JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

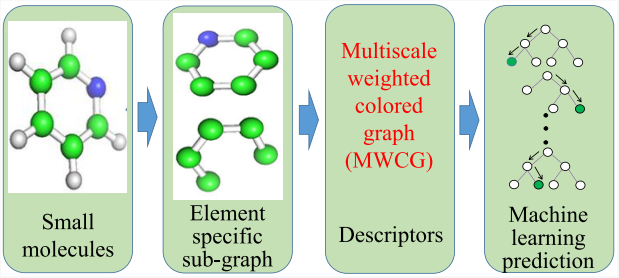
GGL-Tox: Geometric Graph Learning for Toxicity Prediction

Jian Jiang, Rui Wang, and Guo-Wei Wei*

Cite This: J. Chem. Inf. Model. 2021, 61, 1691−1700 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Toxicity analysis is a major challenge in drug design and discovery. Recently significant progress has been made through machine learning due to its accuracy, efficiency, and lower cost. US Toxicology in the 21st Century (Tox21) screened a large library of compounds, including approximately 12 000 environmental chemicals and drugs, for different mechanisms responsible for eliciting toxic effects. The Tox21 Data Challenge offered a platform to evaluate different computational methods for toxicity predictions. Inspired by the success of multiscale weighted colored graph (MWCG) theory in protein−ligand binding affinity predictions, we consider MWCG theory for toxicity analysis. In the present work, we develop a geometric graph learning toxicity (GGL-Tox) model by integrating MWCG features and the gradient boosting decision tree (GBDT) algorithm. The benchmark tests of the Tox21 Data Challenge are employed to demonstrate the utility and usefulness of the proposed GGL-Tox model. An extensive comparison with other state-of-the-art models indicates that GGL-Tox is an accurate and efficient model for toxicity analysis and prediction.

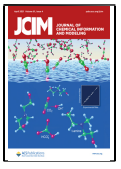


INTRODUCTION

Chemical toxicity is an important measure in environmental, agricultural, and pharmaceutical sciences. In pharmacology, toxicity plays a crucial role in drug discovery and is the major factor for disqualifying most drug candidates. For example, a potential drug for cancer treatment must be studied for its activity at multiple biological targets, including possibly novel targets, rendering a probability of multiple toxicological profiles.[1] Therefore, it is highly desirable to develop novel methods that can determine the fate of chemical compounds, in order to decrease the failure rates in the early stage of drug design and accelerate the approval of promising drug candidates. The traditional paradigm in toxicity testing incorporates *in vivo* animal studies and *in vitro* techniques, which is laborious, expensive, and often impractical for evaluating large numbers of compounds. This approach has been gradually phased out owing to its controversial nature.[2] As a result, *in silico* methods are in great demand for the accurate prediction of toxicity and enable the prioritization of drug candidates for experimental testing. These *in silico* methods typically utilize experimental data generated by *in vivo* and *in vitro* screening technologies and lead to powerful predictive models, which could be used to screen thousands of chemicals for potential unwanted side effects early on during development cycles or to re-evaluate existing ones. Computational approaches are used to utilize limited experimental resources efficiently.

Due to the availability of abundant experimental data, machine learning (ML) algorithms have been widely used in toxicity prediction,[3−5] including k-nearest neighbors (KNN),[6,7] support vectors machine (SVM),[8−10] random forest (RF),[11,12] and many others.[13−16] The traditional machine learning techniques depend heavily on the quantity and quality of training data and domain knowledge-based feature engineering. For example, nonlinear SVM can be capable of dealing with high-dimensional data but may not be robust to the presence of diverse chemical descriptors.[17] Deng and Zhao[18] reported that the computational cost of KNN increases exponentially with the size of the input samples. Recently, deep learning (DL) has attracted much attention for predicting the outcome of biological assays and becomes a key candidate for toxicity prediction due to its ability to bypass feature extraction. Mayr et al. developed the DeepTox pipeline using deep neural networks (DNNs) to study toxicology in the 21st Century (Tox21) Data Challenge 10k library data sets and found that DL outperformed other computational approaches like naive Bayes, SVM, and RF.[14,19] Molecular and biomolecular data sets involve structural complexity, which makes ML performance highly dependent on structural representations.[20] By using a molecular graph encoding convolutional neural network (MGE-CNN) architecture, Xu et al. constructed deepAOT (DL-based acute oral toxicity) models for both quantitative toxicity prediction and toxicant

Received: November 8, 2020
Published: March 15, 2021



https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700

category classification.[21] In addition, Wu and Wei introduced an algebraic topology-based approach that combines multitask DNN and element-specific persistent homology (ESPH) for quantitative toxicity prediction using four benchmark ecotoxicity data sets.[22] More references about molecule structural representation and toxicity prediction can be found in the literature.[23−26]

Graph theories have been widely applied to problems in the biological, physical, social, chemical, and computer sciences. Pairwise relations in reality can be easily represented and analyzed by graphs. For instance, in chemistry and biology, a graph can model the structure of a molecule, where graph vertices indicate atoms and graph edges indicate possible bonds. Graphs have widespread applications in chemical analysis[27] and macromolecular modeling,[28] such as normal-mode analysis (NMA)[29] and elastic network models (ENMs)[30] for modeling protein flexibility and long time dynamics. In particular, graphs bridge the gap between the toxicity of chemical compounds and their structure and functional relationships. The utility of graph theory makes it a popular approach not only for toxicity prediction but also for describing chemical data sets,[31,32] biomolecular data sets,[33,34] protein thermal fluctuations,[35] protein−ligand binding affinity,[36,37] deep learning,[38] and chemical molecule design.[39] Recently, a new graph theory, multiscale weighted colored graph (MWCG), has been proposed for protein flexibility analysis[35] and protein−ligand binding prediction.[36,37] Mathematical properties of MWCGs include low-dimensionality, simplicity, robustness, and invariance of rotations, translations, and reflections. The molecular modeling of MWCGs requires only atomic names and coordinates. Paired with machine learning algorithms, MWCGs were shown to outperform other approaches in the D3R Grand Challenges, a worldwide competition series in computer-aided drug design.[20,40] However, the potential of MWCGs for small molecular property analysis, such as small molecular toxicity prediction, remains unknown.

The objective of the present work is to understand the utility and performance of MWCGs for small molecular representation and modeling. To this end, we consider the Tox21 10k library (12 data sets) as a benchmark to test MWCGs' performance. Many alternative molecular representations, particularly popular two-dimensional (2D) fingerprints,[49−51] are employed to calibrate MWCGs. Additionally, to understand the robustness of MWCGs and other 2D fingerprints, we consider three simple machine learning algorithms, namely, SVM, RF, and gradient boosting decision tree (GBDT), to build a variety of predictive models. Among them, the combination of MWCGs and GBDT, denoted as the geometric graph learning toxicity (GGL-Tox) model, outperforms other models we have examined for the Tox21 10k library. Finally, we compare the performance of GGL-Tox with that of other well-established models in the literature. Our results indicate that GGL-Tox achieves the state-of-the-art in toxicity prediction and classification.

MATERIALS AND METHODS

Tox21 Data Challenge. The Tox21 project,[41−43] started in 2008, is a multiagency collaborative consortium, constituted of the National Institutes of Health (NIH), the Environmental Protection Agency (EPA), the National Toxicity Program (NTP), National Center for Advancing Translational Sciences (NCATS), and the Food and Drug Administration (FDA). The goal is to develop fast and effective approaches for large-scale assessment of toxicity in order to identify chemicals that could be potentially toxic and impair various human biological pathways. This consortium makes use of its combined resources and expertise to predict more effectively how a collection of around 12 000 compounds composed of environmental chemicals and approved drugs will affect human health and the environment. The Tox21 10k library has already been employed in a high-throughput screening (HTS) against a panel of nuclear receptor (NR)[44−46] and stress response (SR) pathway assays.[47] NCATS launched Tox21 Data Challenge 2014 with these publicly available data sets to enlist independent participants to unveil how well they can predict the effects of compounds in cellular and biochemical pathways, causing potential harm. The origin and sample introduction of the Tox21 Data Challenge can be found in section S2 of the Supporting Information.

Multiscale Weighted Colored Graph Theory. A graph can be used to describe the noncovalent interaction of N atoms in proteins with a set of vertices $\mathbb{V}$ and links or edges among them, denoted as $\mathbb{E}$. To improve the graph theory representation, the colored graph has attracted much attention in which different types of elements are labeled. Labeled protein atoms are classified into subgraphs where colored edges correspond to element-specific interactions.

In the present work, we focus on pairwise noncovalent interactions in our subgraph theory description. For a given data set, we first identify statistically a set of commonly occurring chemical element types, say $\mathbb{C} = \{\mathrm{H}, \mathrm{C}, \mathrm{N}, \mathrm{O}, \mathrm{S}, \mathrm{P}, \mathrm{F}, \mathrm{Cl}, \mathrm{Br}, \cdots\}$. For a given molecule or biomolecule in the data set, we denote

$$\mathbb{V} = \{(\mathbf{r}_j, \alpha_j) | \mathbf{r}_j \in \mathbb{R}^3; \quad \alpha_j \in \mathbb{C}; \quad j = 1, 2, ..., N\} \tag{1}$$

as a subset of N atoms (i.e., subgraph vertices) that are members of $\mathbb{C}$. Note that the ith atom is labeled both by its element type α_j and its position $\mathbf{r}_j$. The classification of atoms into chemical element types is a graph coloring, which is important for encoding different types of interactions and gives rise to a basis for the collective coarse-grained description of the data set. We assume that all the pairwise noncovalent interactions between element types $\mathbb{C}_k$ and $\mathbb{C}_{k'}$ in a molecule or molecular complex can be represented by fast-decay weight functions

$$\mathbb{E} = \{\Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \ \eta_{kk'}) | \alpha_i = \mathbb{C}_k, \alpha_j = \mathbb{C}_{k'}; \\ i, j = 1, 2, ..., N; \quad \|\mathbf{r}_i - \mathbf{r}_j\| > r_i + r_j + \sigma\} \tag{2}$$

where $\|\mathbf{r}_i - \mathbf{r}_j\|$ is the Euclidean distance between the ith and jth atoms, r_i and r_j are the atomic radii of ith and jth atoms, respectively, and σ is the mean value of the standard deviations of r_i and r_j in the data set. The distance constraint ($\|\mathbf{r}_i - \mathbf{r}_j\| > r_i + r_j + \sigma$) excludes covalent interactions. Here $\eta_{kk'}$ is a characteristic distance between the atoms, and Φ is a subgraph weight and is chosen to have the following properties:[48]

$$\Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \ \eta_{kk'}) = 1, \\ \text{as } \|\mathbf{r}_i - \mathbf{r}_j\| \to 0 \text{ and } \Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \ \eta_{kk'}) = 0, \\ \text{as } \|\mathbf{r}_i - \mathbf{r}_j\| \to \infty, \ \alpha_i = \mathbb{C}_k, \ \alpha_j = \mathbb{C}_k. \tag{3}$$

Although most radial basis functions can be used, generalized exponential functions and generalized Lorentz

https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700

functions were shown to work very well for biomolecules.[48] We, therefore, have a weighted colored subgraph $G(\mathbb{V}, \mathbb{E})$. To construct element-level collective molecular descriptors, we propose the multiscale weighted colored subgraph rigidity representation (RR) between kth element type $\mathbb{C}_k$ and k'th element type $\mathbb{C}_{k'}$

$$\mathrm{RR}^G(\eta_{kk'}) = \sum_i \mu_i^G(\eta_{kk'}) = \sum_i \sum_j \Phi(\|\mathbf{r}_i - \mathbf{r}_j\|;\ \eta_{kk'}),$$
$$\alpha_i = \mathbb{C}_k,\ \alpha_j = \mathbb{C}_{k'};\quad \|\mathbf{r}_i - \mathbf{r}_j\| > r_i + r_j + \sigma \tag{4}$$

where $\mu_i^G(\eta_{kk'})$ is a geometric subgraph centrality for the ith atom. The physical interpretation of eq 4 is straightforward—the summation over $\mu_i^G(\eta_{kk'})$ in eq 4 leads to the total interaction strength for the selected pair of element types $\mathbb{C}_k$ and $\mathbb{C}_{k'}$, which provides the element-level coarse-grained description of molecular-level properties. Additionally, the above formulation is a generalization of the successful bipartite subgraph used in our earlier predictions of protein−ligand binding affinities and free energy ranking.[37] For a bipartite subgraph, each of its edge connects one atom in the protein and another atom in the ligand. However, in the present work, the subgraph is undirected, and two atoms connected by edges belong to the same molecule. An illustration of the weighted colored subgraph G_{NO} of the uracil molecule ($C_4H_4N_2O_2$) is given in Figure 1.

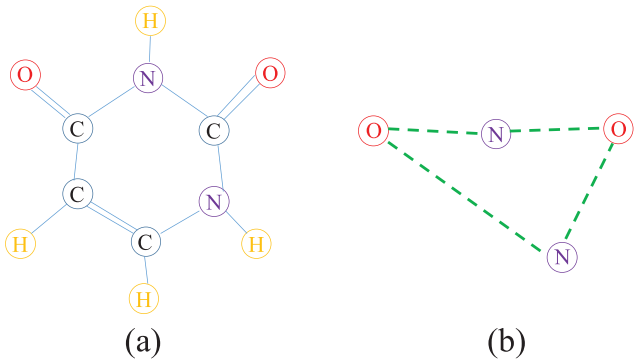


Figure 1. (a) Diagram of the uracil molecular ($C_4H_4N_2O_2$) structure. (b) One weighted colored subgraph G_{NO} consisting of N and O atoms. The edges are labeled by green-dashed lines which are not covalent bonds.

The different selections of characteristic distance $\eta_{kk'}$ give rise to a multiscale description of intermolecular and intramolecular interactions. By appropriate selections of element combinations k and k', the characteristic distance $\eta_{kk'}$, and subgraph weight Φ, we systematically construct a family of collective, scalable, multiscale graph-based molecular and biomolecular descriptors. The proposed multiscale weighted colored subgraph rigidity representation is robust and straightforward—the only required data inputs are atomic names and coordinates. Consequently, our graph approach is very fast. Our fast algorithm has the computational complexity of $O(N)$ and is able to predict B-factors for α-carbons of an HIV virus capsid (313 236 residues) in less than 30 s on a single processor.[48]

In present work, we propose three-scale MWCGs to capture multiscale interactions with each element where three kernels, one exponential kernel and two Lorentz kernels at the same time are considered and construct three sets of feature vectors. Generally, we represent these feature vectors by $\mathrm{RR}^{\alpha_1\alpha_2\alpha_3}_{\beta_1,\eta_1;\beta_2,\eta_2;\beta_3,\eta_3}$ as a straightforward extension of our notation. $\alpha_i = E, L$ (i = 1, 2, 3) is a kernel index indicating either the exponential kernel (E) or Lorentz kernel (L). Correspondingly, β_i is kernel order index such that $\beta_i = \kappa$ when $\alpha_i = E$ and $\beta_i = \nu$ when $\alpha_i = L$. The details of parameters in the three-scale model for Tox21 Data Challenge 2014 can be found in Table_S2 in the Supporting Information.

Since all the chemical compounds in Tox21 Data Challenge 2014 have 53 different types of atoms in total (see Table_S3 in the Supporting Information), the three-scale models consist of 1470 descriptors, where each kernel α_i (i = 1, 2, 3) is used to generate 490 descriptors, presenting cross-correlation between atoms.

2D Molecular Fingerprints. In chemoinformatic studies, molecular representations commonly used in traditional ML-based models are molecular fingerprints,[49−51] physicochemical properties, topological properties, and thermo-dynamics properties.[52] As the property profile of a molecule, the molecular fingerprint plays a crucial role in quantitative structure−activity/property relationship (QSAR/QSPR) analysis. There are four types of 2D molecular fingerprints, namely keys-based fingerprints, pharmacophore fingerprints, topological or path-based fingerprints, and circular fingerprints.[53,54] In the present work, we consider four popular 2D fingerprints, namely molecular access system (MACCS) fingerprint,[55] Estate 1 (electro-topological state) fingerprint, Estate 2 fingerprint,[56] and Morgan fingerprint with radius 2 hashed to 1024 bits length,[50] which are generated by RDKit (version 2018.09.3).[57] MACCS fingerprints are designed on generic substructure keys,[55] and Estate fingerprints can be used to examine whether molecular fragments based on the electronic, topological, and valence state indices of atom types can be useful in prediction of toxic activity.[56] Table_S4 in the Supporting Information summarizes the essential information related to these fingerprints.

Gradient Boosting Decision Tree (GBDT). GBDT is a widely used ensemble algorithm of decision trees that assembles a number of so-called weak "learners" into a prediction model iteratively. As GBDTs are generally robust to outliers and have strong predictive power, which can handle heterogeneous features, they have already achieved good performances in many different applications, such as multiclass classification,[58] learning to rank,[59] and click prediction.[60] In this method, individual decision trees are trained sequentially and are assembled in a stagewise fashion to boost their capability of learning complex features. In general, based on N consecutive decision trees, the prediction of the model with data $\{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^M$ (M is the number of samples) is as follows:

$$\hat{y}_N(\mathbf{x}) = \sum_{n=1}^{N} p_n(\mathbf{x}), \tag{5}$$

where $p_n(\mathbf{x})$ is the predicted labels of the nth tree. Taking regression as an example, a general loss function is given by

$$L_n = \sum_i l_i(y^{(i)}, \hat{y}_n^{(i)}), \tag{6}$$

where $l_i = (y^{(i)} - \hat{y}_n^{(i)})^2/2$ with a square loss taken into consideration. In each iteration, GBDT learns the decision trees by fitting the negative gradients. The total loss function L can be minimized along the following gradient direction:

https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700

Table 1. Comparison of Prediction Results of GBDT with Five Different Molecular Fingerprints in Tox21 Prediction[a]

metrics	fingerprints	1	2	3	4	5	6	7	8	9	10	11	12
ACC	Estate 2	0.882	0.977	0.986	0.926	0.905	0.964	0.946	0.824	0.941	0.963	0.896	0.922
	Estate 1	0.886	0.979	0.986	0.926	0.909	0.962	0.944	0.820	0.941	0.963	0.913	0.933
	MACCS	0.870	0.979	0.982	0.926	0.905	0.964	0.947	0.824	0.942	0.965	0.898	0.935
	MWCG	0.914	0.993	0.995	0.950	0.929	0.986	0.968	0.856	0.962	0.985	0.926	0.963
	hybrid	0.926	0.997	0.991	0.948	0.923	0.986	0.971	0.844	0.959	0.983	0.919	0.955
AUC	Estate 2	0.822	0.706	0.720	0.698	0.739	0.793	0.735	0.714	0.704	0.757	0.876	0.739
	Estate 1	0.874	0.680	0.688	0.619	0.746	0.800	0.726	0.707	0.718	0.729	0.920	0.747
	MACCS	0.865	0.668	0.759	0.760	0.783	0.771	0.737	0.743	0.797	0.828	0.909	0.790
	MWCG	0.887	0.915	0.991	0.834	0.836	0.892	0.773	0.803	0.833	0.978	0.931	0.810
	hybrid	0.884	0.768	0.777	0.752	0.801	0.870	0.695	0.769	0.718	0.967	0.926	0.829
BA	Estate 2	0.628	0.532	0.549	0.589	0.538	0.576	0.544	0.544	0.526	0.543	0.638	0.505
	Estate 1	0.625	0.582	0.537	0.558	0.557	0.523	0.586	0.525	0.526	0.522	0.690	0.512
	MACCS	0.625	0.571	0.524	0.512	0.601	0.524	0.573	0.556	0.540	0.566	0.712	0.536
	MWCG	0.629	0.612	0.599	0.605	0.617	0.663	0.632	0.611	0.634	0.569	0.779	0.610
	hybrid	0.736	0.601	0.588	0.599	0.613	0.663	0.606	0.607	0.541	0.636	0.721	0.535
MCC	Estate 2	0.319	0.211	0.243	0.236	0.208	0.268	0.255	0.158	0.178	0.177	0.364	0.022
	Estate 1	0.329	0.327	0.301	0.298	0.256	0.088	0.251	0.101	0.178	0.142	0.468	0.198
	MACCS	0.323	0.405	0.321	0.341	0.318	0.102	0.356	0.178	0.235	0.275	0.444	0.296
	MWCG	0.373	0.455	0.356	0.411	0.375	0.326	0.467	0.300	0.332	0.182	0.549	0.426
	hybrid	0.535	0.423	0.332	0.400	0.334	0.326	0.429	0.273	0.173	0.277	0.475	0.180

[a]The performance of the models are evaluated by the accuracy (ACC), ROC-curve (AUC), balanced accuracy (BA), and Matthews correlation coefficient (MCC). The numbers from 1 to 12 in the first row correspond to Tox21 datasets of NR-AhR, NR-AR, NR-AR-LBD, NR-Aromatase, NR-ER, NR-ER-LBD, NR-PPAR-gamma, SR-ARE, SR-ATAD5, SR-HSE, SR-MMP, and SR-p53, respectively.

$$-\frac{\partial L_n}{\partial p_n(\mathbf{x}^{(i)})} = y^{(i)} - \hat{y}_n^{(i)}. \tag{7}$$

The main procedure of GBDT is the learning of decision trees, which costs most of the time to find the best split spot. Compared to another popular algorithm in toxicity prediction, i.e., deep neural networks (DNN), GBDT is robust, relatively insensitive to hyperparameters, more suitable for dealing with small data sets, and easy to implement. Additionally, it is faster to train than DNN, which is a major advantage of GBDT. However, it is worth noting that one of the challenges of GBDT is how to balance the trade-off between accuracy and efficiency for big data sets.

Random Forest (RF). In order to explore the advantages and weaknesses of different ML algorithms, another popular ensemble method, namely, random forest (RF) is considered, which has been widely used in solving QSAR prediction problems and usually does not require a complicated feature selection procedure. Particularly, it is insensitive to parameters and robust to redundant features.

RF, developed by Breiman, is a collection of decision trees whose prediction is averaged to get an ensemble performance.[61] Each individual tree, drawn upside down, uses only a subset of samples and features which are both chosen randomly and begin with a trunk that splits into multiple branches before eventually arriving at the leaves. The leaf nodes mean the end point to be predicted, while all other nodes are assigned a molecular feature. In order to construct a robust decision tree, the features that optimally separate the end points are chosen. Optimal features are selected based on the information gain criterion or the Gini coefficient. The main hyperparameters for RF are the number of trees, the number of features considered in each step, the number of samples, the feature choice, and the feature type.

Support Vector Machine (SVM). SVM, developed by Cortes and Vapnik, is a nonprobabilistic kernel-based supervised learning method that maps input vectors into high-dimensional feature space where the decision hyperplane is constructed.[62] Alvarsson et al. introduced the package of LibSVM with a radial basis function kernel to develop the classification models with hyperparameters optimized within the predefined ranges.[63]

RESULTS

In this section, we will present the performance of machine learning models on the Tox21 Challenge Data sets based on GBDT, RF, and SVM algorithms with different molecular representations. We use the area under the ROC curve (AUC), accuracy (ACC), balanced accuracy (BA), and Matthews correlation coefficient (MCC) to evaluate the performance of various models.[64] AUC takes values between (0, 1). A perfect model would have an AUC of 1, and an AUC of 0.5 means a random classifier. ACC is defined as the number of compounds predicted correctly over the total number of compounds. BA is the average of correct prediction for both active and inactive classes, and the MCC value varies from −1 to 1, with −1 implying anticorrelation, 1 implying a perfect classifier, and 0 representing no correlation between the prediction and the known truth. For all experiments in the present study, GBDT, RF, and SVM are implemented by the Scikit-learn package (version 0.20.1).[65]

Table 1 shows the values of ACC, AUC, BA, and MCC with five molecular fingerprints, Estate 1, Estate 2, MACCS, MWCG, as well as hybrid, which is the combination of these four fingerprints, combined with GBDT for toxicity prediction of Tox21 Challenge data sets. Here, the three-scale GGL-Tox model, $RR^{\alpha_1\alpha_2\alpha_3}_{\beta_1,\eta_1,\beta_2,\eta_2,\beta_3,\eta_3}$, is constructed with $\alpha_1 = E$ and $\alpha_2 = \alpha_3 = L$. The hyperparameters of the GBDT algorithm are given in Table S5 in the Supporting Information. Note that the hyperparameters of different machine learning algorithms provided in the Supporting Information are selected based on all training data sets with optimal performance using 5-fold

https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700

cross-validation (CV) and not on test data sets, and the test data is used for the comparison of results. From this table, we can find that in all 12 assays, the MWCG-based GBDT model (i.e., the GGL-Tox model) achieves the highest AUC values. The performance of test data and CV data for all 12 assays with the GGL-Tox model can be found in Table S6 in the Supporting Information. Additionally, the AUC values of the hybrid fingerprint are obtained by combining MGCG and three 2D fingerprints, i.e., Estate 1, Estate 2, and MACCS in 9 assays, except the assays of NR-Aromatase, NR-ppar-gamma, and SR-ATAD5. These results suggest that the MWCG-based GBDT model has the best performance in the toxicity prediction. The behavior of the hybrid fingerprint is the second best, and 2D fingerprints perform the worst. Additionally, we consider another 2D fingerprint, the Morgan fingerprint with radius 2 hashed to 1024 bits in the prediction, which performs similarly with aforementioned fingerprints but is not as good as MWCG descriptors. The details of prediction results and hyperparameters with Morgan fingerprint for 12 data sets can be found in Tables S7 and S8 of the Supporting Information, respectively.

Feature Importance Analysis. From the results of the above section, we find that MWCG fingerprint could achieve much better predictive performance than classical 2D fingerprints. Moreover, although the hybrid fingerprint has more features than MWCG, it does not perform well. In general, different features play different roles in ML predictions, and redundant or noisy features may result in a negative influence in the training process. In this section, we try to understand the outcomes in terms of feature importance, which refers to Gini importance that is weighted by the number of trees in a forest calculated by our baseline algorithm GBDT with scikit-learn package.[61] To get the optimal number of features, we construct a family of models with top N% features based the ranking of feature importance, where N goes from 0 to 100. Once the AUC reaches the maximum value, its corresponding feature size will be the optimized number of features.

Figure 2 shows the influence of the feature size on the value of AUC of 12 data sets with the GBDT algorithm in toxicity prediction, where the feature size of the hybrid fingerprint is 1794. One can find that this method depends on the number of features, and the value of AUC increases quickly as the numbers of features increase for all 12 data sets. For SR-ATAD5, NR-AR-LBD, NR-PPAR-gamma, NR-AhR, SR-HSE, NR-Aromatase, NR-AR, SR-MMP, and SR-ARE data sets, when the number of features increases up to around 10, 40, 40, 41, 51, 70, 79, 80 (0.6%, 2.2%, 2.2%, 2.3%, 2.8%, 3.9%, 4.4%, 4.5% of 1794 features), respectively, the values of AUC reach their respective maximum value. For the NR-ER, SR-p53, and NR-ER-LBD data sets, the respective maximum values of AUC are obtained with the feature size of around 200, 200, and 500, which are respectively about 11.1%, 11.1% and 27.8% of their total numbers of features. In other words, choosing at least 0.6%, or at most 27.8%, of the most important features could optimize the prediction performance. Additionally, it should be noticed that for all 12 data sets, when the maximum values of AUC are reached, the value of AUC for each data set decreases with the increasing of feature size. These results suggest that, for the hybrid fingerprint, some redundant features may exist with high correlation, which leads to a negative influence on the prediction performance to some extent. The method of ranking feature importance is a more efficient and less time-consuming method for ML.

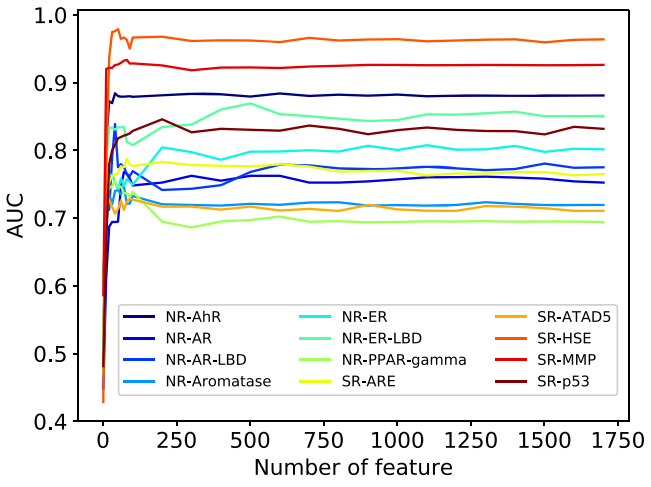


Figure 2. Relationship between AUC value and the number of top features for hybrid fingerprint in 12 data sets obtained using GBDT in toxicity prediction.

Comparison with Other ML Algorithms. In order to compare the predictive results of different ML algorithms, Figure 3 shows the comparison of three algorithms, GBDT, RF, and SVM in association with MWCG representations in toxicity prediction. The values of AUC and ACC of these algorithms and the details of the main hyperparameters of RF and SVM in the implementation can be found in Tables S9–S11 of the Supporting Information, respectively. We find that AUC values with the GBDT algorithm for 11 data sets are larger than those with RF and SVM algorithms, which is obvious and shown with blue in the figure. Only for data set SR-p53 is the AUC value with the GBDT algorithm 6.0% smaller than the maximum AUC value with RF. Additionally, the average AUC values of the NR set are 0.875 for GBDT, 0.831 for RF, and 0.811 for SVM, and the average AUC values of SR set are 0.871 for GBDT, 0.850 for RF, and 0.812 for SVM, respectively. This result indicates that among these three ML algorithms, GBDT performs best, RF is second-best, and SVM is the worst.

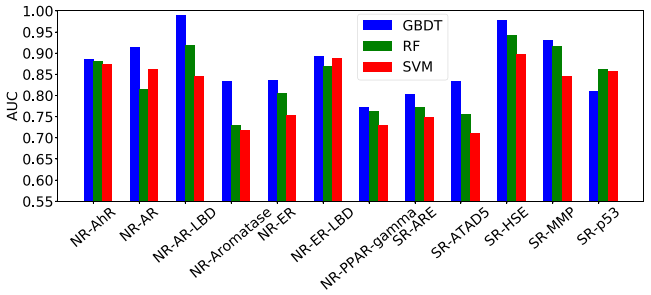


Figure 3. Comparison of prediction results realized by GBDT, RF, and SVM in association with MWCG in toxicity prediction, respectively. AUC values with the GBDT algorithm for 11 data sets are larger than those with RF and SVM algorithms, except for data set SR-p53.

Comparison with Other Models in the Literature. Figure 4 represents the comparison of our GGL-Tox model with others on Tox21 Data Challenge 2014, where our model outperforms other methods with the highest average AUC for SR and NR toxicity data sets marked in red. Notice that in

https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700

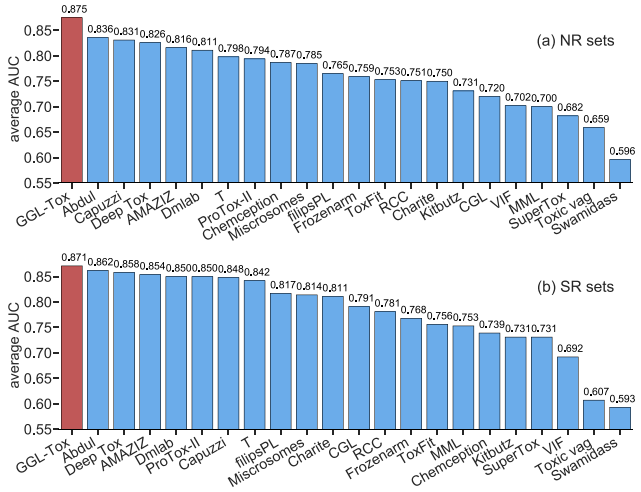


Figure 4. Comparison of prediction results with different methods on Tox21 Data Challenge 2014. The result of our model GGL-Tox is marked in red: (a) NR and (b) SR sets. The average AUC values of other methods are taken from refs 19, 25, 26, and 66−68 in blue.

order to make a fair comparison between our model and other state-of-the-art models, we carried out data preprocessing for each data set prior to modeling steps. The details of this procedure can be found in section S3 of the Supporting Information. Among these methods, Abdul et al.[25] developed a single task-based chemical toxicity prediction framework using only 2D fingerprints as well as shallow neural networks and decision trees, which achieves an average AUC of 0.836 for NR data sets and 0.862 for SR data sets. The winning model of Tox21 Data Challenge 2014 was based on the DeepTox pipeline[19] or deep learning and trained on around 273 577 features, resulting in the average AUC of 0.826 for NR and 0.858 for SR on the final test set. The large numbers of features used by the model made the training process very time-intensive and hard to interpret which features are playing a vital role in decision making. Amaziz et al.[66] introduced consensus models using associative neural network (ANN) to achieve an average AUC of 0.816 for NR and 0.854 for SR data sets. Here, ANN means a combination of an ensemble of feed-forward neural networks and the KNN technique. Unfortunately, information on the total number of features was not reported. Additionally, Capuzzi et al. used DNN with an ensemble of 2489 features to achieve a good overall average AUC of 0.831 for NR and 0.848 for SR.[26]

Our GGL-Tox model constructed from MWCG features and the GBDT algorithm obtains the state-of-the-art average AUC on both NR and SR data sets. The results in Figure 4 indicate that our model is computationally efficient and opens an avenue for interpretability.

Next, the similarity between 12 data sets of Tox21 Data Challenge 2014 was examined in Figure. 5 according to the Tanimoto coefficient ($S_{A,B}$):[69]

$$S_{A,B} = \frac{\sum_{i=1}^{N} x_{iA}x_{iB}}{\sum_{i=1}^{N} (x_{iA})^2 + \sum_{i=1}^{N} (x_{iB})^2 - \sum_{i=1}^{N} x_{iA}x_{iB}}, \quad (8)$$

where x_{iA} (x_{iB}) denotes the ith feature of molecule A (B). $S_{A,B} \in [0, 1]$ is used in the present work to calculate the degree of similarity between two molecule structures. A large average value of $S_{A,B}$ between two data sets means there is a high similarity between them. As shown in Figure 5, there are three

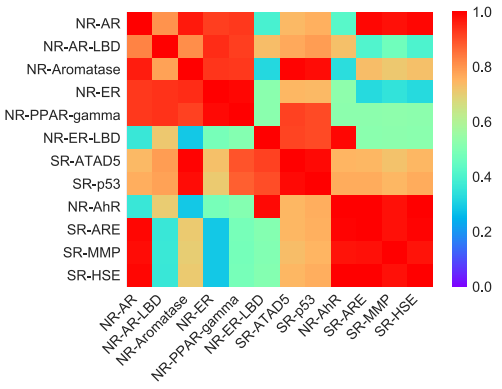


Figure 5. Heatmap of Tanimoto similarity between the different data sets of Tox21 Data Challenge 2014. Three clusters can be found obviously in the diagonal of the heatmap containing 5, 3, and 4 assays, respectively.

clusters. One cluster consists of five assays, NR-AR, NR-AR-LBD, NR-Aromatase, NR-ER, and NR-PPAR-gamma, with an average Tanimoto similarity of $S = 0.914$, which can be seen on the top-left corner of the figure. The other cluster has three assays, NR-ER-LBD, SR-ATAD5, and SR-p53, with an average Tanimoto similarity of $S = 0.936$, which is in the middle of the figure. The last cluster has four assays, NR-AhR, SR-ARE, SR-MMP, and SR-HSE, with an average Tanimoto similarity of $S = 0.987$, which can be observed on the bottom-right corner of the figure. The presence of three clusters indicates that ligand bind domains in each cluster share some structural similarities that can accommodate similar ligands. On the contrary, weak similarities are found between NR-ER and SR-MMP and between NR-ER and SR-MMP, whose similarities are $S = 0.286$ and 0.287, respectively.

Applicability Domain. The "Distance to model" (DM) approach is commonly used in the study of applicability domain assessment of QSAR models. In the present work, the method of Class-lag, a distance-based approach introduced by Sushko et al.,[70,71] is used to estimate the applicability domain of the GGL-Tox model for 12 data sets. For the binary classification problem of toxicity prediction, the label for nontoxic compounds can be marked by +1 when the prediction value is larger than 0, or marked by −1 otherwise, after we normalize the prediction values to the interval of −0.5 and 0.5. Then, the absolute value of the difference between the prediction value and the nearest of the labels can be used as a measure of prediction uncertainty. This measure, named as Class-lag, is calculated as

$$d_i^{\text{Class-lag}} = \min\{|-1 - p(i)|, |1 - p(i)|\}, \quad (9)$$

where i labels the ith compound to be predicted and $p(i)$ is its prediction value. Class-lag indicates the degree of close to nearest class label of a compound.

Figure 6 shows the effect of Class-lag DM of prediction with 12 data sets, where the y-axis corresponds to the values of $1 - d_i^{\text{Class-lag}}$ and the x-axis corresponds to the compound number. Green dots, which are closer to 0.5 in the y-axis, are predicted with better prediction accuracy than red dots that are closer to 0 in the y-axis. In other words, the "prediction certainty area" is labeled with green dots, and the "prediction uncertainty area" is labeled with red dots. Additionally, circles and squares are positive (nontoxic) and negative (toxic) predictions, respec-

https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700

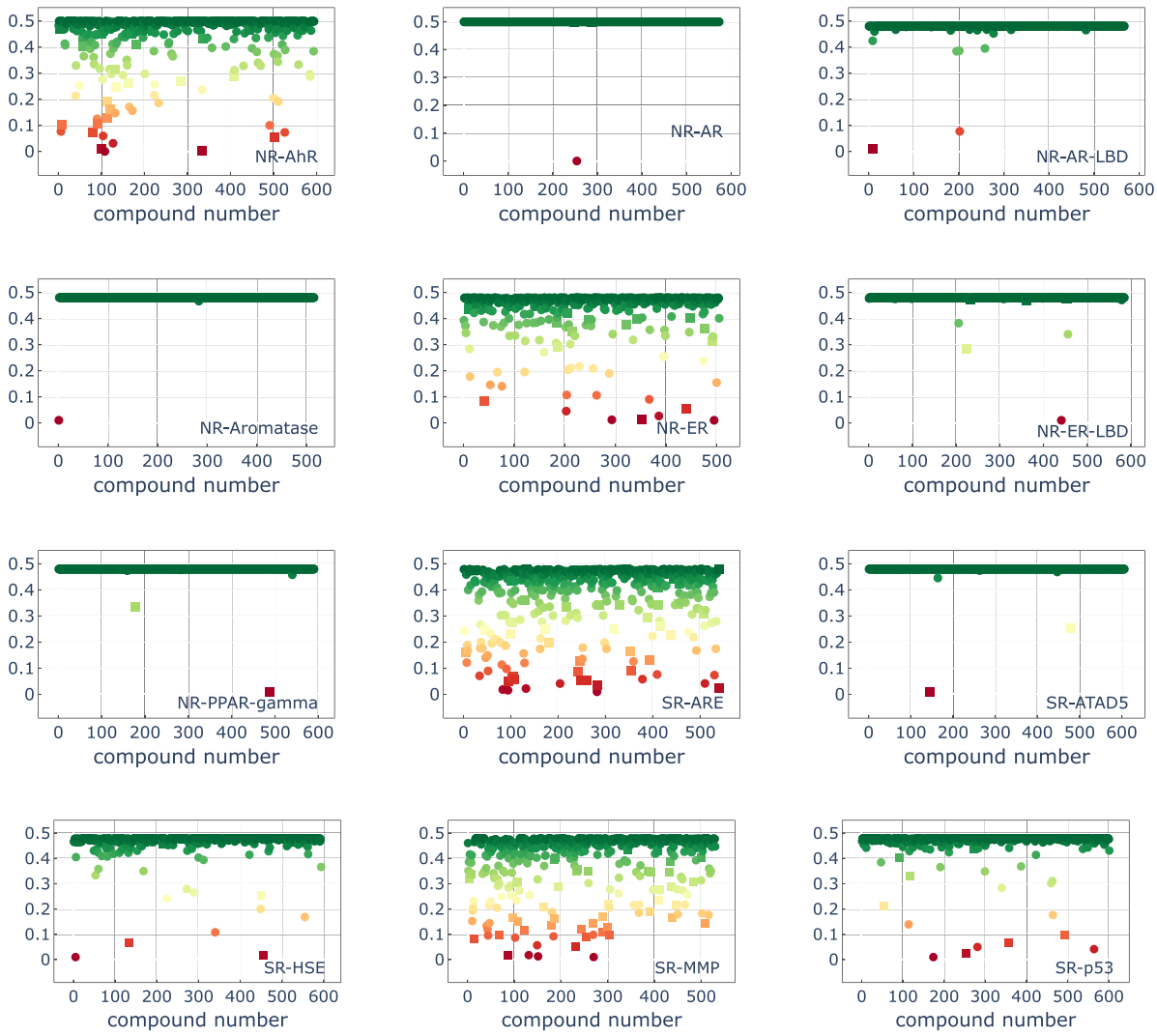


Figure 6. Graphical demonstration of the Class-lag DM. Green indicates the low values of the Class-lag DM, and red shows the high vales of the Class-lag DM. Circles represent the nontoxic compounds, and squares are for toxic compounds. The reliable predictions are colored with green, and the unreliable predictions are indicated with red.

tively. From Figure 6, we can find that most of the predictions have a few of red dots, except with data sets of NR-AhR, NR-ER, SR-ARE, and SR-MMP, which suggests that our GGL-Tox model predictions are reliable.

CONCLUSION

Chemical toxicity is of major concern in environment sciences and a vital factor for drug design and discovery. Despite of significant progress in the last two decades, toxicity analysis and prediction remain a challenging task due to chemical diversity, structural complexity, limited size, and poor quality of current toxicity data sets, which calls for innovative approaches. Geometric graph theories are commonly used in the study of molecular and biomolecular systems. Motivated by the success of the multiscale weighted colored graph (MWCG) approach for B-factor predictions,[35] we propose a geometric graph learning toxicity (GGL-Tox) model for toxicity analysis and prediction. In our GGL-Tox model, molecules are represented by MWCGs and the gradient boosting decision tree (GBDT) is chosen as the machine learning algorithm. Our GGL-Tox model is validated by using the Tox21 Data Challenge 2014, a benchmark tested for toxicity prediction methodologies. MWCGs are compared with standard two-dimensional (2D) fingerprints. The performance of GGL-Tox model is compared with other state-of-the-art methods in toxicity prediction and shows a cutting edge advantage. Additionally, our model shows high flexibility of applicability, which can be applied to toxicity prediction, and other problems, such as the prediction of solubility, solvation, partition coefficient, mutation-induced protein folding stability change, and protein−nucleic acid interactions.

Data and Model Availability. All data sets used in present work can be downloaded from the Web site https://tripod.nih.gov/tox21/challenge/, and the details of data can be found in section S2 of the Supporting Information. Our model is available at the online server https://weilab.math.msu.edu/Tox. Additionally, the codes for calculating the descriptors are available on Github https://github.com/jjlyl/descriptor-calculation.

https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c01294.

Number of data points per assay in Tox21 Challenge Data sets and the target and assay information (Table S1). Sets of parameters in the three-scale GGL-Tox model for data sets in Tox21 Data Challenge 2014 (Table S2). List of atoms in the Tox21 Data Challenge (Table S3). Brief summary of three 2D molecular fingerprints (Table S4). Hyperparameters in the GBDT algorithm with five different fingerprints for 12 data sets in toxicity prediction (Table S5). Performance of test data and cross-validation data for all 12 assays with the GGL-Tox method (Table S6). Prediction results and the hyperparameters of GBDT with Morgan fingerprint in Tox21 prediction (Tables S7 and S8, respectively). Comparison of GBDT, RF, and SVM prediction results with the MWCG fingerprint in toxicity prediction (Table S9). Main hyperparameters of RF and SVM in the implementation (Tables S10 and S11). Sample size for different data sets before and after the data preprocessing (Table S12). Omitted sample IDs in the training and test sets of 12 data sets (Table S13). Comparison of prediction results and the hyperparameters of GBDT with five different molecular fingerprints in toxicity prediction before data preprocessing (Tables S14 and S15, respectively). Details of the Tox21 Data Challenge and data preprocessing (sections S2 and S3), respectively (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Guo-Wei Wei − *Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, United States; Department of Biochemistry and Molecular Biology and Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0002-5781-2937; Email: wei@math.msu.edu

Authors

Jian Jiang − *Research Center of Nonlinear Science, College of Mathematics and Computer Science, Engineering Research Center of Hubei Province for Clothing Information, Wuhan Textile University, Wuhan 430200, P R. China*

Rui Wang − *Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0002-7402-6372

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c01294

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by NIH grant GM126189, NSF grants DMS-2052983, DMS-1761320, and IIS-1900473, NASA grant 80NSSC21M0023, Michigan Economic Development Corporation, George Mason University award PD45722, Bristol-Myers Squibb 65109, and Pfizer. J.J. was supported by the National Natural Science Foundation of China under grant nos. 11972266 and 11971367.

■ REFERENCES

(1) Maziasz, T.; Kadambi, V.-J.; Silverman, L.; Fedyk, E.; Alden, C.-L. Predictive toxicology approaches for small molecule oncology drugs. *Toxicol. Pathol.* **2010**, *38*, 148−164.

(2) Anastas, P.; Teichman, K.; Hubal, E.-C. Ensuring the safety of chemicals. *J. Exposure Sci. Environ. Epidemiol.* **2010**, *20*, 395−396.

(3) Lima, A.-N.; Philot, E.; Trossini, G.-H.-G.; Scott, L.-P.-B.; Maltarollo, V.-G.; Honorio, K. Use of machine learning approaches for novel drug discovery. *Expert Opin. Drug Discovery* **2016**, *11*, 225−239.

(4) Huang, R.; Xia, M.; Nguyen, D.-T.; Zhao, T.; Sakamuru, S.; Zhao, J.; Shahane, S. A.; Rossoshek, A.; Simeonov, A. Tox21 Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. *Front. Environ. Sci.* **2016**, *3*, 85.

(5) Banerjee, P.; Siramshetty, V. B.; Drwal, M. N.; Preissner, R. Computational methods for prediction of in vitro effects of new chemical structures. *J. Cheminf.* **2016**, *8*, 51.

(6) 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.

(7) Cao, D.-S.; Huang, J.-H.; Yan, J.; Zhang, L.-X.; Hu, Q.-N.; Xu, Q.-S.; Liang, Y.-Z. Kernel k-nearest neighbor algorithm as a flexible SAR modeling tool. *Chemom. Intell. Lab. Syst.* **2012**, *114*, 19−23.

(8) Li, X.; Chen, L.; Cheng, F.; Wu, Z.; Bian, H.; Xu, C.; Li, W.; Liu, G.; Shen, X.; Tang, Y. In silico prediction of chemical acute oral toxicity using multi-classification methods. *J. Chem. Inf. Model.* **2014**, *54*, 1061−1069.

(9) Ambe, K.; Ishihara, K.; Ochibe, T.; Ohya, K.; Tamura, S.; Inoue, K.; Yoshida, M.; Tohkin, M. In silico prediction of chemical-induced hepatocellular hypertrophy using molecular descriptors. *Toxicol. Sci.* **2018**, *162*, 667−675.

(10) Koutsoukas, A.; St Amand, J.; Mishra, M.; Huan, J. Predictive toxicology: modeling chemical induced toxicological response combining circular fingerprints with random forest and support vector machine. *Front. Environ. Sci.* **2016**, *4*, 11.

(11) 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. Inf. Model.* **2009**, *49*, 2481−2488.

(12) 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.

(13) Fernandez, M.; Ban, F.; Woo, G.; Hsing, M.; Yamazaki, T.; LeBlanc, E.; Rennie, P. S.; Welch, W. J.; Cherkasov, A. Toxic colors: the use of deep learning for predicting toxicity of compounds merely from their graphic images. *J. Chem. Inf. Model.* **2018**, *58*, 1533−1543.

(14) 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*, 5441−5451.

(15) Drwal, M. N.; Siramshetty, V. B.; Banerjee, P.; Goede, A.; Preissner, R.; Dunkel, M. Molecular similarity-based predictions of the Tox21 screening outcome. *Front. Environ. Sci.* **2015**, *3*, 54.

(16) Zhang, J.; Mucs, D.; Norinder, U.; Svensson, F. LightGBM: An Effective and Scalable Algorithm for Prediction of Chemical Toxicity−Application to the Tox21 and Mutagenicity Data Sets. *J. Chem. Inf. Model.* **2019**, *59*, 4150−4158.

(17) 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. Comput. Sci.* **2003**, *43*, 1947−1958.

(18) Deng, C.-H.; Zhao, W.-L. Fast k-means based on k-NN Graph. *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, 2018; pp 1220−1223.

(19) Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* **2016**, *3*, 80.
(20) Nguyen, D. D.; Cang, Z.; Wei, G.-W. A review of mathematical representations of biomolecular data. *Phys. Chem. Chem. Phys.* **2020**, *22*, 4343−4367.
(21) 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.
(22) Wu, K.; Wei, G.-W. Quantitative toxicity prediction using topology based multitask deep neural networks. *J. Chem. Inf. Model.* **2018**, *58*, 520−531.
(23) Idakwo, G.; Thangapandian, S.; Luttrell, J.; Zhou, Z.; Zhang, C.; Gong, P. Deep Learning-Based Structure-Activity Relationship Modeling for Multi-Category Toxicity Classification: A Case Study of 10K Tox21 Chemicals With High-Throughput Cell-Based Androgen Receptor Bioassay Data. *Front. Physiol.* **2019**, *10*, 1044.
(24) Matsuzaka, Y.; Uesawa, Y. Prediction Model with High-Performance Constitutive Androstane Receptor (CAR) Using DeepSnap-Deep Learning Approach from the Tox21 10K Compound Library. *Int. J. Mol. Sci.* **2019**, *20*, 4855.
(25) Karim, A.; Mishra, A.; Newton, M. H.; Sattar, A. Efficient toxicity prediction via simple features using shallow neural networks and decision trees. *ACS Omega* **2019**, *4*, 1874−1888.
(26) 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**, *4*, 3.
(27) Janezic, D.; Milicevic, A.; Nikolic, S.; Trinajstic, N. *Graph-theoretical matrices in chemistry*; CRC Press, FL, 2015.
(28) Angeleska, A.; Jonoska, N.; Saito, M. DNA recombination through assembly graphs. *Discrete Applied Mathematics* **2009**, *157*, 3020−3037.
(29) Levitt, M.; Sander, C.; Stern, P. S. Protein Normal-mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease andiLysozyme. *J. Mol. Biol.* **1985**, *181*, 423−447.
(30) Atilgan, A. R.; Durell, S.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505−515.
(31) Foulds, L. R. *Graph theory applications*; Springer Science & Business Media, New York, 2012.
(32) Ozkanlar, A.; Clark, A. E. ChemNetworks: A complex network analysis tool for chemical systems. *J. Comput. Chem.* **2014**, *35*, 495−505.
(33) Di Paola, L.; Giuliani, A. Protein contact network topology: a natural language for allostery. *Curr. Opin. Struct. Biol.* **2015**, *31*, 43−48.
(34) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513−530.
(35) Bramer, D.; Wei, G.-W. Multiscale weighted colored graphs for protein flexibility and rigidity analysis. *J. Chem. Phys.* **2018**, *148*, 054103.
(36) Nguyen, D. D.; Xiao, T.; Wang, M.; Wei, G.-W. Rigidity strengthening: A mechanism for protein−ligand binding. *J. Chem. Inf. Model.* **2017**, *57*, 1715−1721.
(37) Nguyen, D. D.; Wei, G.-W. AGL-Score: Algebraic Graph Learning Score for Protein−Ligand Binding Scoring, Ranking, Docking, and Screening. *J. Chem. Inf. Model.* **2019**, *59*, 3291−3304.
(38) Kojima, R.; Ishida, S.; Ohta, M.; Iwata, H.; Honma, T.; Okuno, Y. kGCN: a graph-based deep learning framework for chemical structures. *J. Cheminf.* **2020**, *12*, 32.
(39) Winter, R.; Montanari, F.; Noe, F.; Clevert, D. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci.* **2019**, *10*, 1692−17012.
(40) Nguyen, D. D.; Cang, Z.; Wu, K.; Wang, M.; Cao, Y.; Wei, G.-W. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 71−82.
(41) National Research Council. *Toxicity testing in the 21st century: a vision and a strategy*; National Academies Press, Washington, D.C, 2007.
(42) Collins, F. S.; Gray, G. M.; Bucher, J. R. Transforming environmental health protection. *Science* **2008**, *319*, 906−907.
(43) Krewski, D.; Acosta, D., Jr; Andersen, M.; Anderson, H.; Bailar, J. C., III; 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.; et al. Toxicity testing in the 21st century: a vision and a strategy. *J. Toxicol. Environ. Health, Part B* **2010**, *13*, 51−138.
(44) Huang, R.; Xia, M.; Cho, M.-H.; Sakamuru, S.; Shinn, P.; Houck, K. A.; Dix, D. J.; Judson, R. S.; Witt, K. L.; Kavlock, R. J.; Tice, R.-R.; Austin, C.-P. Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. *Environ. Health Perspect.* **2011**, *119*, 1142−1148.
(45) Huang, R.; Sakamuru, S.; Martin, M. T.; Reif, D. M.; Judson, R. S.; Houck, K. A.; Casey, W.; Hsieh, J.-H.; Shockley, K. R.; Ceger, P.; Fostel, J.; Witt, K. L.; Tong, W.; Rotroff, D. M.; Zhao, T.; Sinn, P.; Simeonov, A.; Dix, D. J.; Austin, C. P.; Kavlock, R. J.; Tice, R. R.; Xia, M. Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. *Sci. Rep.* **2015**, *4*, 5664.
(46) Chen, S.; Hsieh, J.-H.; Huang, R.; Sakamuru, S.; Hsin, L.-Y.; Xia, M.; Shockley, K. R.; Auerbach, S.; Kanaya, N.; Lu, H.; Svoboda, D.; Witt, K. L.; Merrick, B. A.; Teng, C. T.; Tice, R. R. Cell-based high-throughput screening for aromatase inhibitors in the Tox21 10K library. *Toxicol. Sci.* **2015**, *147*, 446−457.
(47) Attene-Ramos, M. S.; Huang, R.; Michael, S.; Witt, K. L.; Richard, A.; Tice, R. R.; Simeonov, A.; Austin, C. P.; Xia, M. Profiling of the Tox21 chemical collection for mitochondrial function to identify compounds that acutely decrease mitochondrial membrane potential. *Environ. Health Perspect.* **2015**, *123*, 49−56.
(48) Opron, K.; Xia, K.; Wei, G.-W. Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis. *J. Chem. Phys.* **2014**, *140*, 234105.
(49) Myint, K. Z.; Xie, X.-Q. *Artificial Neural Networks*; Springer, New York, 2015; pp 149−164.
(50) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(51) Wang, L.; Ma, C.; Wipf, P.; Liu, H.; Su, W.; Xie, X.-Q. TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *AAPS J.* **2013**, *15*, 395−406.
(52) Yao, K.; Parkhill, J. Kinetic energy of hydrocarbons as a function of electron density and convolutional neural networks. *J. Chem. Theory Comput.* **2016**, *12*, 1139−1147.
(53) Lo, Y.-C.; Rensi, S. E.; Torng, W.; Altman, R. B. Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today* **2018**, *23*, 1538−1546.
(54) Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular fingerprint similarity search in virtual screening. *Methods* **2015**, *71*, 58−63.
(55) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273−1280.
(56) Hall, L. H.; Kier, L. B. Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039−1045.
(57) Landrum, G. *RDKit: Open-source cheminformatics*; 2006.
(58) Li, P. Robust logitboost and adaptive base class (abc) logitboost. *arXiv.org* **2012**, 1203.3491.
(59) Burges, C. J. *From ranknet to lambdarank to lambdamart: An overview*, MSR-TR-2010-82; 2010; Vol. *11*, p 81.
(60) Richardson, M.; Dominowska, E.; Ragno, R. Predicting clicks: estimating the click-through rate for new ads. *Proceedings of the 16th international conference on World Wide Web*, 2007; pp 521−530.
(61) Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5−32.

https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700

(62) Cortes, C.; Vapnik, V. Support-vector networks. *Machine learning* **1995**, *20*, 273−297.
(63) Alvarsson, J.; Eklund, M.; Andersson, C.; Carlsson, L.; Spjuth, O.; Wikberg, J. E. Benchmarking study of parameter variation when using signature fingerprints together with support vector machines. *J. Chem. Inf. Model.* **2014**, *54*, 3211−3217.
(64) Idakwo, G.; Thangapandian, S.; Luttrell, J.; Li, Y.; Wang, N.; Zhou, Z.; Hong, H.; Yang, B.; Zhang, C.; Gong, P. Structure-acitivity relationship-based chemical classification of highly imbalanced Tox21 datasets. *J. Cheminf.* **2020**, *12*, 66.
(65) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine learning in Python. *J. Mach Learn Res.* **2011**, *12*, 2825−2830.
(66) Abdelaziz, A.; Spahn-Langguth, H.; Schramm, K.-W.; Tetko, I. V. Consensus modeling for HTS assays using in silico descriptors calculates the best balanced accuracy in Tox21 challenge. *Front. Environ. Sci.* **2016**, *4*, 2.
(67) Goh, G. B.; Siegel, C.; Vishnu, A.; Hodas, N. O.; Baker, N. Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models. *arXiv.org* **2017**, 1706.06689.
(68) Banerjee, P.; Eckert, A. O.; Schrey, A. K.; Preissner, R. ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.* **2018**, *46*, W257−W263.
(69) Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminf.* **2015**, *7*, 20.
(70) Sushko, I.; Novotarskyi, S.; Korner, R.; Pandey, A. K.; Cherkasov, A.; Li, J.; Gramatica, P.; Hansen, K.; Schroeter, T.; Muller, K.; Xi, L.; Liu, H.; Yao, X.; Oberg, T.; Hormozdiari, F.; Dao, P.; Sahinalp, C.; Todeschini, R.; Polishchuk, P.; Artemenko, A.; Kuzmin, V.; Martin, T. M.; Yong, D. M.; Fourches, D.; Muratov, E.; Tropsha, A.; Baskin, I.; Horvath, D.; Marcou, G.; Muller, C.; Varnek, A.; Prokopenko, V. V.; Tetko, I. V. Applicability domains for classification problems: Benchmarking of distance to models for ames mutagenicity set. *J. Chem. Inf. Model.* **2010**, *50*, 2094−2111.
(71) 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*, 1733−1746.

https://dx.doi.org/10.1021/acs.jcim.0c01294
J. Chem. Inf. Model. 2021, 61, 1691−1700