Review

Machine Learning Based Toxicity Prediction: From Chemical Structural Description to Transcriptome Analysis

Yunyi Wu and Guanyu Wang *

Department of Biology, Guangdong Provincial Key Laboratory of Cell Microenviroment and Disease Research, Southern University of Science and Technology, Shenzhen 518055, China; wuyy3@mail.sustc.edu.cn
***** Correspondence: wanggy@sustc.edu.cn; Tel.: +86-755-8801-8215; Fax: +86-755-8801-8000

check for
updates

Abstract: Toxicity prediction is very important to public health. Among its many applications, toxicity prediction is essential to reduce the cost and labor of a drug's preclinical and clinical trials, because a lot of drug evaluations (cellular, animal, and clinical) can be spared due to the predicted toxicity. In the era of Big Data and artificial intelligence, toxicity prediction can benefit from machine learning, which has been widely used in many fields such as natural language processing, speech recognition, image recognition, computational chemistry, and bioinformatics, with excellent performance. In this article, we review machine learning methods that have been applied to toxicity prediction, including deep learning, random forests, k-nearest neighbors, and support vector machines. We also discuss the input parameter to the machine learning algorithm, especially its shift from chemical structural description only to that combined with human transcriptome data analysis, which can greatly enhance prediction accuracy.

1. Introduction

Toxicity evaluation is of fundamental importance in drug development and approval. It is well known that drugs must undergo clinical trials to become legal [1,2]. Unfortunately, clinical trials are always associated with certain degree of risk. It was reported that about half of the new drugs were found to be unsafe or ineffective in late human clinical trials [3]. For example, the drug Sitaxentan (Figure 1) was urgently withdrawn from global markets due to specific and irreversible hepatotoxicity in humans [4,5]. Unsafety of clinical trials highlights the importance of preclinical evaluations, which are absolutely necessary in order to prevent toxic drugs from entering into clinical trials.

The animal trial, a common method of preclinical evaluation, is of limited value. On the one hand, the trial is very expensive and laborious. On the other hand, the results offer little guidance to human toxicity reactions, due to inter-species differences and differential disease models [6,7]. For example, Sitaxentan caused no explicit liver injury in animal experiments [8], whereas the hepatotoxicity was prominent in humans [4,5]. Therefore, animal experiments cannot tell the human body's response to new drugs and offer no risk exemption [6,9].

Figure 1. Chemical structural description of Sitaxentan and Sulfisoxazole. (**a**) The 166-bit molecular access system (MACCS) molecular fingerprints, where the different values are indicated in yellow; (**b**) The undirected graphs with atoms as nodes and bonds as edges; (**c**) The molecular structures of Sitaxentan and Sulfisoxazole, where the cyan regions are their common molecular fragment identified by CNN training; (**d**) Other chemical properties.

To reduce the expenses and uncertainties inherent of animal experiments, it is crucial to perform high-throughput computer toxicity predictions. One dominant and most developed toxicity prediction method is Quantitative Structure-Activity Relationships (QSAR) based on chemical structural parameters [10]. This method uses statistics to establish, for a drug compound, a quantitative relationship between the structural or physicochemical characteristics and its physiological activities [11]. From the relationship, one can predict the physiological activities or other properties of the compound, including

toxicity. The earliest and widely used QSAR method was the Hansch approach, as proposed in 1962 [12], which assumes independence of the factors modulating the compounds' biological activities. It relies on methods that are related to free energy and statistical methods, such as linear regression, to obtain the QSAR model [12]. The Free-Wilson method, as proposed in 1964, directly used molecular structure as a variable for regression analysis of physiological activity [13]. In the 1980s, QSAR regression analysis began its application in drug toxicity prediction [14–16]. At the turn of the 21st century, researchers performed toxicity prediction based on single or multiple physicochemical mechanisms [17]. For the single mechanism, linear regression analysis, multivariate analysis, and neural network models were primarily used. For the multiple mechanisms, knowledge based systems were often used besides statistical approaches. Nowadays, with the amount of data increasing explosively, it becomes more and more difficult to maintain completeness of knowledge bases; thus, knowledge-based systems are difficult to complete highly automated work with high volume of data [18]. Meanwhile, statistical approaches, such as linear regression analysis, multivariate analysis, and early shallow neural network models are difficult to extract more abstract features, and are thus difficult to predict with high accuracy.

To address these new challenges, researchers made great efforts to improve both the prediction model (development of machine learning) and inputs to the prediction model (characterization of chemical structure descriptors). The two lines of works interacted with each other and synergistically promoted the field of computer-based toxicity prediction. They are discussed, respectively, in the following.

2. Machine Learning

Machine learning is a branch of artificial intelligence that uses sophisticated algorithms to give computers the ability to learn from the data and make predictions [19]. Main algorithms of machine learning, evolved from the study of cluster analysis and pattern recognition, include artificial neural networks (ANN), decision trees, support vector machines (SVM), and Bayesian classifiers [20]. Besides cluster analysis and pattern recognition, these algorithms have been widely linked to data mining [21].

Due to merits of machine learning, such as fastness, cost-effectiveness, and high accuracy, more and more researchers use machine learning to predict toxicity [22]. Researchers have used a combination of algorithms, such as genetic algorithm (GA) [23,24], random forest (RF) [25–27], artificial neural network (ANN) [28–30], and other machine learning algorithms [31–33] to optimize traditional QSAR models in predicting a drug's toxicity or other biological activities. Different machine learning methods perform differently. Factors such as datasets and computational representations can significantly affect the performance.

2.1. Shallow Architectures

In 1957, Rosenblatt put forward a perceptron model simulating the structure of a neuron, which can be used as a binary classifier [34]. Widrow and Hoff first used Delta rules to train the perceptron and laid the foundation for linear classifier [35]. In 1967, Cover and Hart proposed the nearest neighbor algorithm, which allows for computers to classify sample points according to spatial features [36]. In 1986, Quilan proposed the decision tree algorithm [37]. In 1995, Cortes et al. came up with SVM, the key idea of which was to find a boundary that divides two categories with the largest distance. Besides the linear classification, SVM can be applied to high dimensional nonlinear classification [38]. In 2001, Breiman gave rise to the RF algorithm [39], which is a classifier with multiple decision trees. Individual trees output their respective prediction category, which then vote to determine the final category output of the classifier [40]. It is widely used in solving multiclass problems. SVMs and RFs are both based on statistics; they thus perform well in structured and denser datasets.

In 1986, Hinton et al. invented the back-propagating algorithm (BP) of multi-layer perceptron (MLP) with a sigmoid activation function to perform nonlinear mapping, and used ANN effectively to solve the problem of nonlinear classification and training [41]. Soon, in 1991, it was pointed out that BP with sigmoid activation function has the vanishing gradient problem and is thus difficult to follow deeper and more abstract training [42]. These ANN architectures are thus called shallow learning.

2.2. Deep Learning

In 2011, the ReLU (Rectified Linear Unit) activation function was first proposed [43], which solved the vanishing gradient problem inherent of the sigmoid function. This breakthrough signified the birth of deep learning [44]. Algorithms that are based on ReLU activation function have obtained compelling performance in the field of image recognition [45,46].

As an extension to ANN, deep learning has become a very successful branch of machine learning. It innovates many fields, including pattern recognition, speech recognition [47,48], natural language processing [49,50], image and video recognitions [45,51,52], and life science [53,54]. Deep learning excels when the working data are unstructured, sparse, and large. In recent years, two neural network models, recurrent neural networks (RNN) [55,56], and convolutional neural networks (CNN) [57,58], have been commonly used in deep learning. The former is more suitable for prediction or recognition of sequences, such as natural language processing [59] and time series prediction [60,61]. The latter is more suitable for the recognition of spatial arrangement features, such as the shapes in graphics and images [62].

With the increase of computer speed, the deployment of large-scale distributed clusters [63] and GPUs [64], and the emergence of numerous optimization algorithms [65], deep learning training time reduced greatly and it is now useful to both bioinformatics [66,67] and chemoinformatics [68,69].

3. Chemical Structure Descriptors

Information for toxicity prediction is primarily from the drug compound's chemical structure. To be understood by computers, the chemical structures need to be represented by numbers or characters, the so-called chemical descriptors. Only after chemical structures are converted into descriptors, can the computers efficiently process a large amount of structures, via the computers' high-throughput data processing capacity.

Cammarata and Menon first proposed a molecule-based pattern structure, and established an 8-bit digital chemical descriptor [70,71]. Later, researchers added first-order molecular connectivity values to the existing descriptor indices, for the structural classification of compounds [72]. In addition, a lot of researchers have applied quantum chemistry in order to calculate molecular descriptors (e.g., [73]). By 2000, atoms and bond multiplicity were added to describe the structural parameters of the topology; molecular hydrological, steric, or electronic descriptors were added to explore the relationship between biological activity and chemical structure as well [74]. Around 2001, researchers began to take the three-dimensional (3D) structure of molecules into account to establish 3D-QSAR chemical descriptors [75,76]; some went a step further to generate four-dimensional (4D)-QSAR chemical descriptors by adding molecular dynamics (MD) trajectories and topological information [77].

The descriptor types vary from simple features, like atomic counts or molecular weights to structural features [78]. Different combinations of chemical descriptors and machine learning models might perform differently.

3.1. Traditional Chemical Descriptors

Traditional chemical descriptors are those that are calculated mainly based on molecular structure-derived information, like atomic types, atomic charges, or atomic distances. Table 1 presents the main types of traditional chemical descriptors that are categorized by the calculation parameters [79]. Among them, molecular fingerprints are the most widely used, which are in the form of an array of numbers. They use information, such as atomic attributes, atomic environments, bond properties, and bond position to encode chemical structures [80]. Among them, the 166-bit molecular access system (MACCS) is a typical one (Figure 1a). Each of the 166 bits encodes a specific structural characteristic, such as: whether or not the number of methyl groups in the molecule is greater than 1? whether or not the molecule is aromatic [81]?

Table 1. The main types of traditional chemical descriptors [79].

Descriptor Type	Descriptor Name	Description
Fingerprint-based	ECFP4	atom type, extended connectivity fingerprint, maximum distance = 4
	FCFP4	functional-class-based, extended connectivity fingerprint, maximum distance = 4
	MACCS	166 predefined MDL keys (public set)
Connectivity-matrix-based	BCUT	atomic charges, polarizabilities, H-bond donor and acceptor abilities, and H-bonding modes of intermolecular interaction
Shape-based	rapid overlay of chemical structures (ROCS), combo Tanimoto (shape and electrostatic score)	shape-based molecular similarity method; molecules are described by smooth Gaussian function and pharmacophore points
	PMI	normalized principal moment-of-inertia ratios
Pharmacophore-based	GpiDAPH3	graph-based 3-point pharmacophore, eight atom types computed from three atom properties (in pi system, donor, acceptor)
	TGD	typed graph distances, atom typing (donor, acceptor, polar, anion, cation, hydrophobe)
	TAD	typed atom distances, atom typing (donor, acceptor, polar, anion, cation, hydrophobe)
Bioactivity-based	Bayes affinity fingerprints	bioactivity model based on multicategory Bayes classifier trained on data from ChEMBL v. 14
Physicochemical-property-based	prop2D	physicochemical properties (such as molecular weight, atom counts, partial charges, hydrophobicity etc.)

The importance of molecular fingerprint is easily seen: for those active substances whose functional groups happen to locate at "ortho" or "meta" positions, their toxicity can usually be predicted correctly with MACCS or extended connectivity fingerprint (ECFP) [82]. Autoencoder and convolution based methods are used to predict the chemical properties where chemicals are signified by vectors of fixed length, just like MACCS [68]. In experiments involving combinations of molecular fingerprint and machine learning, Pubchem-SVM and MACCS-RF are the two best combinations. The merits of SVM and RF are apparent. SVM performs the best among many machine learning models, including SVM, RF, k-nearest neighbor (k-NN), and naive Bayes [83]. On the other hand, RF is structured by many decision trees, which are trees with "yes" and "no" as their leaves. Since "yes" and "no" are represented by 1 and 0, respectively, RFs correspond naturally to molecular fingerprints or other chemical descriptors, which consist of many binary digits (0 or 1).

3.2. Deep-Minded Chemical Descriptor

Molecular fingerprints encode chemical structures in great detail (every atom or bond), which may sometimes be unnecessary or even disadvantageous (complicated and inefficient). To obtain a coarse grained, but more deep-mined model, researchers characterized molecules by deep learning architectures, such as RNN and CNN.

One learning method is based on the two-dimensional planar molecular structure, whereby the entire molecule is converted into an undirected graph (Figure 1b). With atoms as nodes and bonds as edges, each node is sequentially traversed [68,84,85]. This would permit an understanding of the relationship between structure and reactivity [86]. Being sensitive to time sequence or succession, RNN and its variant long short-term memory (LSTM) are used to construct this kind of molecular fragments [84,85].

Two-dimensional fragments can be constructed directly from the molecule (Figure 1c), without sequentially traversing every atom in the undirected graph by RNN. CNN classifies the molecules into molecular fragments, which are chemical substructures that are not naturally classified according to the functional groups, but they are adjusted constantly by the "learning" machine. The final molecular fragments should be more interpretable and readable [87]. Using CNN to automatically

construct abstract chemical fragments, the deep learning model showed very high performance in toxicity prediction based on high-throughput data, with an average area under the curve (AUC) of 0.846 [88]. The AUC value is the probability, according to the result of the current algorithm, that the positive sample is ranked before the negative sample when both samples are randomly picked by the algorithm [89]. The greater the AUC value, the more likely the current classification algorithm placing the positive sample before the negative one, and the better the classification.

Figure 1c gives an example of CNN identifying the same substructure (colored in cyan) from two different molecules. It can identify even smaller substructures. After extensive data training, CNN can identify those substructure or molecular fragments that might make a molecule toxic. When working on new test sets, CNN usually predicts with high accuracy [88].

3.3. Chemical Properties

Being determined by molecular structures, chemical properties (molecular weight, degradation rate, solubility coefficient in different solvents, molar index, permeability, etc.) can also be used for classification and prediction (e.g., Figure 1d) [90,91]. The use of molecular descriptor parameters that are derived from electronegativity and covalent radii of forming atoms and interatomic distances can also improve prediction by ANNs [92]. Molecular fingerprints that based on both simple molecular properties and characteristics derived from two-dimensional molecular structures, such as measurements of lipophilicity (LogP and LogD) and topological polar surface area (TPSA), were combined with a variety of machine learning models (e.g., RF, SVM, k-NN) for toxicity prediction and classification. By comparing their performances, it was found that RF usually outperformed [91]. Using the k-NN algorithm, Chavan et al. even tried to predict the chronic toxicity of chemical substances by combining acute toxicity information with molecular fingerprints such as MACCS and CDK [93]. These studies demonstrated that chemical properties can help improve accuracy of toxicity prediction.

3.4. Examples of Chemical Structural Description

Sitaxentan is a drug to treat pulmonary arterial hypertension (PAH) and Sulfisoxazole is a sulfonamide antimicrobial with some hepatotoxicity implications [94]. Their structural descriptions are presented in Figure 1. The two drugs have 22/166 different places and 144/166 identical places in the MACCS molecular fingerprint (Figure 1a). The explicit binary structure of the MACCS molecular fingerprint is well-suited to the structural characteristics of the decision tree algorithm; thus, RF outperformed other machine learning models when dealing with MACCS. Figure 1b displays the undirected graphs of Sitaxentan and Sulfisoxazole, with atoms as nodes and bonds as edges. Every node corresponds to a vector whose terminal point is just the node. The vector can be constructed from the undirected graph by determining the paths of all the other nodes to the terminal point. Finally, all of the vectors are added to form the molecular structure vector of the corresponding molecule [68]. In Figure 1c, the cyan region indicates the same substructure of the two molecules that are identified by CNN. Figure 1d gives the other chemical properties of these two molecules.

4. Chemical Structure Based Toxicity Prediction by Machine Learning

After using computer-readable and interpretable methods to represent the molecular structure, a machine learning model is trained to predict toxicity.

4.1. Data Collection

Accuracy of toxicity prediction depends on the amount of data being collected. During the past years, extensive data collections have resulted in some mainstream toxicity databases (Table 2). Toxicology data network (TOXNET), which was created in 1985, is among the world's largest collection of toxicology databases. The first database that was added to the network was the Hazardous Substances Data Bank (HSDB), which contains acute-toxicity information [95,96]. Toxicity ForeCaster

(ToxCast) is also a widely used high-throughput toxicity database. It is a part of the Toxicology in the 21st Century (Tox21), whose screening workflow is represented in Figure 2. Tox21 contains both acute and chronic toxicity information.

Table 2. The mainstream data resources of toxicity chemicals.

| Database	Database Description	Online Websites	Reference
TOXNET	A collection of toxicity databases.	https://toxnet.nlm.nih.gov/	[97]
ToxCast	High-throughput toxicity data on thousands of chemicals.	https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data	[98]
Tox21	(1) Chemical Effects in Biological Systems; (2) Individual data and summaries from National Toxicology Program studies; (3) The growth, survival, pathology and other toxicology data.	https://ntp.niehs.nih.gov/results/dbsearch/index.html	[99,100]
PubChem	(1) Chemical structures; (2) Identifiers; (3) Chemical and physical properties; (4) Biological activities; (5) Toxicity data (6) Patents and health, safety and so on.	https://pubchem.ncbi.nlm.nih.gov/	[94]
DrugBank	Detailed drug data and corresponding drug target information.	https://www.drugbank.ca/	[101]
ToxBank Data Warehouse	Data for systemic toxicity.	http://www.toxbank.net/data-warehouse	[102]
ECOTOX	Single chemical environmental toxicity data on aquatic life, terrestrial plants and wildlife.	https://cfpub.epa.gov/ecotox/index.html	[103]
SuperToxic	Toxic compound data from literature and web sources.	http://bioinformatics.charite.de/supertoxic/	[104]
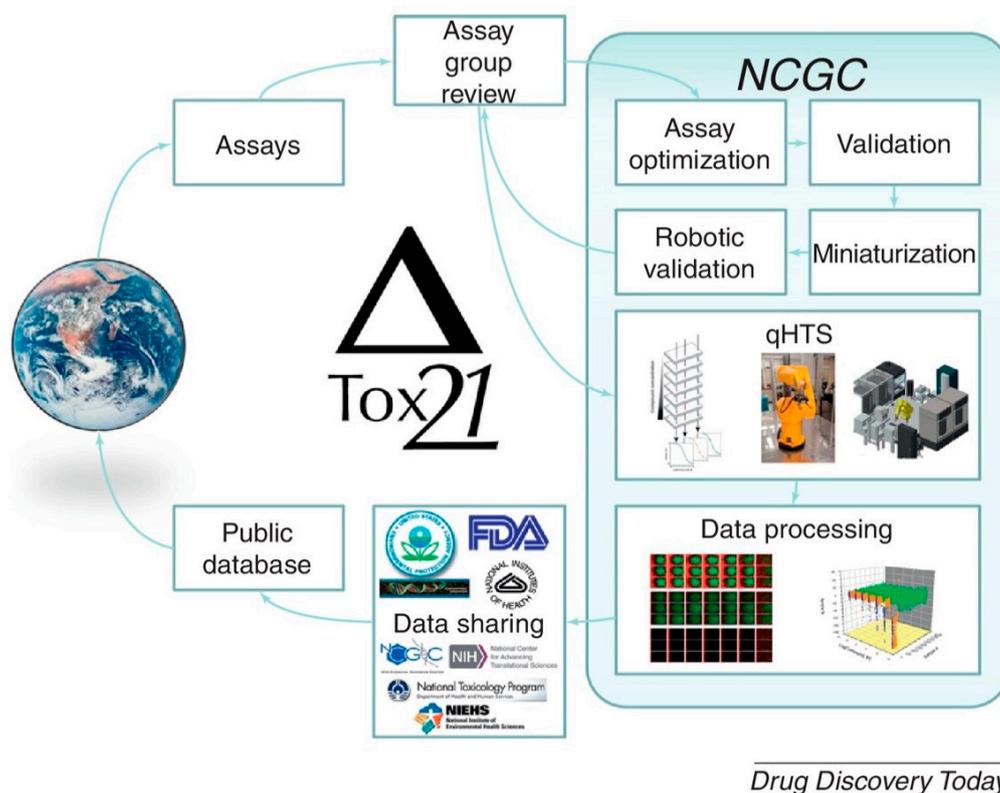
Figure 2. Tox21 screening workflow in drug discovery (qHTS: quantitative high-throughput screening; NCGC: NIH Chemical Genomics Center) [105].

4.2. Performance

The prediction model, which was obtained by combining machine learning and the molecular descriptors, is similar to QSAR, which has long been used to study the quantitative relationship between molecular structure and biological activity [106]. The latter includes toxicity and environmental behavior of chemicals, which makes QSAR one conventional method to predict toxicity [107,108]. Here, we mainly discuss QSAR studies that are based on the two-dimensional structure of chemical molecules combined with biological activity parameters. In the earliest days, researchers used simple pattern recognition methods, such as k-NN, to classify and predict compound toxicity. But, simple pattern recognition is difficult to process asymmetric data, in which positive samples are far less than negative ones, or vice versa [109]. Asymmetric data are ubiquitous in the toxicity database, because non-toxic compounds are not specifically labeled in the database. Fortunately, ANNs and algorithms of the decision tree class, including random forests, can well classify and predict asymmetric data or imbalanced data showing a strong generalization ability [110–112]. For example, with the loss function improving, deep neural networks (DNNs) exhibited excellent performance for classifying even extremely imbalanced data [113].

With molecular fingerprints ECFP6, FP2, MACCS combined with ANN models, the two-dimensional QSAR virtual screening can achieve an average r test value (which measures regression fitness) of 0.75 [114,115]. Deep learning multi-task neural networks worked so well that the AUC value for toxicity QSAR prediction of NIH/3T3 cells (mouse embryonic fibroblast) can reach 0.9, which is slightly higher than the AUC of 0.87 in random forests, in which molecular fingerprints as input of the model [116]. Besides ANNs, RFs have also been successfully applied to QSAR predictions. Using a molecular fingerprint or a simplex representation of molecular structure to store chemical molecular structure information, such as atom type and other physical-chemical characteristics of an atom, RF was validated on the QSAR external test set [25,117]. In addition, Wu et al. recently improved traditional molecular descriptors using element specific persistent homology (ESPH) and auxiliary descriptors, where ESPH includes topological information from intermolecular interactions and homology analysis on each component of molecules. On this basis, they performed RF, Gradient Boosting Decision Tree, single-task deep learning, multi-task deep learning, multi-task deep learning methods, and achieved the highest degree of fitness and accuracy [118].

Table 3 presents the AUC values of different machine learning models combined with different molecular descriptors. One sees that traditional machine learning methods such as SVM and RF have higher AUC values than deep learning algorithms. The reason might be that currently available toxicity datasets are not sufficiently large to support deep learning algorithms to further improve their accuracy. Otherwise, the accuracy of deep learning would increase markedly due to semi-supervised learning characteristics.

Table 3. Comparison of area under the curve (AUC) scores among different combinations of molecular descriptors and machine learning models.

	Molecular Descriptor	Model	AUC	Reference
Shallow architectures	Dragon descriptors (2489 descriptors)	RF	0.81	[119]
	Pubchem keys	SVM	0.948	[83]
	MACCS fingerprints	RF	0.947	[83]
Deep learning	Molecular fragments learned by CNN	DNN	0.837	[88]
	Unidirectional graph learned by CNN	Graph CNN	0.867	[120]
	LSTM graph	One-shot learning	0.84	[84]

5. Acute (Immediate) Toxicity Prediction

Toxicity can be divided into acute toxicity and chronic toxicity. The latter includes toxicity to reproduction, mutagenicity, and carcinogenicity [121]. Acute toxicity is usually measured by LD50

(Lethal Dose 50) for drug testing and LC50 (Lethal Concentration 50) for environmental sciences [122]. In 1997, Gute and Basak used the simplest linear regression to predict acute aquatic toxicity [123]. In 2000, Basak et al. used ANN to predict LC50 of benzene derivatives [124]. After the development of machine learning, in 2011, Lu et al. used k-NN combined linear regression model to predict acute oral toxicity in rats and achieved a R square value of 0.712, for which they utilized the local chemical structure that was represented by molecular fingerprints [31]. Martin et al. used the global hierarchical clustering method to predict acute toxicity of pesticides and obtained better results than linear regression [125].

Recently, Liu et al. compared performance of shallow architectures, such as RF and k-NN, with DNN in acute toxicity prediction based on extremely unbalanced datasets. For the sake of fairness, they used the chemical descriptor of ECFP uniformly. It was found that RF and DNN performed better on the global dataset, while k-NN performed better on the unbalanced acute toxicity datasets. This result also highlights the importance of neighbor information in acute toxicity prediction [126]. In order to adapt the chemical descriptor to the prediction model, Xu et al. used an enhanced molecular graph encoding convolutional neural networks (MGE-CNN) (the gray box in Figure 3) to process the standard molecular structure data, and finally obtained the fingerprint. The fingerprint was further mined both forwardly and backwardly, which yielded the deep-minded fingerprint (the array of black dots in Figure 3). The deep-minded fingerprint was then tested by the regression model (the blue circle) and the multiclass/multitask models (the green circles), which yielded a classification accuracy up to 95.0% and a regression R square value of 0.861 [127].
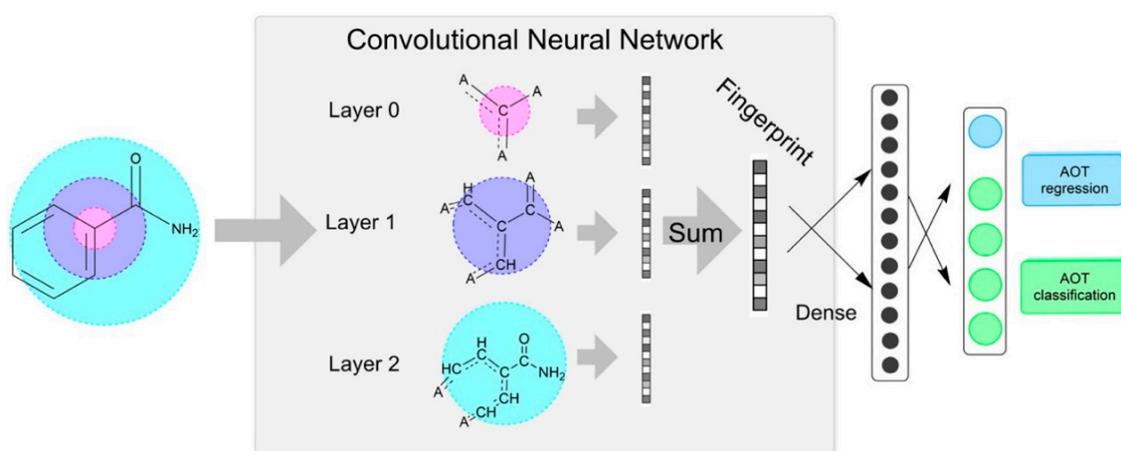

Figure 3. An acute oral toxicity prediction. The prediction starts from a chemical molecular structure in the simplified molecular-input line-entry system (SMILES) format, as an input to the MEG-CNN, where the pink, purple, and cyan circles represent the first, second, and third iterations, respectively. During each iteration, the chemical structure is processed by the convolutional kernel according to the atom degree to obtain the corresponding pre-fingerprint. All of the pre-fingerprints are integrated to generate the fingerprint, which was further processed to generate the deep-mined fingerprint. The deep-minded fingerprint was then tested by the regression model (the blue circle) and the multiclass/multitask models (the green circles) [127].

6. Chronic (Delayed) Toxicity Prediction

6.1. Prediction Based on Chemical Structure

When compared with acute toxicity, chronic toxicity is more latent and hard to discover. Chavan et al. classified the LD50 values of compounds using k-NN. Based on the classification, they predicted the LOEL (lowest observed effect level), which was then used to measure chronic toxicity. The R square value of the test set was only 0.54, however [93]. In 2017, Li et al. used machine learning models,

such as RF, SVM, and k-NN to predict the oral LOAEL (lowest observed adverse effect level) of rats. The method k-NN obtained the best performance, yielding AUC values up to 0.814 [128].

6.2. Prediction with Cellular Transcriptome Information

Chemical structure based toxicity prediction is only the first step of drug evaluation. The subsequent steps include cell, animal, and clinical toxicity tests. Because drugs are designed for humans, toxicity testing on human cells is both clinically relevant and cost effective. Whole genome expression, or transcriptome expression, reflects the state changes of a cell, either in vivo or in vitro. For example, if a cell has a high expression of a proto-oncogene, then the chance is high of the cell's carcinogenesis. Therefore, machines should fully exploit gene expression data for feature selection and classification in drug trials [129]. Deep-sequencing RNA-Seq technology has led to an unrivaled explosion in the amount of data, which would help researchers to gain a deeper understanding of biological mechanisms (e.g., changes of cellular signaling pathways) of toxic compounds, such as Benzo[a]pyrene. This would, in turn, help researchers to better characterize harmful effects that are caused by chemicals [130].

These technical developments make the following strategies practical. One can induce changes of whole genome expression of cultured human cells of a specific type by adding a test drug to the culture. By analyzing changes in the transcriptome, toxicity of the drug to the cell type, and to the corresponding organ, can be predicted [131]. Schwartz et al. used both toxic and non-toxic compounds to treat 3D-cultured human pluripotent stem cell-derived neural cells, then used RNA-Seq to determine the whole genome expression profile, and then used SVM to classify the chemicals according to their toxicity. The scheme gained an average AUC value of 0.91 [132]. Yamane et al. used chemicals to treat human embryonic stem cells and analyzed their transcriptomes. By classifying the chemicals into different categories, such as neurotoxins, genotoxic carcinogens, and non-genotoxic carcinogens, and by analyzing gene interaction networks, they gained much richer information, which greatly improved the accuracy of toxicity prediction and even allowed for them to predict the delayed chemical toxicity with SVM [133]. What underlay their success was the fact that delayed toxicity is associated with changes in gene expression, which can, in turn, affect the expression of downstream genes [134,135]. Although the number of affected genes is small at the induction, much greater gene expression changes will occur 24 h after induction [136]. Therefore, the accurate prediction of late-onset chemical toxicity might be ascribed to the analysis of gene interaction networks: alterations that are caused by a compound propagate through gene-gene interactions; and, the chain reactions finally lead to genome instability and cytotoxicity. Because gene expression is not immediate, toxicity onset is often delayed and it is difficult to detect immediately after the induction. Following the same logic, the degree of toxicity would positively correlate with the degree of connectivity of the genetic network, because the number of affected genes would increase explosively as the complexity of the network increases [137,138].

Based on a large-scale dataset of gene expression, and by using drugs' chemical structure as the input and the altered gene expression as the output, Liu et al. established a variable-nearest neighbor model to predict the QSAR between chemical structures and gene expression profiles, and obtained an AUC value of more than 0.7 [139].

7. An in Silico Platform of Deep Learning Based Toxicity Prediction

On the basis of the above researches, we are establishing a pertinent system encompassing all of the major aspects of toxicity prediction: chemical structure, gene expression, deep learning, etc. Besides immediate toxicity prediction, delayed toxicity can also be predicted (Figure 4). In this system, drug molecular structures are represented by chemical fragments learned by CNN [88]. Gene expression data are mainly obtained by splicing gene embedding identified by RNA-Seq.
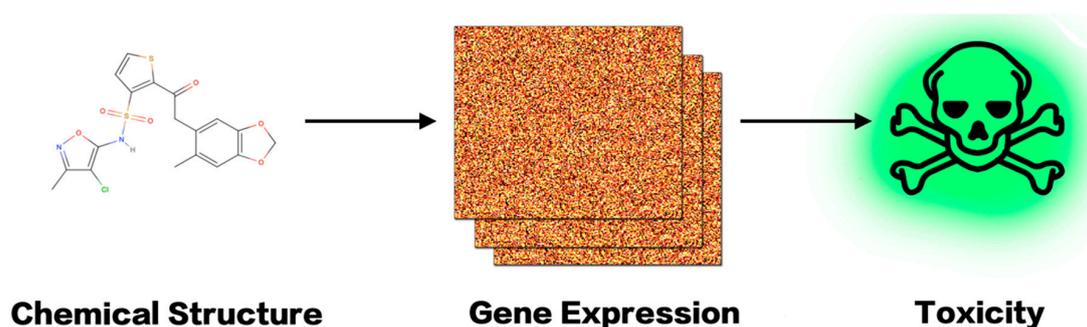
Figure 4. Toxicity prediction with gene expression data.

7.1. Collection of Gene Expression Data

Table 4 represents the databases we are using to gain gene expression data after drug treatment to the cells. Among the databases, CMap is the most popular one to analyze the relationship between transcriptome data and drugs [140].

Table 4. Databases of drug induced gene expression.

Database	Description		Websites	References
GEO database	Gene expression data of drug-treated samples in subsets.		https://www.ncbi.nlm.nih.gov/geo/	[141,142]
Connectivity Map (CMap)	(1)	Genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules;	https://portals.broadinstitute.org/cmap/	[140]
	(2)	Simple pattern-matching of functional connections between drugs, genes and diseases through the transitory feature of common gene-expression changes.		
DSigDB	(1)	Drug and small molecule-related genes based on quantitative inhibition;	http://tanlab.ucdenver.edu/DSigDB	[143]
	(2)	Drug-induced gene expression changes data.		
LINCS Canvas Browser (LCB)	(1)	Experiment data about the landmark gene expression changes in response to a drug;	http://www.maayanlab.net/LINCS/LCB	[144]
	(2)	Both gene expression records before and after drug application.		
Therapeutic target database (TTD)	(1)	Drug resistance mutations in drug-target genes;	http://bidd.nus.edu.sg/group/ttd/ttd.asp	[145]
	(2)	Drug resistance mutations in regulatory genes;		
	(3)	Differential expression profiles of drug-targets in the disease-relevant drug-targeted tissues of different diseases;		
	(4)	Expression profiles of drug-targets in the non-targeted tissues of healthy individuals;		
	(5)	Target combinations of different drugs.		
Comparative Toxicogenomics Database (CTD)	(1)	Cross-species chemical-gene/protein interactions data;	http://ctdbase.org/	[146]
	(2)	Chemical- and gene-disease relationships.		
Drug-Path	Drug-induced pathways.		http://www.cuilab.cn/drugpath	[147]
CancerDR	(1)	Anticancer drugs and their effectiveness against cancer cell lines;	http://crdd.osdd.net/raghava/cancerdr/	[148]
	(2)	Drug target gene information like function, structure, and gene sequences in respective cancer cell lines.		
KEGG DRUG	(1)	Chemical structures and/or chemical components;	https://www.genome.jp/kegg/drug/	[149]
	(2)	The interaction network with target molecules, metabolizing enzymes, and other drugs;		
	(3)	The chemical structure transformation network in the history of drug development.		

7.2. Representation of Gene Expression Data

Each of these human gene embeddings can be represented by a 300-dimensional gene vector trained from 984 datasets of the GEO database based on gene co-expression patterns [150]. This vector

representation reflects gene functions indirectly. Besides this co-expression based gene embedding, there are other methods for vector representation of genes. One method is similar to word2vec used in natural language processing [151,152]. The method word2vec converts words into vectors that are computer understandable by using shallow neural networks with a large amount of neurons. In another method, vectors are constructed based on a similarity of different gene annotations in Gene Ontology, which allows for the quantification of similarities between genes [153]. This representation directly reflects gene functions and indirectly reflects gene interactions. Besides the use of gene vectors, the dimension of RNA-Seq data can be reduced by techniques, such as Stacked Denoising Autoencoder (SDAE), which allows for the discovery of gene interaction patterns [154] and specific gene expression patterns [155] by extracting features from RNA-Seq data by a supervised learning classification model. By scoring pathway activation and regarding "landmark genes" as new features to perform dimensionality reduction, Aliper et al. combined processed gene expression data with DNNs to identify the pharmacological properties of multiple drugs under different biological systems and conditions [156].

With gene expression data at hand and with chemical structures digitalized, one can use the system to find deeper and intrinsic links between the two through machine learning models (Figure 5), by either establishing the association with chemical structures as input and gene expressions as output (from structure to effect), or vice versa (from effect to structure). The former can help with QSAR prediction, including toxicity, while the latter can help with the design of inducing drugs based on the desired changes of gene expression pattern.
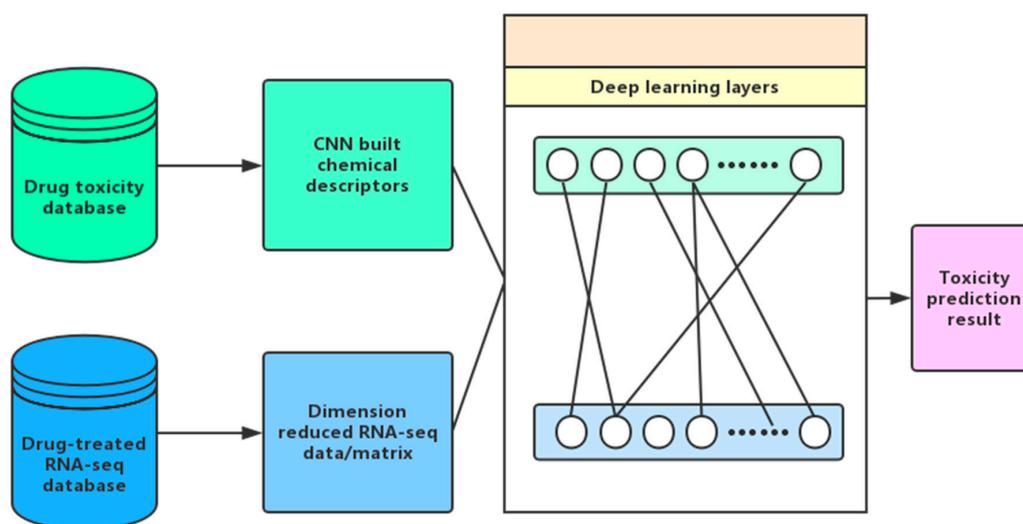

Figure 5. Toxicity prediction with RNA-seq data.

7.3. Toxicity Prediction

Incorporating genetic information would render more accurate toxicity prediction and QSAR construction [133]. The fundamental reason is that changes in gene expression provide biological information, which is much richer and more complex than the simple molecular structure and chemical properties. Furthermore, the biological information is not only at the molecular level, involving only a single pair of drug-protein interaction, but also at the systems level with a drug targeting the whole gene interaction network, affecting the whole cell and even the whole organism.

One can not only distinguish between toxic and non-toxic, but also perform classified toxicity prediction (neurotoxins, carcinogens, etc.). For example, Gayvert et al. performed classified toxicity prediction on FDA-approved drugs and drugs that had failed to pass toxicity-tests, with the RF supervised learning algorithm. The learning was from multiple sources: chemical structure

characterizations, the median value of the expression of the drug targeted genes from the transcriptome of various tissues, the frequency or possibility of functional mutations (i.e., drug induced gene mutations that lead to loss of function). They finally obtained an AUC value of about 0.8263 [157]. Calculation of median expression of drug target genes is useful, but they may ignore tissue specificity and differential toxic reactions. For example, a toxic drug may induce high expression of a particular gene only in the liver, but not in the other organs or tissues. The median value of the gene expression, being based on the whole body measurement, is thus very low and cannot reflect the drug's toxicity specific to the liver. In this event, the use of tissue transcriptome data might be more specific and can help to extract more relevant features.

When compared with the random forest approach, the deep learning approach can handle higher throughput and larger amounts of data, and be capable to deal with higher-level and more abstract features, resulting in a better performance after subsequent data accumulation.

8. Summary

The 21st century has witnessed the rapid development of artificial intelligence, including machine learning. This rapid development is partly stimulated by its many important applications, one of which is drug toxicity prediction in silico [88,127,158]. Together with "Big Data" science [159], machine learning techniques may provide much more information about toxicity than ever before.

In this article, we have reviewed machine learning methods that have been applied to toxicity prediction. We have also discussed the input parameter to the machine learning algorithm, especially its shift from chemical structural description only to that combined with human transcriptome data analysis, which can greatly enhance prediction accuracy.

The merits of machine learning based toxicity prediction are summarized, as follows. Firstly, many harmful and risky animal or clinical trials can be spared, due to toxicity predicted by computers. Secondly, in silico prediction is risk-free, low-costly, and of high throughput. Thirdly, because human transcriptome data are often used, the prediction is essentially based on system-level complexities; the method is thus more global than those studying single protein related toxicity. Finally, due to its capacity of extracting complex and abstract features in pharmacology and bioinformatics applications [160], machine learning may eventually become completely in silico, as the data continue to expand and the accuracy continues to improve.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ting, N. (Ed.) Introduction and New Drug Development Process. In *Dose Finding in Drug Development*; Springer: New York, NY, USA, 2006; pp. 1–17.
2. Janodia, M.D.; Sreedhar, D.; Virendra, L.; Ajay, P.; Udupa, N. Drug Development Process: A review. *Pharm. Rev.* **2007**, *5*, 2214–2221.
3. Hwang, T.J.; Carpenter, D.; Lauffenburger, J.C.; Wang, B.; Franklin, J.M.; Kesselheim, A.S. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. *JAMA Intern. Med.* **2016**, *176*, 1826–1833. [CrossRef] [PubMed]
4. Erve, J.C.; Gauby, S.; Maynard, M.J., Jr.; Svensson, M.A.; Tonn, G.; Quinn, K.P. Bioactivation of sitaxentan in liver microsomes, hepatocytes, and expressed human P450s with characterization of the glutathione conjugate by liquid chromatography tandem mass spectrometry. *Chem. Res. Toxicol.* **2013**, *26*, 926–936. [CrossRef] [PubMed]
5. Galiè, N.; Hoeper, M.M.; Simon, J.; Gibbs, R.; Simonneau, G. Liver toxicity of sitaxentan in pulmonary arterial hypertension. *Eur. Heart J.* **2011**, *32*, 386–387. [CrossRef] [PubMed]

6.	Johnson, D.E. Fusion of nonclinical and clinical data to predict human drug safety. *Expert Rev. Clin. Pharmacol.* **2013**, *6*, 185–195. [CrossRef] [PubMed]

7.	Akhtar, A. The Flaws and Human Harms of Animal Experimentation. *Camb. Q. Healthc. Ethics* **2015**, *24*, 407–419. [CrossRef] [PubMed]

8.	Owen, K.; Cross, D.M.; Derzi, M.; Horsley, E.; Stavros, F.L. An overview of the preclinical toxicity and potential carcinogenicity of sitaxentan (Thelin®), a potent endothelin receptor antagonist developed for pulmonary arterial hypertension. *Regul. Toxicol. Pharmacol.* **2012**, *64*, 95–103. [CrossRef] [PubMed]

9.	Thomas, R.S.; Paules, R.S.; Simeonov, A.; Fitzpatrick, S.C.; Crofton, K.M.; Casey, W.M.; Mendrick, D.L. The US Federal Tox21 Program: A strategic and operational plan for continued leadership. *Altex* **2018**, *35*, 163–168. [CrossRef] [PubMed]

10.	Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R. QSAR Modeling: Where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977–5010. [CrossRef] [PubMed]

11.	Roy, K.; Kar, S.; Das, R.N. Chapter 7—Validation of QSAR Models. In *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*; Roy, K., Kar, S., Das, R.N., Eds.; Academic Press: Boston, MA, USA, 2015; pp. 231–289.

12.	Hansch, C.; Maloney, P.P.; Fujita, T.; Muir, R.M. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178–180. [CrossRef]

13.	Free, S.M.; Wilson, J.W. A Mathematical Contribution to Structure-Activity Studies. *J. Med. Chem.* **1964**, *7*, 395–399. [CrossRef] [PubMed]

14.	Quinn, F.R.; Neiman, Z.; Beisler, J.A. Toxicity and quantitative structure-activity relationships of colchicines. *J. Med. Chem.* **1981**, *24*, 636–639. [CrossRef] [PubMed]

15.	Denny, W.A.; Cain, B.F.; Atwell, G.J.; Hansch, C.; Panthananickal, A.; Leo, A. Potential antitumor agents. 36. Quantitative relationships between experimental antitumor activity, toxicity, and structure for the general class of 9-anilinoacridine antitumor agents. *J. Med. Chem.* **1982**, *25*, 276–315. [CrossRef] [PubMed]

16.	Denny, W.A.; Atwell, G.J.; Cain, B.F. Potential antitumor agents. 32. Role of agent base strength in the quantitative structure-antitumor relationships for 4′-(9-acridinylamino) methanesulfonanilide analogs. *J. Med. Chem.* **1979**, *22*, 1453–1460. [CrossRef] [PubMed]

17.	Barratt, M.D. Prediction of toxicity from chemical structure. *Cell Biol. Toxicol.* **2000**, *16*, 1–13. [CrossRef] [PubMed]

18.	Compton, P.; Preston, P.; Edwards, G.; Kang, B. Knowledge Based Systems That Have Some Idea of Their Limits. *CIO* **2000**, *15*, 57–63.

19.	Mitchell, T.M. *Machine Learning*; McGraw Hill: Ridge, IL, USA, 1997; Volume 45, pp. 870–877.

20.	Bishop, C.M. *Pattern Recognition and Machine Learning*, 1st ed.; Springer: New York, NY, USA, 2006; p. 738.

21.	Fürnkranz, J.; Gamberger, D.; Lavrač, N. Machine Learning and Data Mining. *Comput. Study* **2010**, *42*, 110–114.

22.	Yang, H.; Sun, L.; Li, W.; Liu, G.; Tang, Y. Corrigendum: In Silico Prediction of Chemical Toxicity for Drug Design Using Machine Learning Methods and Structural Alerts. *Front. Chem.* **2018**, *6*, 129. [CrossRef] [PubMed]

23.	Hemmateenejad, B.; Akhond, M.; Miri, R.; Shamsipur, M. Genetic algorithm applied to the selection of factors in principal component-artificial neural networks: Application to QSAR study of calcium channel antagonist activity of 1,4-dihydropyridines (nifedipine analogous). *Cheminform* **2003**, *34*, 1328–1334. [CrossRef]

24.	Hoffman, B.T.; Kopajtic, T.; And, J.L.K.; Newman, A.H. 2D QSAR Modeling and Preliminary Database Searching for Dopamine Transporter Inhibitors Using Genetic Algorithm Variable Selection of Molconn Z Descriptors. *J. Med. Chem.* **2000**, *43*, 4151–4159. [CrossRef] [PubMed]

25.	Polishchuk, P.G.; Muratov, E.N.; Artemenko, A.G.; Kolumbin, O.G.; Muratov, N.N.; Kuz'Min, V.E. Application of random forest approach to QSAR prediction of aquatic toxicity. *J. Chem. Inform. Model.* **2009**, *49*, 2481–2488. [CrossRef] [PubMed]

26.	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. Inform. Comput. Sci.* **2015**, *43*, 1947. [CrossRef] [PubMed]

27. Svetnik, V.; Liaw, A.; Tong, C.; Wang, T. Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules. In Proceedings of the Multiple Classifier Systems, International Workshop, MCS 2004, Cagliari, Italy, 9–11 June 2004; Roli, F., Kittler, J., Windeatt, T., Eds.; Springer: Berlin/Heidelberg, Germany; Cagliari, Italy, 2004; pp. 334–343.

28. Agrafiotis, D.K.; Cedeño, W.; Lobanov, V.S. On the use of neural network ensembles in QSAR and QSPR. *J. Chem. Inform. Comput. Sci.* **2002**, *42*, 903–911. [CrossRef]

29. Wikel, J.H.; Dow, E.R. The use of neural networks for variable selection in QSAR. *Bioorgan. Med. Chem. Lett.* **1993**, *3*, 645–651. [CrossRef]

30. Lu, X.; Ball, J.W.; Dixon, S.L.; Jurs, P.C. Quantitative structure-activity relationships for toxicity of phenols using regression analysis and computational neural networks. *Environ. Toxicol. Chem.* **1998**, *13*, 841–851.

31. Lu, J.; Peng, J.; Wang, J.; Shen, Q.; Bi, Y.; Gong, L.; Zheng, M.; Luo, X.; Zhu, W.; Jiang, H. Estimation of acute oral toxicity in rat using local lazy learning. *J. Cheminform.* **2014**, *6*, 26. [CrossRef] [PubMed]

32. Mazzatorta, P.; Cronin, M.T.D.; Benfenati, E. A QSAR Study of Avian Oral Toxicity using Support Vector Machines and Genetic Algorithms. *Qsar Comb. Sci.* **2010**, *25*, 616–628. [CrossRef]

33. Srinivasan, A.; King, R.D. *Using Inductive Logic Programming to construct Structure-Activity Relationshipsp*; AAAI: Menlo Park, CA, USA, 1999; pp. 64–73.

34. Rosenblatt, F. *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*; MIT Press: Cambridge, MA, USA, 1988; pp. 386–408.

35. Widrow, B.; Hoff, M.E. Adaptive Switching Circuits. In *Neurocomputing: Foundations of Research*; Ire Wescon Conv. Rec; MIT Press: Cambridge, MA, USA, 1966.

36. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **2002**, *13*, 21–27. [CrossRef]

37. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]

38. Cortes, C.; Vapnik, V.; Cortes, C.; Vapnik, V.; Llorens, C.; Vapnik, V.N.; Cortes, C.; Côrtes, M.V.C.B. Support-vector networks. *Mach. Learn.* **1995**, *20*, 27–297. [CrossRef]

39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

40. Tin Kam, H. Random Decision Forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; Volume 1, pp. 278–282.

41. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Read. Cognit. Sci.* **1986**, *323*, 399–421. [CrossRef]

42. Hochreiter, S. The Vanishing Gradient Problem during Learning Recurrent Neural Nets and Problem Solutions. *Int. J. Uncertain. Fuzz. Knowl.-Based Syst.* **1998**, *6*, 107–116. [CrossRef]

43. Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; Geoffrey, G., David, D., Miroslav, D., Eds.; PMLR Proceedings of Machine Learning Research, 2011; Volume 15, pp. 315–323.

44. Zahangir Alom, M.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Hasan, M.; Van Esesn, B.C.; Awwal, A.A.S.; Asari, V.K. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. *arXiv* **2018**, arXiv:1803.01164.

45. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Lake Tahoe, NV, USA, 2012; Volume 1, pp. 1097–1105.

46. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Kai, L.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.

47. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [CrossRef]

48. Dahl, G.E.; Acero, A. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 30–42. [CrossRef]

49. Luong, T.; Socher, R.; Manning, C.D. Better Word Representations with Recursive Neural Networks for Morphology. In Proceedings of the Seventeenth Conference on Computational Natural Language Learning, Sofia, Bulgaria, 8–9 August 2013; pp. 104–113.

50.　Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 27: 28th Annual Conference on Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13 December 2014; Volume 4, pp. 3104–3112.

51.　Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Volume 8689, pp. 818–833.

52.　Donahue, J.; Hendricks, L.A.; Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Darrell, T.; Saenko, K. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; p. 677.

53.　Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **2016**, *12*, 878. [CrossRef] [PubMed]

54.　Webb, S. Deep learning for biology. *Nature* **2018**, *554*, 555–557. [CrossRef] [PubMed]

55.　Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. [CrossRef] [PubMed]

56.　Pineda, F.J. Recurrent Backpropagation and the Dynamical Approach to Adaptive Neural Computation. *Neural Comput.* **1989**, *1*, 161–172. [CrossRef]

57.　Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113. [CrossRef] [PubMed]

58.　LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. In *The Handbook of Brain Theory and Neural Networks*; Michael, A.A., Ed.; MIT Press: Cambridge, MA, USA, 1998; pp. 255–258.

59.　Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078.

60.　Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci. Rep.* **2018**, *8*, 6085. [CrossRef] [PubMed]

61.　Madhavan, P.G. Recurrent neural network for time series prediction. In Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Societ, San Diego, CA, USA, 31 October 1993; pp. 250–251.

62.　Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial Transformer Networks. *arXiv* **2015**, arXiv:1506.02025.

63.　Dean, J.; Corrado, G.S.; Monga, R.; Chen, K.; Devin, M.; Le, Q.V.; Mao, M.Z.; Ranzato, M.A.; Senior, A.; Tucker, P.; et al. Large scale distributed deep networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Lake Tahoe, NV, USA, 2012; Volume 1, pp. 1223–1231.

64.　Raina, R.; Madhavan, A.; Ng, A.Y. Large-scale deep unsupervised learning using graphics processors. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; ACM: Montreal, QC, Canada, 2009; pp. 873–880.

65.　Bottou, L. *Large-Scale Machine Learning with Stochastic Gradient Descent*; Physica-Verlag HD: Heidelberg, Germany, 2010; pp. 177–186.

66.　Min, S.; Lee, B.; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **2016**, *18*, 851–869. [CrossRef] [PubMed]

67.　Kuzminykh, D.; Polykovskiy, D.; Kadurin, A.; Zhebrak, A.; Baskov, I.; Nikolenko, S.; Shayakhmetov, R.; Zhavoronkov, A. 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks. *Mol. Pharm.* **2018**. [CrossRef] [PubMed]

68.　Lusci, A.; Pollastri, G.; Baldi, P. Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules. *J. Chem. Inform. Model.* **2013**, *53*, 1563–1575. [CrossRef] [PubMed]

69.　Kim, I.-W.; Oh, J.M. Deep learning: From chemoinformatics to precision medicine. *J. Pharm. Investig.* **2017**, *47*, 317–323. [CrossRef]

70.　Cammarata, A.; Menon, G.K. Pattern recognition. Classification of therapeutic agents according to pharmacophores. *J. Med. Chem.* **1976**, *19*, 739–748. [CrossRef] [PubMed]

71.　Menon, G.K.; Cammarata, A. Pattern recognition II: Investigation of structure—Activity relationships. *J. Pharm. Sci.* **1977**, *66*, 304–314. [CrossRef] [PubMed]

72. Henry, D.R.; Block, J.H. Classification of drugs by discriminant analysis using fragment molecular connectivity values. *J. Med. Chem.* **1979**, *22*, 465–472. [CrossRef] [PubMed]

73. Karelson, M.; Lobanov, V.S.; Katritzky, A.R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027–1044. [CrossRef] [PubMed]

74. Devillers, J.; Balaban, A.T. *Topological Indices and Related Descriptors in QSAR and QSPAR*; CRC Press: Boca Raton, FL, USA, 2000.

75. Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inform. Comput. Sci.* **2002**, *42*, 693–705. [CrossRef]

76. Kiss, L.E.; Kövesdi, I.; Rábai, J. An improved design of fluorophilic molecules: Prediction of the ln P fluorous partition coefficient, fluorophilicity, using 3D QSAR descriptors and neural networks. *J. Fluor. Chem.* **2001**, *108*, 95–109. [CrossRef]

77. Ataide Martins, J.P.; Ma, R.D.O.; Ms, O.D.Q. Web-4D-QSAR: A web-based application to generate 4D-QSAR descriptors. *J. Comput. Chem.* **2018**, *39*, 917–924. [CrossRef] [PubMed]

78. Roy, K.; Kar, S.; Das, R.N. Chapter 2—Chemical Information and Descriptors. In *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*; Roy, K., Kar, S., Das, R.N., Eds.; Academic Press: Boston, MA, USA, 2015; pp. 47–80.

79. Koutsoukas, A.; Paricharak, S.; Galloway, W.R.J.D.; Spring, D.R.; Ijzerman, A.P.; Glen, R.C.; Marcus, D.; Bender, A. How Diverse Are Diversity Assessment Methods? A Comparative Analysis and Benchmarking of Molecular Descriptor Space. *J. Chem. Inform. Model.* **2014**, *54*, 230–242. [CrossRef] [PubMed]

80. Durant, J.L.; Leland, B.A.; Henry, D.R.; Nourse, J.G. Reoptimization of MDL Keys for Use in Drug Discovery. *Cheminform* **2003**, *34*, 1273–1280. [CrossRef]

81. Greg Landrum. *Source Code for Module rdkit.Chem.MACCSkeys*; Greg Landrum: Basel, Switzerland, 2011.

82. Banerjee, P.; Siramshetty, V.B.; Drwal, M.N.; Preissner, R. Computational methods for prediction of in vitro effects of new chemical structures. *J. Cheminform.* **2016**, *8*, 51. [CrossRef] [PubMed]

83. Fan, D.; Yang, H.; Li, F.; Sun, L.; Di, P.; Li, W.; Tang, Y.; Liu, G. In silico prediction of chemical genotoxicity using machine learning methods and structural alerts. *Toxicol. Res.* **2018**, *7*, 211–220. [CrossRef]

84. Altae-Tran, H.; Ramsundar, B.; Pappu, A.S.; Pande, V. Low Data Drug Discovery with One-Shot Learning. *Acs Cent. Sci.* **2016**, *3*, 283–293. [CrossRef] [PubMed]

85. Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L. Deep Learning for Drug-Induced Liver Injury. *J. Chem. Inform. Model.* **2015**, *55*, 2085–2093. [CrossRef] [PubMed]

86. Dias, J.R.; Milne, G.W.A. Chemical Applications of Graph Theory. *J. Chem. Inform. Model.* **1976**, *32*, 210–242. [CrossRef]

87. Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Hirzel, T.; Adams, R.P. Convolutional networks on graphs for learning molecular fingerprints. In Proceedings of the International Conference on Neural Information Processing Systems, Istanbul, Turkey, 9–12 November 2015; pp. 2224–2232.

88. Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. DeepTox: Toxicity Prediction using Deep Learning. *Front. Environ. Sci.* **2016**, *3*. [CrossRef]

89. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]

90. Marvuglia, A.; Kanevski, M.; Benetto, E. Machine learning for toxicity characterization of organic chemical emissions using USEtox database: Learning the structure of the input space. *Environ. Int.* **2015**, *83*, 72–85. [CrossRef] [PubMed]

91. Sharma, A.K.; Srivastava, G.N.; Roy, A.; Sharma, V.K. ToxiM: A Toxicity Prediction Tool for Small Molecules Developed Using Machine Learning and Chemoinformatics Approaches. *Front. Pharmacol.* **2017**, *8*, 880. [CrossRef] [PubMed]

92. Cherkasov, A. Inductive QSAR Descriptors. Distinguishing Compounds with Antibacterial Activity by Artificial Neural Networks. *Int. J. Mol. Sci.* **2005**, *6*, 63–86. [CrossRef]

93. Chavan, S.; Friedman, R.; Nicholls, I.A. Acute Toxicity-Supported Chronic Toxicity Prediction: A k-Nearest Neighbor Coupled Read-Across Strategy. *Int. J. Mol. Sci.* **2015**, *16*, 11659–11677. [CrossRef] [PubMed]

94. Sunghwan, K.; Thiessen, P.A.; Bolton, E.E.; Jie, C.; Gang, F.; Asta, G.; Han, L.; He, J.; He, S.; Shoemaker, B.A. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.

95. Fonger, G.C. Hazardous substances data bank (HSDB) as a source of environmental fate information on chemicals. *Toxicology* **1995**, *103*, 137–145. [CrossRef]

96. Fonger, G.C.; Hakkinen, P.; Jordan, S.; Publicker, S. The National Library of Medicine's (NLM) Hazardous Substances Data Bank (HSDB): Background, Recent Enhancements and Future Plans. *Toxicology* **2014**, *325*, 209–216. [CrossRef] [PubMed]

97. Fonger, G.C.; Stroup, D.; Thomas, P.L.; Wexler, P. TOXNET: A computerized collection of toxicological and environmental health information. *Toxicol. Ind. Health* **2000**, *16*, 4–6. [CrossRef] [PubMed]

98. Kavlock, R.; Chandler, K.; Houck, K.; Hunter, S.; Judson, R.; Kleinstreuer, N.; Knudsen, T.; Martin, M.; Padilla, S.; Reif, D. Update on EPA's ToxCast Program: Providing High Throughput Decision Support Tools for Chemical Risk Management. *Chem. Res. Toxicol.* **2012**, *25*, 1287–1302. [CrossRef] [PubMed]

99. 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*, 756–765. [CrossRef] [PubMed]

100. National Toxicology Program. *A National Toxicology Program for the 21st Century: A Roadmap for the Future*; National Toxicology Program: Research Triangle Park, NC, USA, 2004.

101. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2017**, *46*, D1074–D1082. [CrossRef] [PubMed]

102. Kohonen, P.; Benfenati, E.; Bower, D.; Ceder, R.; Crump, M.; Cross, K.; Grafstrom, R.C.; Healy, L.; Helma, C.; Jeliazkova, N.; et al. The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing. *Mol. Inform.* **2013**, *32*, 47–63. [CrossRef] [PubMed]

103. U.S. Environmental Protection Agency. *ECOTOX User Guide: ECOTOXicology Knowledgebase System, version 4.0*; U.S. Environmental Protection Agency: Washington, DC, USA, 2018.

104. Schmidt, U.; Struck, S.; Gruening, B.; Hossbach, J.; Jaeger, I.S.; Parol, R.; Lindequist, U.; Teuscher, E.; Preissner, R. SuperToxic: A comprehensive database of toxic compounds. *Nucleic Acids Res* **2009**, *37*, D295–D299. [CrossRef] [PubMed]

105. 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. The Tox21 robotic platform for the assessment of environmental chemicals—From vision to reality. *Drug Discov. Today* **2013**, *18*, 716–723. [CrossRef] [PubMed]

106. Hansch, C. Quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.* **1969**, *2*, 232–239. [CrossRef]

107. Bradbury, S.P. Predicting modes of toxic action from chemical structure: An overview. *SAR QSAR Environ. Res.* **1994**, *2*, 89–104. [CrossRef] [PubMed]

108. Cronin, M.T.D.; Dearden, J.C. QSAR in Toxicology. 1. Prediction of Aquatic Toxicity. *QSAR Comb. Sci.* **2010**, *14*, 1–7. [CrossRef]

109. Dunn, W.J., III. QSAR approaches to predicting toxicity. *Toxicol. Lett.* **1988**, *43*, 277–283.

110. Kumar, R.S.; Anitha, Y. An Efficient Approach for Asymmetric Data Classification. *Int. J. Innov. Res. Adv. Eng.* **2014**, *1*, 157–161.

111. Yi, L.M.; Hong, G.; Feldkamp, L.A. Neural Learning from Unbalanced Data. *Appl. Intell.* **2004**, *21*, 117–128.

112. Chen, C.; Breiman, L. *Using Random Forest to Learn Imbalanced Data*; University of California: Berkeley, CA, USA, 2004.

113. Wang, S.; Liu, W.; Wu, J.; Cao, L.; Meng, Q.; Kennedy, P.J. Training deep neural networks on imbalanced data sets. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 4368–4374.

114. Myint, K.Z.; Wang, L.; Tong, Q.; Xie, X.Q. Molecular fingerprint-based artificial neural networks QSAR for ligand biological activity predictions. *Mol. Pharm.* **2012**, *9*, 2912–2923. [CrossRef] [PubMed]

115. Myint, K.Z.; Xie, X.Q. Ligand biological activity predictions using fingerprint-based artificial neural networks (FANN-QSAR). *Methods Mol. Biol.* **2015**, *1260*, 149–164. [PubMed]

116. Dahl, G.E.; Jaitly, N.; Salakhutdinov, R. Multi-task Neural Networks for QSAR Predictions. *arXiv* **2014**, arXiv:1406.1231.

117. Lee, K.; Lee, M.; Kim, D. Utilizing random Forest QSAR models with optimized parameters for target identification and its application to target-fishing server. *BMC Bioinform.* **2017**, *18* (Suppl. 16), 567. [CrossRef] [PubMed]

118. Wu, K.; Wei, G.W. Quantitative toxicity prediction using topology based multi-task deep neural networks. *J. Chem. Inform. Model.* **2018**, *58*, 520–531. [CrossRef] [PubMed]

119. Capuzzi, S.J.; Politi, R.; Isayev, O.; Farag, S.; Tropsha, A. QSAR Modeling of Tox21 Challenge Stress Response and Nuclear Receptor Signaling Toxicity Assays. *Front. Environ. Sci.* **2016**. [CrossRef]

120. Kearnes, S.; Mccloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: Moving beyond fingerprints. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 1–14. [CrossRef] [PubMed]

121. Binetti, R.; Costamagna, F.M.; Marcello, I. Exponential growth of new chemicals and evolution of information relevant to risk control. *Ann. dell'Istituto Super. di Sanita* **2008**, *44*, 13–15.

122. Trevan, J.W. The Error of Determination of Toxicity. *Proc. R. Soc. Lond.* **1927**, *101*, 483–514. [CrossRef]

123. Gute, B.D.; Basak, S.C. Predicting acute toxicity (LC50) of benzene derivatives using theoretical molecular descriptors: A hierarchical QSAR approach. *SAR QSAR Environ. Res.* **1997**, *7*, 117–131. [CrossRef] [PubMed]

124. Basak, S.C.; Grunwald, G.D.; Gute, B.D.; Balasubramanian, K.; Opitz, D. Use of statistical and neural net approaches in predicting toxicity of chemicals. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 885–890. [CrossRef] [PubMed]

125. Martin, T.M.; Lilavois, C.R.; Barron, M.G. Prediction of pesticide acute toxicity using two-dimensional chemical descriptors and target species classification. *SAR QSAR Environ. Res.* **2017**, *28*, 1–15. [CrossRef] [PubMed]

126. Liu, R.; Madore, M.; Glover, K.P.; Feasel, M.G.; Wallqvist, A. Assessing Deep and Shallow Learning Methods for Quantitative Prediction of Acute Chemical Toxicity. *Toxicol. Sci.* **2018**, *164*, 512–526. [CrossRef] [PubMed]

127. 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*, 2672–2685. [CrossRef] [PubMed]

128. Li, X.; Zhang, Y.; Chen, H.; Li, H.; Zhao, Y. In silico prediction of chronic toxicity with chemical category approaches. *RSC Adv.* **2017**, *7*, 41330–41338. [CrossRef]

129. Liu, J.; Xu, C.; Yang, W.; Shu, Y.; Zheng, W.; Zhou, F. Multiple similarly effective solutions exist for biomedical feature selection and classification problems. *Sci. Rep.* **2017**, *7*, 12830. [CrossRef] [PubMed]

130. Van, D.J.; Gaj, S.; Lienhard, M.; Albrecht, M.W.; Kirpiy, A.; Brauers, K.; Claessen, S.; Lizarraga, D.; Lehrach, H.; Herwig, R. RNA-Seq provides new insights in the transcriptome responses induced by the carcinogen benzo[a]pyrene. *Br. J. Dermatol.* **2012**, *130*, 568–577.

131. Liu, R.; Yu, X.; Wallqvist, A. Using Chemical-Induced Gene Expression in Cultured Human Cells to Predict Chemical Toxicity. *Chem. Res. Toxicol.* **2016**, *29*, 1883. [CrossRef] [PubMed]

132. Schwartz, M.P.; Hou, Z.; Propson, N.E.; Zhang, J.; Engstrom, C.J.; Santos, C.V.; Jiang, P.; Nguyen, B.K.; Bolin, J.M.; Daly, W. Human pluripotent stem cell-derived neural constructs for predicting neural toxicity. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12516–12521. [CrossRef] [PubMed]

133. Yamane, J.; Aburatani, S.; Imanishi, S.; Akanuma, H.; Nagano, R.; Kato, T.; Sone, H.; Ohsako, S.; Fujibuchi, W. Prediction of developmental chemical toxicity based on gene networks of human embryonic stem cells. *Nucleic Acids Res.* **2016**, *44*, 5515–5528. [CrossRef] [PubMed]

134. Ippolito, D.L.; Abdulhameed, M.D.; Tawa, G.J.; Baer, C.E.; Permenter, M.G.; Mcdyre, B.C.; Dennis, W.E.; Boyle, M.H.; Hobbs, C.A.; Streicker, M.A. Gene Expression Patterns Associated With Histopathology in Toxic Liver Fibrosis. *Toxicol. Sci.* **2016**, *149*, 67–88. [CrossRef] [PubMed]

135. Smith, J.B.; Lanitis, E.; Dangaj, D.; Buza, E.; Poussin, M.; Stashwick, C.; Scholler, N.; Powell, D.J. Tumor Regression and Delayed Onset Toxicity Following B7-H4 CAR T Cell Therapy. *Mol. Therapy J. Am. Soc. Gene Therapy* **2016**, *24*, 1987. [CrossRef] [PubMed]

136. Zhang, J.D.; Berntenis, N.; Roth, A.; Ebeling, M. Data mining reveals a network of early-response genes as a consensus signature of drug-induced in vitro and in vivo toxicity. *Pharmacogenomics J.* **2014**, *14*, 208–216. [CrossRef] [PubMed]

137. Isik, Z.; Baldow, C.; Cannistraci, C.V.; Schroeder, M. Drug target prioritization by perturbed gene expression and network information. *Sci. Rep.* **2015**, *5*, 17417. [CrossRef] [PubMed]

138. Kotlyar, M.; Fortney, K.; Jurisica, I. Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods* **2012**, *57*, 499–507. [CrossRef] [PubMed]

139. Liu, R.; Abdulhameed, M.D.M.; Wallqvist, A. Molecular Structure-Based Large-Scale Prediction of Chemical-Induced Gene Expression Changes. *J. Chem. Inform. Model.* **2017**, *57*, 2194–2201. [CrossRef] [PubMed]

140. Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.P.; Subramanian, A.; Ross, K.N. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**, *313*, 1929–1935. [CrossRef] [PubMed]

141. 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. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2011**, *39*, 1005–1010. [CrossRef] [PubMed]

142. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [CrossRef] [PubMed]

143. Yoo, M.; Shin, J.; Kim, J.; Ryall, K.A.; Lee, K.; Lee, S.; Jeon, M.; Kang, J.; Tan, A.C. DSigDB: Drug signatures database for gene set analysis. *Bioinformatics* **2015**, *31*, 3069–3071. [CrossRef] [PubMed]

144. Duan, Q.; Flynn, C.; Niepel, M.; Hafner, M.; Muhlich, J.L.; Fernandez, N.F.; Rouillard, A.D.; Tan, C.M.; Chen, E.Y.; Golub, T.R. LINCS Canvas Browser: Interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res.* **2014**, *42*, W449. [CrossRef] [PubMed]

145. Li, Y.H.; Yu, C.Y.; Li, X.X.; Zhang, P.; Tang, J.; Yang, Q.; Fu, T.; Zhang, X.; Cui, X.; Tu, G. Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* **2017**, *46*, D1121–D1127.

146. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; King, B.L.; McMorran, R.; Wiegers, J.; Wiegers, T.C.; Mattingly, C.J. The Comparative Toxicogenomics Database: Update 2017. *Nucleic Acids Res* **2017**, *45*, D972–D978. [CrossRef] [PubMed]

147. Zeng, H.; Qiu, C.; Cui, Q. Drug-Path: A database for drug-induced pathways. *Database* **2015**, *2015*, bav061. [CrossRef] [PubMed]

148. Kumar, R.; Chaudhary, K.; Gupta, S.; Singh, H.; Kumar, S.; Gautam, A.; Kapoor, P.; Raghava, G.P.S. CancerDR: Cancer Drug Resistance Database. *Sci. Rep.* **2013**, *3*, 1445. [CrossRef] [PubMed]

149. 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*, 355–360. [CrossRef] [PubMed]

150. Du, J.; Jia, P.; Dai, Y.; Tao, C.; Zhao, Z.; Zhi, D. Gene2Vec: Distributed Representation of Genes Based on Co-Expression. *bioRxiv* **2018**. [CrossRef]

151. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *Comput. Sci. arXiv* **2013**, arXiv:1301.3781.

152. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 3111–3119.

153. Duong, D.; Eskin, E.; Li, J. A novel Word2vec based tool to estimate semantic similarity of genes by using Gene Ontology terms. *bioRxiv* **2017**. [CrossRef]

154. Danaee, P.; Ghaeini, R.; Hendrix, D.A. A Deep Learning Approach For Cancer Detection and Relevant Gene Identification. *Pac. Symp. Biocomput.* **2016**, *22*, 219–229.

155. Sharifi-Noghabi, H.; Liu, Y.; Erho, N.; Shrestha, R.; Alshalalfa, M.; Davicioni, E.; Collins, C.C.; Ester, M. Deep Genomic Signature for early metastasis prediction in prostate cancer. *bioRxiv* **2018**. [CrossRef]

156. Aliper, A.; Plis, S.; Artemov, A.; Ulloa, A.; Mamoshina, P.; Zhavoronkov, A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* **2016**, *13*, 2524–2530. [CrossRef] [PubMed]

157. Gayvert, K.M.; Madhukar, N.S.; Elemento, O. A Data-Driven Approach to Predicting Successes and Failures of Clinical Trials. *Cell Chem. Biol.* **2016**, *23*, 1294–1301. [CrossRef] [PubMed]

158. Zhen, X.; Chen, J.; Zhong, Z.; Hrycushko, B.A.; Zhou, L.; Jiang, S.B.; Albuquerque, K.; Gu, X. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study. *Phys. Med. Biol.* **2017**, *62*, 8246–8263. [CrossRef] [PubMed]

159. Zhu, H.; Zhang, J.; Kim, M.T.; Boison, A.; Sedykh, A.; Moran, K. Big data in chemical toxity research: The use of high-throughput screening assays to identify potential toxicants. *Chem. Res. Toxicol.* **2014**, *27*, 1643–1651. [CrossRef] [PubMed]

160. Pasturromay, L.A.; Cedrón, F.; Pazos, A.; Portopazos, A.B. Deep Artificial Neural Networks and Neuromorphic Chips for Big Data Analysis: Pharmaceutical and Bioinformatics Applications. *Int. J. Mol. Sci.* **2016**, *17*, 1313. [CrossRef] [PubMed]