.

(107) Marchese Robinson, R. L.; Palczewska, A.; Palczewski, J.; Kidley, N. Comparison of the Predictive Performance and Interpretability of Random Forest and Linear Models on Benchmark Data Sets. *J. Chem. Inf. Model.* **2017**, *57* (8), 1773−1792.

(108) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comp. Sci.* **2003**, *43* (6), 1947−1958.

(109) Yu, F.; Wei, C.; Deng, P.; Peng, T.; Hu, X. Deep exploration of random forest model boosts the interpretability of machine learning studies of complicated immune responses and lung burden of nanoparticles. *Sci. Adv.* **2021**, *7* (22), No. eabf4130.

(110) Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R. P.; Tang, J.; Liu, H. Feature selection: A data perspective. *ACM Comput. Surv.* **2018**, *50* (6), 1−45.

(111) Blum, A. L.; Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **1997**, *97* (1−2), 245−271.

(112) Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3* (Mar), 1157−1182.

(113) Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Netw.* **2015**, *61*, 85−117.

(114) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521* (7553), 436−444.

(115) Ching, T.; Himmelstein, D. S.; Beaulieu-Jones, B. K.; Kalinin, A. A.; Do, B. T.; Way, G. P.; Ferrero, E.; Agapow, P. M.; Zietz, M.; Hoffman, M. M.; Xie, W.; Rosen, G. L.; Lengerich, B. J.; Israeli, J.; Lanchantin, J.; Woloszynek, S.; Carpenter, A. E.; Shrikumar, A.; Xu, J.; Cofer, E. M.; Lavender, C. A.; Turaga, S. C.; Alexandari, A. M.; Lu, Z.; Harris, D. J.; DeCaprio, D.; Qi, Y.; Kundaje, A.; Peng, Y.; Wiley, L. K.; Segler, M. H. S.; Boca, S. M.; Swamidass, S. J.; Huang, A.; Gitter, A.; Greene, C. S. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **2018**, *15* (141), 20170387.

(116) Burrell, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data Soc.* **2016**, *3* (1), 205395171562251.

(117) LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. In *The handbook of brain theory and neural networks*; MIT Press: Cambridge, MA, 1998; pp 255−258.

(118) Fukushima, K.; Miyake, S. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **1980**, *36* (4), 193−202.

(119) Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63* (11), 139−144.

(120) Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv*, December 20, 2013, 1312.6034, ver. *1*. DOI: 10.48550/arXiv.1312.6034.

(121) Springenberg, J. T.; Dosovitskiy, A.; Brox, T.; Riedmiller, M. Striving for simplicity: The all convolutional net. *arXiv*, December 24, 2014, 1412.6806, ver. *1*. DOI: 10.48550/arXiv.1412.6806.

(122) Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Muller, K. R.; Samek, W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS One* **2015**, *10* (7), No. e0130140.

(123) Ancona, M.; Ceolini, E.; Öztireli, C.; Gross, M.Towards better understanding of gradient-based attribution methods for deep neural networks. *arXiv*, November 16, 2017, 1711.06104, ver. *1*. DOI: 10.48550/arXiv.1711.06104.

(124) Olden, J. D.; Jackson, D. A. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model.* **2002**, *154* (1−2), 135−150.

(125) Adel, T.; Ghahramani, Z.; Weller, A. Discovering interpretable representations for both deep generative and discriminative models. In *Proceedings of the 35th International Conference on Machine Learning*, Vol. *80*; PMLR, 2018; pp 50−59.

(126) Mahendran, A.; Vedaldi, A. Understanding deep image representations by inverting them. *2015 Proceedings of the IEEE conference on computer vision and pattern recognition* **2015**, 5188−5196.

(127) Webel, H. E.; Kimber, T. B.; Radetzki, S.; Neuenschwander, M.; Nazare, M.; Volkamer, A. Revealing cytotoxic substructures in molecules using deep learning. *J. Comput. Aided Mol. Des.* **2020**, *34* (7), 731−746.

(128) Preuer, K.; Klambauer, G.; Rippmann, F.; Hochreiter, S.; Unterthiner, T. Interpretable deep learning in drug discovery. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer Nature: Cham, Switzerland, 2019; pp 331−345.

(129) Alber, M.; Lapuschkin, S.; Seegerer, P.; Hägele, M.; Schütt, K. T.; Montavon, G.; Samek, W.; Müller, K.-R.; Dähne, S.; Kindermans, P.-J. iNNvestigate neural networks! *J. Mach. Learn. Res.* **2019**, *20* (93), 1−8.

(130) Waring, J. F.; Jolly, R. A.; Ciurlionis, R.; Lum, P. Y.; Praestgaard, J. T.; Morfitt, D. C.; Buratto, B.; Roberts, C.; Schadt, E.; Ulrich, R. G. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharmacol.* **2001**, *175* (1), 28−42.

(131) Russo, D. P.; Strickland, J.; Karmaus, A. L.; Wang, W.; Shende, S.; Hartung, T.; Aleksunes, L. M.; Zhu, H. Nonanimal Models for Acute Toxicity Evaluations: Applying Data-Driven Profiling and Read-Across. *Environ. Health Perspect.* **2019**, *127* (4), 47001.

(132) Ciallella, H. L.; Russo, D. P.; Sharma, S.; Li, Y.; Sloter, E.; Sweet, L.; Huang, H.; Zhu, H. Predicting Prenatal Developmental Toxicity Based On the Combination of Chemical Structures and Biological Data. *Environ. Sci. Technol.* **2022**, *56* (9), 5984−5998.

(133) Martin, T. M.; Harten, P.; Venkatapathy, R.; Das, S.; Young, D. M. A hierarchical clustering methodology for the estimation of toxicity. *Toxicol. Mech. Method.* **2008**, *18* (2−3), 251−266.

(134) Ball, N.; Cronin, M. T.; Shen, J.; Blackburn, K.; Booth, E. D.; Bouhifd, M.; Donley, E.; Egnash, L.; Hastings, C.; Juberg, D. R. T4 report: Toward good read-across practice (GRAP) guidance. *ALTEX* **2016**, *33* (2), 149.

(135) Luechtefeld, T.; Maertens, A.; Russo, D. P.; Rovida, C.; Zhu, H.; Hartung, T. Analysis of public oral toxicity data from REACH registrations 2008−2014. *ALTEX* **2016**, *33* (2), 111.

(136) Wu, S.; Fisher, J.; Naciff, J.; Laufersweiler, M.; Lester, C.; Daston, G.; Blackburn, K. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem. Res. Toxicol.* **2013**, *26* (12), 1840−1861.

(137) Solimeo, R.; Zhang, J.; Kim, M.; Sedykh, A.; Zhu, H. Predicting chemical ocular toxicity using a combinatorial QSAR approach. *Chem. Res. Toxicol.* **2012**, *25* (12), 2763−2769.

(138) Gallegos-Saliner, A.; Poater, A.; Jeliazkova, N.; Patlewicz, G.; Worth, A. P. Toxmatch—A chemical classification and activity prediction tool based on similarity measures. *Regul. Toxicol. Pharmacol.* **2008**, *52* (2), 77−84.

(139) Dimitrov, S.; Diderich, R.; Sobanski, T.; Pavlov, T.; Chankov, G.; Chapkanov, A.; Karakolev, Y.; Temelkov, S.; Vasilev, R.; Gerova, K.; et al. QSAR Toolbox−workflow and major functionalities. *SAR QSAR Environ. Res.* **2016**, *27* (3), 203−219.

(140) Low, Y.; Sedykh, A.; Fourches, D.; Golbraikh, A.; Whelan, M.; Rusyn, I.; Tropsha, A. Integrative chemical−biological read-across approach for chemical hazard classification. *Chem. Res. Toxicol.* **2013**, *26* (8), 1199−1208.

(141) Guo, Y.; Zhao, L.; Zhang, X.; Zhu, H. Using a hybrid read-across method to evaluate chemical toxicity based on chemical structure and biological data. *Ecotox. Environ. Safe.* **2019**, *178*, 178−187.

(142) Koren, Y.; Bell, R.; Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **2009**, *42* (8), 30−37.

(143) Stein-O'Brien, G. L.; Arora, R.; Culhane, A. C.; Favorov, A. V.; Garmire, L. X.; Greene, C. S.; Goff, L. A.; Li, Y.; Ngom, A.; Ochs, M. F.; Xu, Y.; Fertig, E. J. Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends Genet.* **2018**, *34* (10), 790−805.

(144) Saltelli, A. Sensitivity analysis for importance assessment. *Risk Anal.* **2002**, *22* (3), 579−590.

(145) Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global sensitivity analysis: the primer*; Wiley: New York, 2008; p 1−160.

(146) Hamby, D. M. A review of techniques for parameter sensitivity analysis of environmental models. *Environ. Monit. Assess.* **1994**, *32* (2), 135−154.

(147) Homma, T.; Saltelli, A. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.* **1996**, *52* (1), 1−17.

(148) Carriger, J. F.; Martin, T. M.; Barron, M. G. A Bayesian network model for predicting aquatic toxicity mode of action using two dimensional theoretical molecular descriptors. *Aquat. Toxicol.* **2016**, *180*, 11−24.

(149) Svetnik, V.; Wang, T.; Tong, C.; Liaw, A.; Sheridan, R. P.; Song, Q. Boosting: an ensemble learning tool for compound classification and QSAR modeling. *J. Chem. Inf. Model.* **2005**, *45* (3), 786−799.

(150) Goldstein, A.; Kapelner, A.; Bleich, J.; Pitkin, E. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *J. Comput. Graph. Stat.* **2015**, *24* (1), 44−65.

(151) Krause, J.; Perer, A.; Ng, K. Interacting with predictions: Visual inspection of black-box machine learning models. *Proceedings of the 2016 CHI conference on human factors in computing systems* **2016**, 5686−5697.

(152) Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189−1232.

(153) Molnar, C.; Casalicchio, G.; Bischl, B. iml: An R package for interpretable machine learning. *J. Open Source Softw.* **2018**, *3* (26), 786.

(154) Greenwell, B. M. pdp: An R package for constructing partial dependence plots. *R J.* **2017**, *9* (1), 421.

(155) Herman, J.; Usher, W. SALib: An open-source Python library for sensitivity analysis. *J. Open Source Softw.* **2017**, *2* (9), 97.

(156) Pianosi, F.; Sarrazin, F.; Wagener, T. A Matlab toolbox for global sensitivity analysis. *Environ. Model. Softw.* **2015**, *70*, 80−85.

(157) Andrews, R.; Diederich, J.; Tickle, A. B. Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowl. Based Syst.* **1995**, *8* (6), 373−389.

(158) Craven, M.; Shavlik, J. Extracting tree-structured representations of trained networks. *Adv. Neural Inf. Process. Syst.* **1995**, *8*, 1−7.

(159) Ribeiro, M. T.; Singh, S.; Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* **2016**, 1135−1144.

(160) Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V. *Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more*; O'Reilly Media: Sebastopol, CA, **2019**; pp 163−177.

(161) Ribeiro, M. T.; Singh, S.; Guestrin, C. Anchors: High-precision model-agnostic explanations. *Proceedings of the AAAI conference on artificial intelligence* **2018**, 1527−1535.

(162) Klaise, J.; Van Looveren, A.; Vacanti, G.; Coca, A. Alibi Explain: Algorithms for Explaining Machine Learning Models. *J. Mach. Learn. Res.* **2021**, *22* (1), 8194−8200.

(163) Leist, M.; Ghallab, A.; Graepel, R.; Marchan, R.; Hassan, R.; Bennekou, S. H.; Limonciel, A.; Vinken, M.; Schildknecht, S.; Waldmann, T.; Danen, E.; van Ravenzwaay, B.; Kamp, H.; Gardner, I.; Godoy, P.; Bois, F. Y.; Braeuning, A.; Reif, R.; Oesch, F.; Drasdo, D.; Hohme, S.; Schwarz, M.; Hartung, T.; Braunbeck, T.; Beltman, J.; Vrieling, H.; Sanz, F.; Forsby, A.; Gadaleta, D.; Fisher, C.; Kelm, J.; Fluri, D.; Ecker, G.; Zdrazil, B.; Terron, A.; Jennings, P.; van der Burg, B.; Dooley, S.; Meijer, A. H.; Willighagen, E.; Martens, M.; Evelo, C.; Mombelli, E.; Taboureau, O.; Mantovani, A.; Hardy, B.; Koch, B.; Escher, S.; van Thriel, C.; Cadenas, C.; Kroese, D.; van de Water, B.; Hengstler, J. G. Adverse outcome pathways: opportunities, limitations and open questions. *Arch. Toxicol.* **2017**, *91* (11), 3477−3505.

(164) Plant, N. J. An introduction to systems toxicology. *Toxicol. Res.* **2015**, *4* (1), 9−22.

(165) Hartung, T.; FitzGerald, R. E.; Jennings, P.; Mirams, G. R.; Peitsch, M. C.; Rostami-Hodjegan, A.; Shah, I.; Wilks, M. F.; Sturla, S. J. Systems Toxicology: Real World Applications and Opportunities. *Chem. Res. Toxicol.* **2017**, *30* (4), 870−882.

(166) National Research Council. *Toxicity testing in the 21st century: a vision and a strategy*; National Academies Press: Washington, D.C., 2007; pp 1−163.

(167) Clippinger, A. J.; Allen, D.; Behrsing, H.; BéruBé, K. A.; Bolger, M. B.; Casey, W.; DeLorme, M.; Gaça, M.; Gehen, S. C.; Glover, K. Pathway-based predictive approaches for non-animal assessment of acute inhalation toxicity. *Toxicol. In Vitro* **2018**, *52*, 131−145.

(168) Knapen, D.; Vergauwen, L.; Villeneuve, D. L.; Ankley, G. T. The potential of AOP networks for reproductive and developmental toxicity assay development. *Reprod. Toxicol.* **2015**, *56*, 52−55.

(169) Gijbels, E.; Vinken, M. An Update on Adverse Outcome Pathways Leading to Liver Injury. *Appl. In Vitro Toxicol.* **2017**, *3* (4), 283−285.

(170) Patlewicz, G.; Simon, T. W.; Rowlands, J. C.; Budinsky, R. A.; Becker, R. A. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regul. Toxicol. Pharmacol.* **2015**, *71* (3), 463−477.

(171) Spinu, N.; Cronin, M. T. D.; Enoch, S. J.; Madden, J. C.; Worth, A. P. Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. *Arch. Toxicol.* **2020**, *94* (5), 1497−1510.

(172) Jia, X.; Wen, X.; Russo, D. P.; Aleksunes, L. M.; Zhu, H. Mechanism-driven modeling of chemical hepatotoxicity using structural alerts and an in vitro screening assay. *J. Hazard. Mater.* **2022**, *436*, 129193.

(173) Knapen, D.; Angrish, M. M.; Fortin, M. C.; Katsiadaki, I.; Leonard, M.; Margiotta-Casaluci, L.; Munn, S.; O'Brien, J. M.; Pollesch, N.; Smith, L. C.; Zhang, X.; Villeneuve, D. L. Adverse outcome pathway networks I: Development and applications. *Environ. Toxicol. Chem.* **2018**, *37* (6), 1723−1733.

(174) Judson, R. S.; Magpantay, F. M.; Chickarmane, V.; Haskell, C.; Tania, N.; Taylor, J.; Xia, M.; Huang, R.; Rotroff, D. M.; Filer, D. L.; Houck, K. A.; Martin, M. T.; Sipes, N.; Richard, A. M.; Mansouri, K.; Setzer, R. W.; Knudsen, T. B.; Crofton, K. M.; Thomas, R. S. Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor. *Toxicol. Sci.* **2015**, *148* (1), 137−154.

(175) Ciallella, H. L.; Russo, D. P.; Aleksunes, L. M.; Grimm, F. A.; Zhu, H. 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.* **2021**, *55* (15), 10875−10887.

(176) Ideker, T.; Galitski, T.; Hood, L. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2001**, *2*, 343−372.

(177) Waters, M.; Boorman, G.; Bushel, P.; Cunningham, M.; Irwin, R.; Merrick, A.; Olden, K.; Paules, R.; Selkirk, J.; Stasiewicz, S.; Weis, B.; Van Houten, B.; Walker, N.; Tennant, R. Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *EHP Toxicogenomics* **2003**, *111* (1T), 15−28.

(178) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene ontology: tool for the unification of biology. *Nat. Genet.* **2000**, *25* (1), 25−29.

(179) The Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **2021**, *49* (D1), D325−D334.

(180) Fabregat, A.; Jupe, S.; Matthews, L.; Sidiropoulos, K.; Gillespie, M.; Garapati, P.; Haw, R.; Jassal, B.; Korninger, F.; May, B.; et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **2018**, *46* (D1), D649−D655.

(181) Ma, J.; Yu, M. K.; Fong, S.; Ono, K.; Sage, E.; Demchak, B.; Sharan, R.; Ideker, T. Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* **2018**, *15* (4), 290−298.

(182) Kuenzi, B. M.; Park, J.; Fong, S. H.; Sanchez, K. S.; Lee, J.; Kreisberg, J. F.; Ma, J.; Ideker, T. Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell* **2020**, *38* (5), 672−684.

(183) Hao, Y.; Romano, J. D.; Moore, J. H. Knowledge-guided deep learning models of drug toxicity improve interpretation. *Patterns* **2022**, *3* (9), 100565.

(184) Elmarakeby, H. A.; Hwang, J.; Arafeh, R.; Crowdis, J.; Gang, S.; Liu, D.; AlDubayan, S. H.; Salari, K.; Kregel, S.; Richter, C.; Arnoff, T. E.; Park, J.; Hahn, W. C.; Van Allen, E. M. Biologically informed deep neural network for prostate cancer discovery. *Nature* **2021**, *598* (7880), 348−352.

(185) Lin, C. H.; Lichtarge, O. Using Interpretable Deep Learning to Model Cancer Dependencies. *Bioinformatics* **2021**, *37* (17), 2675−2681.

(186) Jornod, F.; Rugard, M.; Tamisier, L.; Coumoul, X.; Andersen, H. R.; Barouki, R.; Audouze, K. AOP4EUpest: mapping of pesticides in adverse outcome pathways using a text mining tool. *Bioinformatics* **2020**, *36* (15), 4379−4381.

(187) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (43), 15545−15550.

(188) Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsdottir, H.; Tamayo, P.; Mesirov, J. P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **2011**, *27* (12), 1739−1740.

(189) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27−30.

(190) Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38* (suppl_1), D355−D360.

(191) Martens, M.; Ammar, A.; Riutta, A.; Waagmeester, A.; Slenter, D. N.; Hanspers, K. R.; A, M.; Digles, D.; Lopes, E. N.; Ehrhart, F.; Dupuis, L. J.; Winckers, L. A.; Coort, S. L.; Willighagen, E. L.; Evelo, C. T.; Pico, A. R.; Kutmon, M. WikiPathways: connecting communities. *Nucleic Acids Res.* **2021**, *49* (D1), D613−D621.

(192) Cerami, E. G.; Gross, B. E.; Demir, E.; Rodchenkov, I.; Babur, O.; Anwar, N.; Schultz, N.; Bader, G. D.; Sander, C. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **2011**, *39* (Database), D685−D690.

(193) Martini, C.; Liu, Y. F.; Gong, H.; Sayers, N.; Segura, G.; Fostel, J. CEBS update: curated toxicology database with enhanced tools for data integration. *Nucleic Acids Res.* **2022**, *50* (D1), D1156−D1163.

(194) Bastian, M.; Heymann, S.; Jacomy, M. Gephi: an open source software for exploring and manipulating networks. *Proceedings of the international AAAI conference on web and social media* **2009**, 361−362.

(195) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13* (11), 2498−2504.

(196) Azodi, C. B.; Tang, J.; Shiu, S. H. Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends Genet.* **2020**, *36* (6), 442−455.

(197) Piir, G.; Kahn, I.; Garcia-Sosa, A. T.; Sild, S.; Ahte, P.; Maran, U. Best Practices for QSAR Model Reporting: Physical and Chemical Properties, Ecotoxicity, Environmental Fate, Human Health, and Toxicokinetics Endpoints. *Environ. Health Perspect.* **2018**, *126* (12), 126001.

(198) Wu, Y.; Wang, G. Machine Learning Based Toxicity Prediction: From Chemical Structural Description to Transcriptome Analysis. *Int. J. Mol. Sci.* **2018**, *19* (8), 2358.

(199) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model.* **2010**, *50* (7), 1189−1204.

(200) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *J. Chem. Inf. Model.* **2016**, *56* (7), 1243−1252.

(201) Luechtefeld, T.; Maertens, A.; Russo, D. P.; Rovida, C.; Zhu, H.; Hartung, T. Analysis of publically available skin sensitization data from REACH registrations 2008−2014. *ALTEX* **2016**, *33* (2), 135.

(202) Kim, M. T.; Wang, W.; Sedykh, A.; Zhu, H., Curating and Preparing High-Throughput Screening Data for Quantitative Structure-Activity Relationship Modeling. In *High-Throughput Screening Assays in Toxicology*, 2016/08/16 ed.; Methods in Molecular Biology, Vol. *1473*; Humana Press: New York, 2016; pp 161−172.

(203) Papadatos, G.; Gaulton, A.; Hersey, A.; Overington, J. P. Activity, assay and target data curation and quality in the ChEMBL database. *J. Comput. Aided Mol. Des.* **2015**, *29* (9), 885−896.

(204) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2020**, *2* (10), 573−584.

(205) Sheridan, R. P. Interpretation of QSAR models by coloring atoms according to changes in predicted activity: how robust is it? *J. Chem. Inf. Comp. Model.* **2019**, *59* (4), 1324−1337.

(206) Mirza, B.; Wang, W.; Wang, J.; Choi, H.; Chung, N. C.; Ping, P. Machine Learning and Integrative Analysis of Biomedical Big Data. *Genes* **2019**, *10* (2), 87.

(207) Tetko, I. V.; Sushko, I.; Pandey, A. K.; Zhu, H.; Tropsha, A.; Papa, E.; Oberg, T.; Todeschini, R.; Fourches, D.; Varnek, A. Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection. *J. Chem. Inf. Model.* **2008**, *48* (9), 1733−1746.

(208) Zhu, H.; Martin, T. M.; Ye, L.; Sedykh, A.; Young, D. M.; Tropsha, A. Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. *Chem. Res. Toxicol.* **2009**, *22* (12), 1913−1921.

(209) Hanser, T.; Barber, C.; Guesné, S.; Marchaland, J. F.; Werner, S. Applicability domain: towards a more formal framework to express the applicability of a model and the confidence in individual predictions. In *Advances in computational toxicology*; Springer Nature: Cham, Switzerland, 2019; pp 215−232.