

Article

# Toxicity Prediction Method Based on Multi-Channel Convolutional Neural Network

Qing Yuan, Zhiqiang Wei, Xu Guan, Mingjian Jiang, Shuang Wang, Shugang Zhang and Zhen Li \*

College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China; yuanqing@stu.ouc.edu.cn (Q.Y.); weizhiqiangouc@163.com (Z.W.); guanxu@stu.ouc.edu.cn (X.G.); jmj@stu.ouc.edu.cn (M.J.); ws@stu.ouc.edu.cn (S.W.); zhangshugang@hotmail.com (S.Z.)

\* Correspondence: lizhen0130@gmail.com; Tel./Fax: +86-532-66781793

Received: 15 July 2019; Accepted: 13 September 2019; Published: 17 September 2019



**Abstract:** Molecular toxicity prediction is one of the key studies in drug design. In this paper, a deep learning network based on a two-dimension grid of molecules is proposed to predict toxicity. At first, the van der Waals force and hydrogen bond were calculated according to different descriptors of molecules, and multi-channel grids were generated, which could discover more detail and helpful molecular information for toxicity prediction. The generated grids were fed into a convolutional neural network to obtain the result. A Tox21 dataset was used for the evaluation. This dataset contains more than 12,000 molecules. It can be seen from the experiment that the proposed method performs better compared to other traditional deep learning and machine learning methods.

**Keywords:** deep learning; Tox21; toxicity prediction; convolutional neural networks

## 1. Introduction

Our daily lives are exposed to a large number of compounds including the environment, foods, skin care products and medicines. In order to protect human health and the human body from potentially harmful substances, these compounds must pass reliable adverse reaction tests, especially toxicity tests. In addition, the actual chemical experiments could test the toxicity of compounds, but they consume considerable manpower, financial and material resources, and many experiments are very time-consuming and may be involved in ethical issues. With the rapid development of artificial intelligence, computers are used to accomplish these tasks. This is a powerful tool for compound property prediction. However, the accuracy of the computer-aided method is still a problem which needs to be solved.

From the end of the last century to the present, as machine learning continues to flourish in various fields, its application in biomedicine has become more widespread [1]. Further, artificial intelligence has been widely used in drug design [2]. For example, in terms of the binding between proteins and small molecules, an energy-based method for the prediction of protein-ligand binding sites was proposed [3]. Michael S. Lee et al. [4] researched the calculation of the binding affinity through the path and endpoint approaches. George E. Dahl et al. [5] proposed multitask neural networks for quantitative structure–activity relationship (QSAR) predictions.

For the QSAR predictions including toxicity predictions, the molecular feature and fingerprints are very important for the performance of the prediction. The simplified molecular input line entry system (SMILES) [6] is a popular language used to represent molecules. Some methods focused on the 2D structure of the molecule and others focused on the 3D information collection. For example, the Barnard Chemical Information Ltd. (BCI) fingerprint is generated according to six different families, including augmented atoms, atom/bond sequences, atom pairs, ring composition fragments, ring fusion fragments and ring ortho fragments [7]. As a molecule is a graph, which means the connection

between each atom could provide useful information, there are some methods researching on how to utilize graph information. For example, the extended-connectivity circular fingerprints (ECFP) [8] is the standard circular fingerprints based on the Morgan algorithm [9]. It selects each heavy atom as the start node and generates the bag of fragments iteratively. Each fragment is hashed into a representation with a fixed-length. The ECFP could be calculated very rapidly and the information of substructures collected. Furthermore, in order utilize the 3D information, Mahendra Awale et al. [10] proposed a fingerprint combining the molecular shape extracted from the 3D-structures and the 2D atom pair fingerprint.

All these features could be fed into certain machine learning methods for prediction. C.Y. Zhao [11] used a support vector machine (SVM) for toxic activity prediction, and compared it with two other methods, the multiple linear regression (MLR) and the radial basis function neural network (RBFNN). It was demonstrated that the SVM performed best among these three methods of different data sets. Furthermore, Hui Zhang et al. [12] proposed a GA-CG-SVM method which utilized the genetic algorithm (GA) to select the features and GA to optimize the parameters. A k-nearest neighbor (kNN) model was introduced for the prediction of acute contact toxicity of pesticides [13]. Decision trees were utilized to estimate sensory irritation potency of volatile organic chemicals [14]. In addition, random forest (RF) [15], which combines multiple decision trees, was also used for toxicity prediction. Renee Solimeo et al. [16] combined the kNN method and RF method together to predict chemical ocular toxicity. On the other hand, some generative machine learning models have also been used in toxicity prediction. The naïve Bayes model was utilized for predicting the carcinogenicity of chemicals and provided guidance for drug design [17]. The linear discriminant analysis (LDA) was also introduced to predict idiosyncratic hepatotoxicity [18]. However, because traditional machine learning methods are based on hand-craft features mentioned before, it is often difficult to find important attributes and positional information that truly affect the toxicity of the molecule, which may adversely affect the results.

With the development of the deep learning method, it plays a significant role in the research fields of biology and medicine. For example, recurrent neural networks (RNN) were utilized to generate focused molecule libraries [19], and a deep neural network with a multi-task learning setting was used for a binding prediction [20]. Moreover, Mohammed AlQuraishi et al. [21] also proposed the end-to-end differentiable learning method to predict the protein structure. In the prediction of molecular activity, a structure-based convolutional neural network was proposed to predict the biological activity of small molecules in drug discovery applications, known as AtomNet [22]. In recent years, the research on the toxicity prediction of deep learning on molecules has been continuously proposed. With the help of deep learning, the features could be extracted automatically. Youjun Xu et al. [23] proposed a regression model and multiclass models for toxicity prediction with the assistance of automatic chemical feature extraction and deep learning. Kedi Wu et al. [24] used topology and multitask deep neural networks for quantitative toxicity prediction. Swamidass et al. [25] designed a 4-layer deep neural network (DNN) to predict sites of epoxidation, which provides atomic-level and molecule-level precise information based on 702 epoxidation reactions. Abdul Karim et al. [26] combined the strings, images and numerical features together, and proposed a multimodal deep learning method for toxicity prediction.

For the deep learning method, the dataset selection is crucial to evaluate the performance of the method. The IGC50 dataset [24] is a popular dataset which records acute aquatic toxicity of the compound on the tetrahymena pyriformis population. It contains 1792 compounds which are stored in SMILES format. The fully convolutional neural network (FCNN), the convolutional neural network (CNN), RNN and the combination of these methods were implemented to test the performance of the IGC50 dataset [26]. Another dataset for the prediction of toxicity is the dataset of drug-induced liver injury [27] which contains 1087 compounds for training and 120 compounds for testing. Furthermore, an RNN-based method has been implemented on this dataset for toxicity prediction [28]. The ToxCast dataset collects toxicology data through *in vitro* high-throughput screening, which covers 600 experiments on 8615 compounds. For each assay, only 205 molecules on average could be used for training and

testing [29]. The Tox 21 [30] is the most famous dataset used for the prediction of toxicity. It contains more than 12,000 compounds with 12 properties. With the assistance of this dataset, this study could evaluate the method with respect to 12 properties, which could demonstrate the generalization of the method. There are also many deep learning methods performed on the Tox21 dataset, such as CNN [31], the pre-trained autoencoder [32], the associative neural networks (ASNN) [33] and DNN [34], which could be used for a horizontal comparison. Therefore, the Tox21 dataset was selected as the evaluation dataset in this paper.

Although there are some methods focusing on toxicity prediction using deep learning methods, most of them have utilized the SIMLEs string as the input of the network ignoring the spatial information of the molecule, which is an important factor of molecular activity. At the same time, the grid method has been utilized to extract the characteristics of the molecule in different grid locations, which could introduce the spatial information to the molecular feature. For example,  $K_{DEEP}$  [35] constructed a multi-channel three-dimensional grid for characterizing the binding sites of proteins and small molecules, a Van der Waals force-like method was used to calculate the contribution of all molecules to each voxel of grid within a certain range, and the grid was fed into a deep neural network to predict the location of the binding site. Although the grid method provides a novel view to represent the structure of the molecule, only the van der Waals force is not enough to represent the structures of the molecule.

Therefore, a grid-based method is introduced to extract the molecular features and study the nature of toxicity in this paper. Multi-channel data is utilized to improve the prediction performance. It can be seen from the experimental results that the proposed method is more accurate and robust. The contributions of the method are listed as follows:

- A two-dimensional grid approach for toxicity prediction of the molecule is proposed.
- In order to improve the accuracy of toxicity prediction, Van der Waals forces and hydrogen bonding properties are combined to improve generalization performance.
- The grid calculation of the constructed molecules reveals the contribution of each atom of the molecule to the prediction. The different grid weight values are calculated by this method, indicating that atoms with higher grid weight values correspond to higher contributions in prediction.

## 2. Methods

### 2.1. Dataset

The Tox21 [14,30] dataset is currently the most widely used toxicity research database in the world. The molecules in the Tox21 database could obtain the exact values of corresponding properties through experimental detection, which could be used to verify our method.

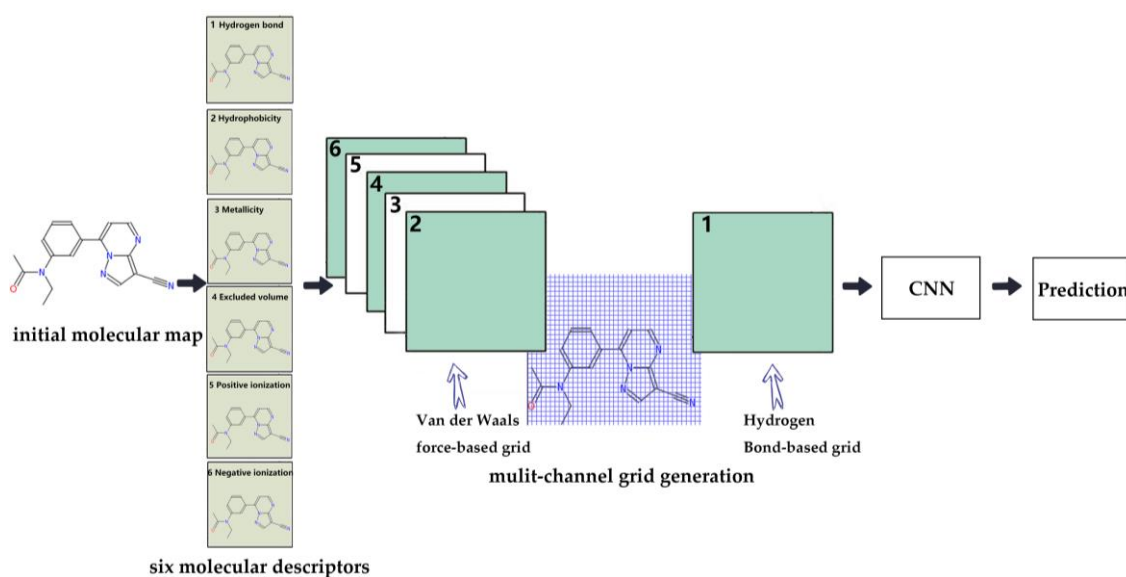
There are 12 toxic substances in Tox21 are shown in Table 1, including the stress response effects (SR) and the nuclear receptor effects (NR). The SR includes five types (ARE, HSE, ATAD5, MMP, p53), and NR includes seven types (ER-LBD, ER, Aromatase, AhR, AR, AR-LBD, PPAR). Regarding the property ER of NR, Y Asako et.al used the high-performance prediction based on chemical structures [36]. Both the SR and NR effects are closely related to human health. For example, the activation of nuclear receptors can disrupt endocrine system function [37], and the activation of stress response pathways can lead to liver damage or cancer [38]. The Tox21 database contains the results of high-throughput screening for these 12 toxic effects. Based on this database, a corresponding grid-based deep learning model was constructed in this paper to predict whether a certain property of a molecule was toxic or not.

**Table 1.** Content of 12 properties. The left side are the seven toxic properties of the nuclear receptor effects (NR), and the right side are the five toxic properties of the heat stress response effects (SR).

Nuclear Receptor Panel		Stress Response Panel	
ER-LBD	estrogen receptor alpha, luciferase	HSE	heat shock factor response element genotoxicity indicated by ATAD5 mitochondrial membrane potential
ER	estrogen receptor alpha	ATAD5	
AhR	aryl hydrocarbon receptor	MMP	
AR	androgen receptor	p53	DNA damage p53 pathway nuclear factor (erythroid-derived 2)-like 2 antioxidant responsive element
PPAR.g	peroxisome proliferator-activated receptor gamma	ARE	
AR-LBD	androgen receptor, luciferase		
Aromatase	cytochrome P450 enzymes		

## 2.2. Method Overview

The whole process of the proposed structure is described as follow. First, the appropriate molecular descriptors are selected and extracted. These are beneficial for prediction. Secondly, the molecular grid is constructed to obtain the corresponding data according to the molecular descriptor. Then, the data are fed into the convolution neural network to obtain the prediction results. Moreover, in order to solve the problem of the unbalanced distribution of the positive and negative data sets in the Tox21 database, the method of data enhancement is introduced. The flow chart of this method is shown in Figure 1.

**Figure 1.** The work flow of the proposed method.

## 2.3. Construction of Molecular Descriptors

An important step in predicting the toxicity of a molecule is to construct suitable molecular descriptors. Moreover, in order to ensure that the descriptors could be fed into a deep learning network, the size of them should be fixed for all molecules. Therefore, six multi-channel grid descriptors were introduced in this paper. Based on the molecular grid descriptors [35] which have been used for predicting binding sites of proteins, the proposed grid descriptors in this paper were especially designed and suitable for molecular toxicity.

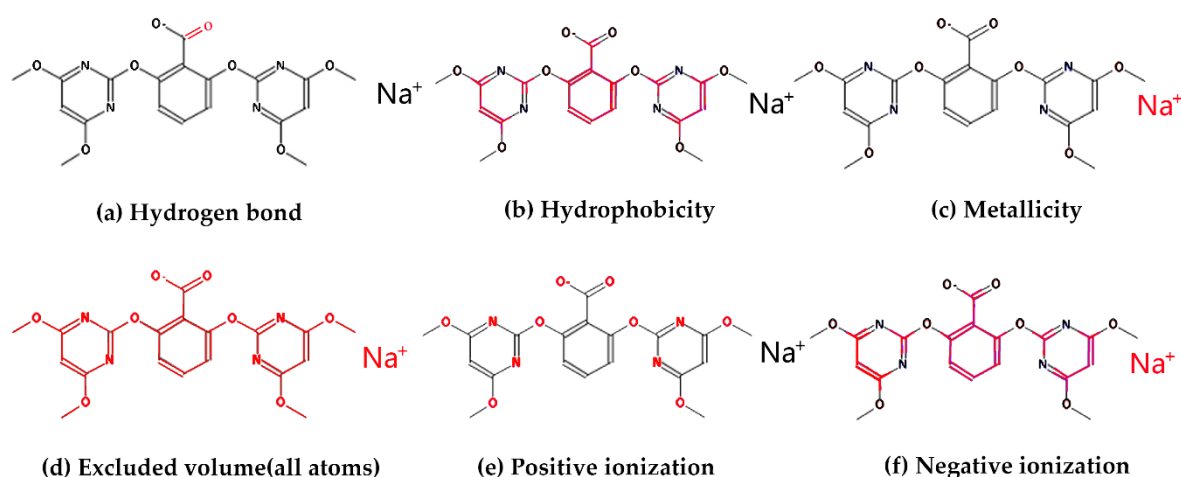
There are many properties of a single molecule or between multi-molecules used for property prediction, such as hydrophobicity, hydrogen bonding, ionization and metallicity. In the authors' opinion, more properties could be helpful in distinguishing different activities of molecules. In order

to obtain the property information of the molecule more precisely, six descriptors were introduced. These are shown in Table 2.

**Table 2.** The molecular descriptors.

Channel	Descriptor
1	Hydrogen bond
2	Hydrophobicity
3	Metallicity
4	Excluded volume
5	Positive ionization
6	Negative ionization

Based on the above description of molecular descriptors and the characteristics of the atoms contained in the data, each molecule could be divided into multi-channel data. For each descriptor, only related atoms are left in the channel. A case of molecule multi-channel descriptors is shown in Figure 2. A molecule in the NR-AhR property was selected. The corresponding atoms and bonds are highlighted in each channel.



**Figure 2.** Multi-channel molecular descriptors. (a) the descriptor of Hydrogen bond, (b) the descriptor of Hydrophobicity, (c) the descriptor of Metallicity, (d) the descriptor of Excluded volume, (e) the descriptor of Positive ionization, (f) the descriptor of Negative ionization.

#### 2.4. Construction of Molecular Grid

All channels extracted from the molecules are not suitable to input a learning network directly. Therefore, six 2D grids were constructed over the channels respectively. Moreover, in order to deep discover the structure and property information of a molecule and ease the sparse problem, two calculation methods were introduced to calculate and generate grids with respect to different channels.

The Van der Waals force-based grid: The Van der Waals force is important to the molecular property, where all atom types are involved. The contribution of the Van der Waals force from each atom of each grid pixel is calculated through Equation (1)

$$W(g)_{vdw} = \sum_{i=1}^N (1 - \exp(-(\frac{r_{vdw}(a_i)}{r_a(g, a_i)})^{12})) \quad (1)$$

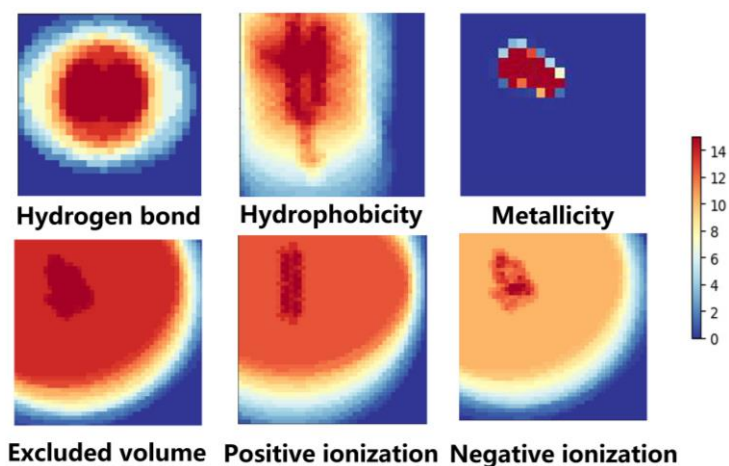
where  $a_i$  represents the  $i$ th atom of a molecule, and  $r_{vdw}(a)$  represents the van der Waals force radius of each type of atom.  $r_a(g, a)$  represents the distance of each atom  $a$  from the specified grid pixel  $g$ .  $W(g)_{vdw}$  represents the sum of the attribute values of pixel  $g$  from all atoms of within the specified range. The Van der Waals force was applied on the first five properties in Table 2 to generate grids.

The hydrogen bond-based grid: The information of the hydrogen bond donor and the corresponding atom of the hydrogen bond acceptor were utilized to generate another grid through Equation (2), which was applied on the last property of Table 2.

$$W(g)_{hbond} = \sum_{i=1}^N \left( \left( \frac{C}{r(g, a_i)^{12}} \right) - \frac{D}{r(g, a_i)^{10}} \right) \quad (2)$$

where  $W(g)_{hbond}$  represents the hydrogen bonds of all atoms in each grid pixel  $g$ .  $a_i$  represents each atom,  $r(g, a)$  represents the Euclidean distance of each atom to each grid pixel point.  $C$  is the well depth parameter assigned according to the hydrogen bonds with oxygen and nitrogen, and  $D$  is the well depth parameter assigned according to the hydrogen bonds with sulfur. The hydrogen bond is calculated through Autodock [39].

For each grid, it is generated according to its corresponding method. These 2D grids descriptors for each molecule are more clear and more specific. The final grid construction process is shown in Figure 3. The grid size is  $24\text{\AA} \times 24\text{\AA}$ , and the resolution of grid was set as  $0.5\text{\AA} \times 0.5\text{\AA}$ . In the experiment section, the performance of different resolutions were tested and displayed. The performance of  $0.5\text{\AA} \times 0.5\text{\AA}$  is the best among them. With the assistance of the grid, the structure and chemical information of each molecule were extracted. This can be fed into a convolutional neural network for training.



**Figure 3.** Heat map of six descriptors. The distribution of a two-dimensional grid of the five channels (positive/negative ionization, excluded volume, metallicity and hydrophobicity) were calculated using Van der Waals force-based grid method, and one channel about the hydrogen bond used the hydrogen bond-based grid method.

### 2.5. Convolutional Neural Network Architecture

The deep neural network, especially the convolutional neural network, is a feedforward neural network whose artificial neural unit can respond to a surrounding unit in a part of the coverage [40]. A convolutional neural network consists of one or more convolutional layers and a fully connected layer (corresponding to a classical neural network). Which is helpful for a two-dimensional structure as the input data. Compared with the other depth and feedforward neural networks, convolutional neural networks consider fewer parameters, which is an attractive deep learning structure.

In this paper, the convolutional neural network was used to predict the toxicity of the molecule and identify the key functional part of the molecule. The reason why the authors choose a simple version CNN is that the calculated parameters are relatively large for some complicated networks. Reducing the number of network layers could ensure accuracy and prevent over-fitting [41]. Through



the experiment, this study found that four layers of the structure is the optimal selection to obtain the best results. Figure 4 shows the structure of the convolutional neural network.

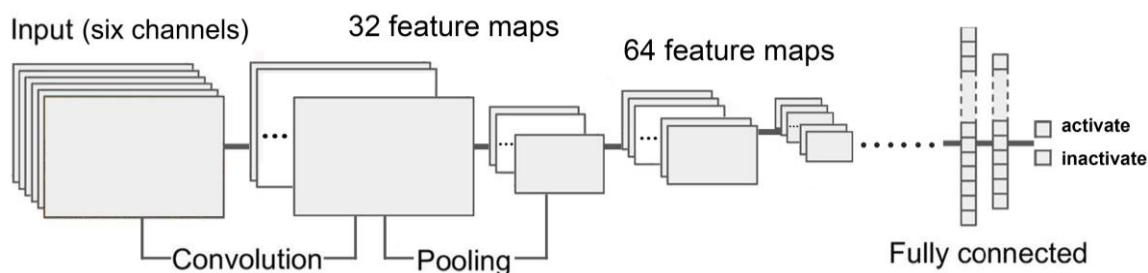


Figure 4. The structure of convolutional neural network (CNN).

During the input process, each molecule is described as a multi-channel specific two-dimensional array. Both the input layer and the hidden layer use the same activation function  $f$ .

Loss Function: The loss function is used to estimate the degree of inconsistency between the predicted value  $y(p)$  of the model and the true value  $Y$ , which is a binary classification in the toxicity classification. The binary cross entropy is selected as the loss function which are defined as follows:

$$loss = - \sum_{i=1}^n [(\hat{y}_i \log y_i + (1 - \hat{y}_i) \log(1 - \hat{y}_i))] \quad (3)$$

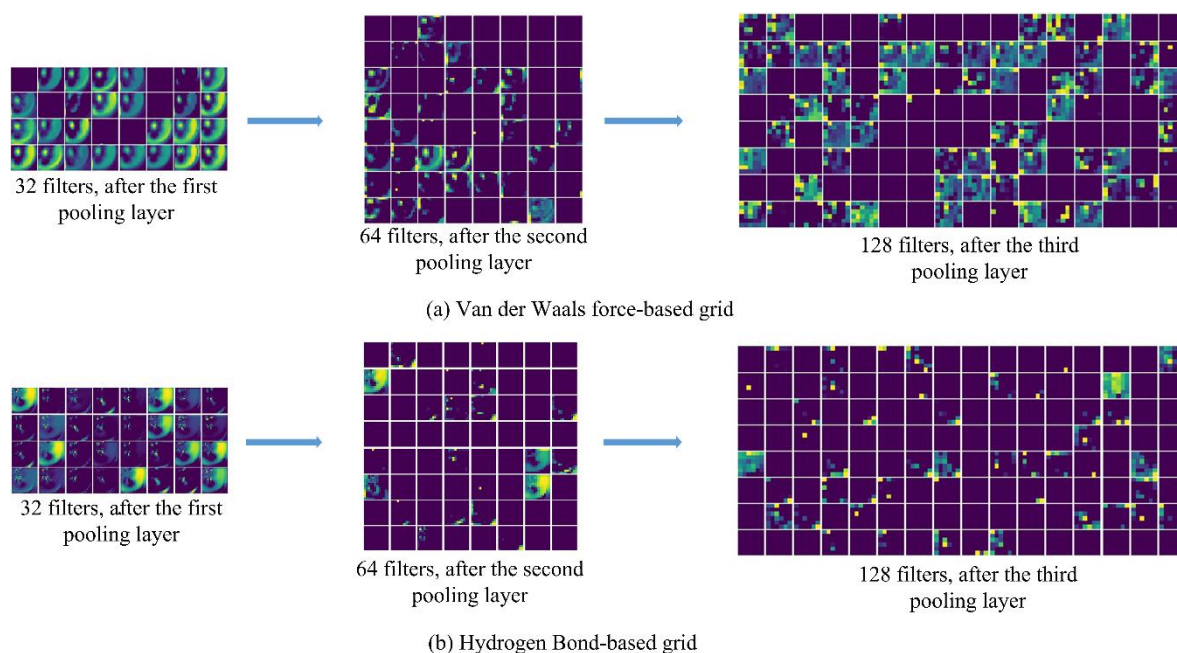
where  $n$  represents the number of categories of the classification, and  $\hat{y}_i$  represents the label. For a binary classification,  $\hat{y}_i$  represents 0 or 1.  $y_i$  represents the corresponding probability of the label.

In addition, the advantage of the CNN is that different layers of the structure reveal the significance of a molecule. Therefore, the largest and most concentrated values of the feature map of each layer could be regarded as the important positions of the molecule.

The six descriptors were calculated by two equations in Section 2.4, and the data of six channels were obtained and sent to the convolution neural network for training, and the feature map of the output data of each layer was obtained. The results of each layer after processing by the CNN are shown in Figure 5. The top part of the figure is the grid from the excluded volume, the first one is the visualization output through 32 filters of a two-dimensional convolution network structure, and the second and third are the results through 64 filters and 128 filters, as shown in Figure 5a. The second part of the figure is the grid from the hydrogen bond. Similarly, the feature maps of the output from each layer are shown in Figure 5b.

## 2.6. Data Augmentation

Due to the imbalance between the active and inactive data in the Tox21 dataset, data augmentation was implemented in the training process. There are 12 properties in Tox21 indicating whether a property is active or inactive of a molecule. However, not all molecules record all 12 properties performances, and the distribution of positive and negative samples of some properties are uneven which can affect the performance of the training network. For example, the distribution of positive and negative samples of NR-AR-LBD and SR-ATAD5 is very unbalanced, such as, in the training data of AR-LBD, the positive-negative ratio is 236:7798, that is, the number of negative samples is approximately 33 times that of the positive samples. Therefore, the synthetic minority oversampling technique (SMOTE) [42] was introduced in this paper. This is an up-sampling method for data processing of the unbalanced data. The core of the method is to generate more samples using existing samples. SMOTE is the most common method, which uses the interpolation between a small number of samples to generate additional samples. Moreover, the weights of positive samples and one-hot coding are also increased to solve the unbalance problem.



**Figure 5.** Visualization of training process. Three feature maps output after each pooling layer with different filters are displayed. The part (a) is grid from excluded volume, and the part (b) is grid from Hydrogen Bond.

### 3. Experiment

#### 3.1. Data Processing

The introduction of the training set, test dataset and verification set are shown in Table 3. In the table, the number of active samples, the number of total samples and the proportion of positive samples in the training sets, test set and verification set are displayed. Through the previous introduction, the problem of the dataset has already been discussed. It is obvious that the distributions of the active dataset and inactive dataset of 12 properties are uneven.

**Table 3.** Data distribution. The table shows the size of the training data set and test data set for each property, as well as the distribution of validation data set for each property.

	Train Active/Total	Test Active/Total	Verification Active/Total
AhR	776/7404 (10.48%)	73/600 (12.17%)	31/274 (11.31%)
AR	285/8472 (3.36%)	12/576 (2.08%)	3/294 (1.57%)
AR-LBD	236/7798 (3.03%)	8/574 (1.39%)	4/255 (1.19%)
ARE	940/6524 (14.41%)	93/546 (17.03%)	48/236 (20.34%)
Aromatase	298/6621 (4.50%)	36/519 (6.94%)	18/216 (8.33%)
ATAD5	265/8228 (3.22%)	38/612 (6.21%)	25/274 (9.12%)
ER	716/6967 (10.28%)	49/509 (9.63%)	27/267 (10.11%)
ER-LBD	343/7960 (4.31%)	19/589 (3.23%)	10/289 (3.46%)
PPAR.g	192/7432 (2.58%)	31/595 (5.21%)	15/269 (5.58%)
HSE	348/7398 (4.70%)	22/601 (3.66%)	10/269 (3.71%)
MMP	966/6619 (14.59%)	60/536 (11.19%)	38/240 (15.83%)
p53	447/7828 (5.71%)	40/605 (6.61%)	28/271 (10.33%)

Moreover, the data enhancement methods were used in the training data, and batch normalization [43] and dropout [44] were also introduced to prevent the data from over-fitting.

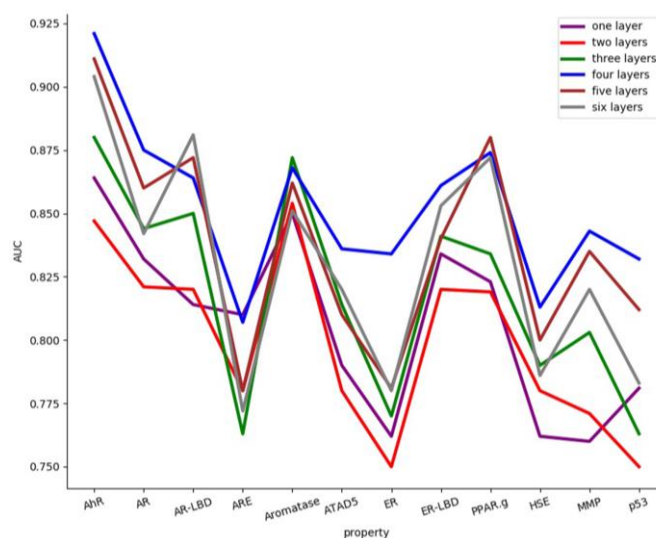
Before the features of the data were extracted, all data were standardized and cleaned. First, the Standardizer software (version 19.18.0, ChemAxon, Budapest, Hungary) was used for standardization.



Each molecule was added explicit hydrogens and aromatized. Then, the duplicated data were cleaned according to InChIKeys values [45]. The rule for deleting molecules is that for different molecules with the same InChIKeys, these molecules are deleted when they have different values for the same property.

### 3.2. Results of Different Layers in Convolutional Networks

Through the introduction of the CNN, we designed four hidden layers CNN were designed. In order to select the most appropriate number of layers, a comparative experiment with different layers was implemented. The structure of the convolution neural network was chosen with more than four layers (five layers and six layers) and less than four layers (one layer, two layers and three layers) for comparison. The area under curve (AUC) results indicate that there are approximately 12 properties as shown in Figure 6.



**Figure 6.** The area under the curve (AUC) of 12 properties with different hidden layers. The different colors represent 12 AUC values with different hidden layers.

It is obvious that when the number of hidden layers is set to 4, the best result is obtained. It is possible that when the hidden layers are relatively small, the network may not be trained well and the network performance is very poor. If the number of hidden layers is too large, the network system error can be reduced, and the network training time can be prolonged. Moreover, the training process can easily fall into the local minimum and cannot find the optimal point.

### 3.3. Hyperparameter and Loss

All 12 toxic properties can be predicted by the proposed model. For the different properties, there are some parameter adjustments during the test, as shown in Table 4.

**Table 4.** Hyperparameter settings of the neural networks.

Hyperparameter	Values are Selected
number of hidden units	64, 128, 256, 512, 1024, 2048
numbers of hidden layers	1, 2, 3, 4, 5, 6
learning rate	$10^{-3}$ , $10^{-4}$ , $10^{-5}$
dropout rate	20% input layer et.al
L2 weight decay	$10^{-2}$ , $10^{-3}$

From the data in Table 4, it can be seen that the hyperparameter settings of each property have different choices. The setting of the hyperparameter for the training process of each property are also provided in Appendix A.

The loss during the training process was also recorded. In order to pay more attention to the changes of loss, the loss curves for each property are shown in Figure 7.

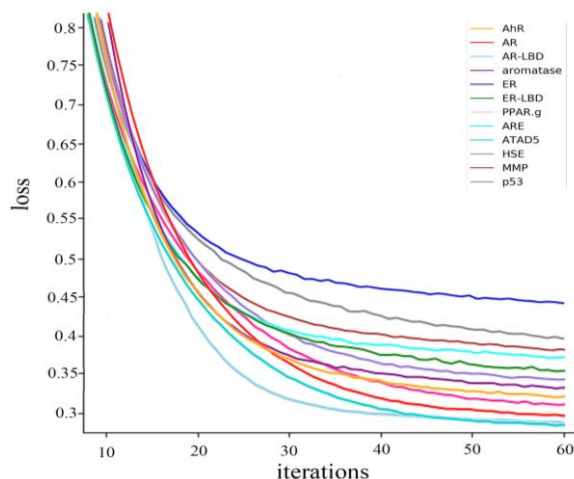


Figure 7. Training loss curve.

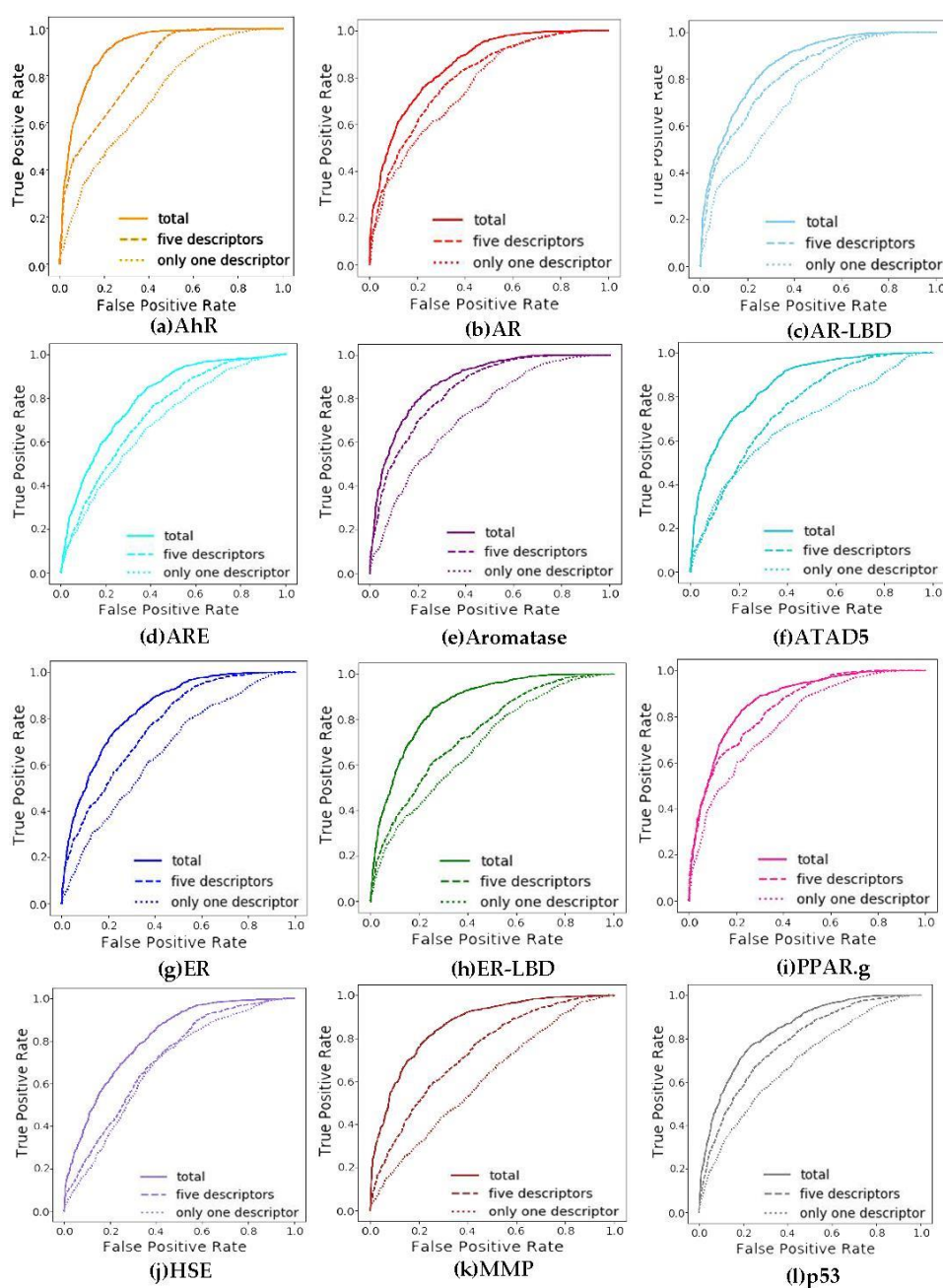
According to the 60 iterations of training in Figure 7, the loss curves of each property show a downward trend and become convergent after 30 iterations.

### 3.4. Evaluation on Different Channels

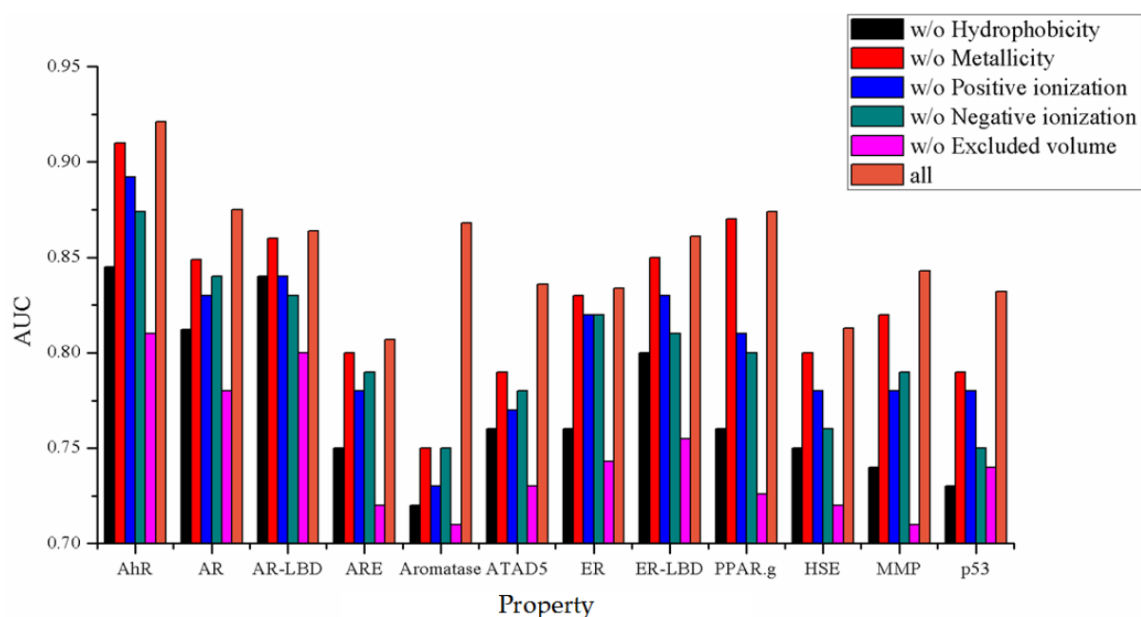
However, this study also realized the importance of extracting features. Two methods were used in the feature extraction part, the first one was used to extract the van der Waals force and the second one was used to calculate the hydrogen bond. It is important to demonstrate that combining two grids provides a better way to explicate the nature of a molecule. Therefore, the comparisons between the different grid methods including only each one of the grid calculation methods and the combination of both methods are shown in Figure 8. It was demonstrated that the performance with two methods fusion is better than any one of them. It is possible that the combined results of the two methods are most suitable for feature extraction which reveal the nature of toxicity.

Each property contains three receiver operating characteristic (ROC) curves, in which the solid line represents the result of the calculation of the six descriptors, the dashed line represents the result after the calculation of the five descriptors by Van der Waals force, and the dotted line represents the result after the calculation of a descriptor refined the hydrogen bond.

From the comparison of ROC curves in Figure 8, it can be seen that the best choice is to extract the properties of six channels. In addition, the Van der Waals force equation was used to calculate the five descriptors, and the hydrogen bond equation was used to calculate the other descriptor. Compared with the two methods, the former has a greater positive impact on the results. In order to explore which of the five channels using the Van der Waals force method has the most significant impact on the results, this study adopted a comparative experiment. Any four channels of the Van der Waals force method were used and combined with hydrogen bonds for comparison. The experimental results are shown in Figure 9. From Figure 9, it is obvious that the properties excluded the volume and hydrophobicity which have great influence on the results, and combining all the descriptors together obtains the best results.



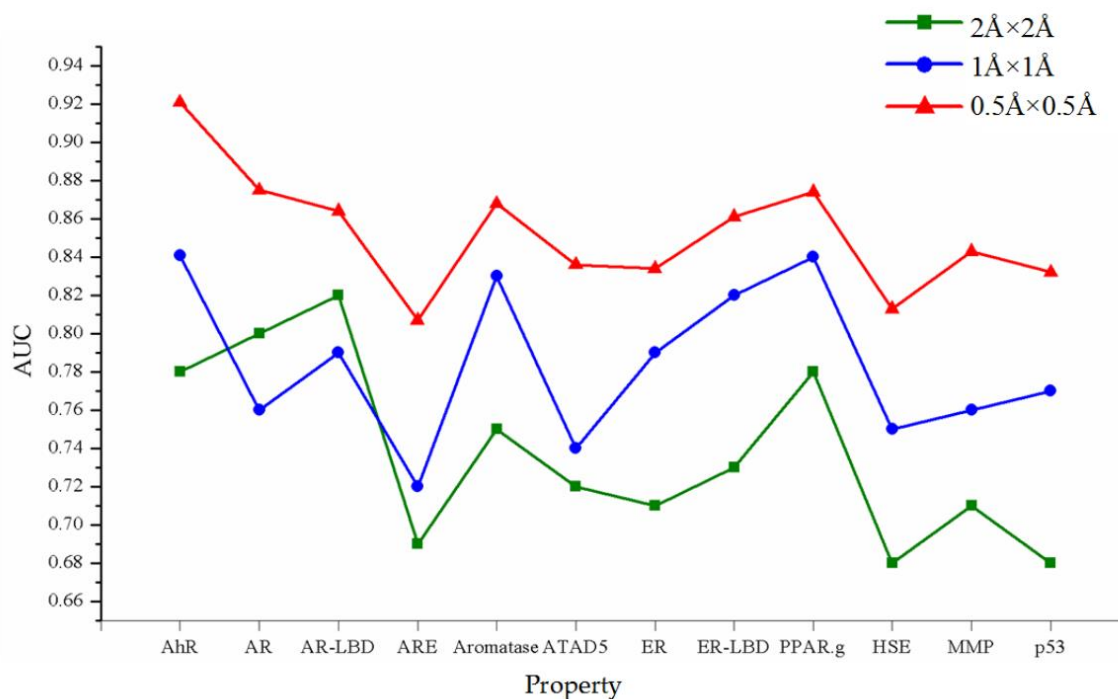
**Figure 8.** The receiver operating characteristic (ROC) curves of the 12 properties. (a) ROC curves of AhR, (b) ROC curves of AR, (c) ROC curves of AR-LBD, (d) ROC curves of ARE, (e) ROC curves of Aromatase, (f) ROC curves of ATAD5, (g) ROC curves of ER, (h) ROC curves of ER-LBD, (i) ROC curves of PPAR.g, (j) ROC curves of HSE, (k) ROC curves of MMP, (l) ROC curves of p53.



**Figure 9.** The results of combining all descriptors (brown) and removing one of the five descriptors.

### 3.5. Evaluation on Different Resolutions

Moreover, the resolution is also important for prediction accuracy. Three resolutions were tested including  $0.5\text{\AA} \times 0.5\text{\AA}$ ,  $1\text{\AA} \times 1\text{\AA}$  and  $2\text{\AA} \times 2\text{\AA}$ . The results of different resolutions are shown in Figure 10. It can be seen from Figure 10 that as the resolution value becomes higher, the performance of prediction becomes more accurate. When the resolution is  $0.5\text{\AA} \times 0.5\text{\AA}$ , the average AUC value of 12 properties is 0.84, and when the resolution is  $1\text{\AA} \times 1\text{\AA}$  and  $2\text{\AA} \times 2\text{\AA}$ , the average AUC values are 0.81 and 0.76 respectively. Therefore, the resolution of  $0.5\text{\AA} \times 0.5\text{\AA}$  is the most suitable for model training.



**Figure 10.** The results of the different resolutions. The red line represents the result of the resolution of  $0.5\text{\AA} \times 0.5\text{\AA}$ , while the yellow and green lines represent  $1\text{\AA} \times 1\text{\AA}$  and  $2\text{\AA} \times 2\text{\AA}$  respectively.

### 3.6. Comparison of Different Methods

In this section, the most common algorithms and their reported results have been introduced for comparison, including RF, SVM, multi-task learning, and several deep learning methods that appeared in the Tox21 Challenge. A random forest is a classifier that contains multiple decision trees, and the output category is determined by the mode of the category output by the individual tree [41]. The SVM is a generalized linear classifier that performs binary classification on data in supervised learning. As most of the molecules in the dataset are labeled with multiple toxic properties, a multitask classifier can be applied for prediction which is implemented in DeepChem [46]. Chemception developed the convolutional neural networks [31] to predict toxicity using 2D images of compounds without explicitly calculating chemical descriptors and achieved an average AUC 0.77. Mykola Galushka et al. utilized a pre-trained auto-encoder [32] to build classifiers to predict the toxicity property. The AMAZIZ team used consensus modeling for predictions based on the ASNN [33]. Capuzzi et al. used the DNN [34] with molecular descriptors and achieved an average AUC of 0.83 for both 12 properties. The comparison and results are summarized in Table 5.

**Table 5.** AUC results of different methods. In the methods listed in the table, the best value of each property is marked in bold. The results of all other methods except ours are obtained through their publications.

	Traditional Machine Learning				Deep Learning			
	SVM	Multitask	RF	Chemception	Auto-Encoder (MLP)	ASNN	DNN	Our Method
AhR	0.823	0.853	0.892	0.80	0.84	0.853	0.90	<b>0.921</b>
AR	0.643	0.734	0.718	0.757	0.81	0.736	0.83	<b>0.875</b>
AR-LBD	0.832	0.844	0.740	0.886	0.85	0.650	<b>0.89</b>	0.864
ARE	0.768	0.792	0.785	0.654	0.75	0.729	0.76	<b>0.807</b>
Aromatase	0.700	0.795	0.783	0.799	0.82	0.737	0.86	<b>0.868</b>
ATAD5	0.777	0.786	0.828	0.776	0.78	0.741	<b>0.85</b>	0.836
ER	0.615	0.637	0.769	0.694	0.74	0.749	0.74	<b>0.834</b>
ER-LBD	0.715	0.773	0.750	0.762	0.79	0.715	0.83	<b>0.861</b>
PPAR.g	0.644	0.741	0.777	0.819	0.69	0.785	0.70	<b>0.874</b>
HSE	0.704	0.700	0.775	0.717	0.70	0.799	0.79	<b>0.813</b>
MMP	0.830	0.822	0.926	0.755	0.85	0.904	<b>0.95</b>	0.843
p53	0.738	0.697	0.782	0.776	0.75	0.765	<b>0.84</b>	0.832
average	0.73	0.76	0.79	0.77	0.78	0.76	<b>0.83</b>	<b>0.85</b>

It could be seen from Table 5 that the proposed multi-channel grid CNN method performs best among all other methods overall. For properties of AhR, AR, ARE, Aromatase, ER, ER-LBD, PPAR.g and HSE, our method did get the first place. The average AUC values of each method were compared. Our method is approximately 0.8, 0.7, 0.9, and 0.2 higher than Chemception, Auto-encoder, ASNN, DNN methods respectively, which are also based on deep learning methods. The AUC value of the proposed method is also 0.6 higher than that of the random forest method.

## 4. Conclusions

In daily life, people are exposed to many molecules every day through food, drugs, etc. Therefore, the toxicity test of a molecule is crucial to human life. Through the Tox21 database, the nature of the 12 toxic substances can be studied which is very meaningful work. Previously, the traditional machine learning method was not considered to discover the nature of molecular data. The method proposed in this paper extracted the descriptors of the atomic type, distance, coordinate position and other information according to the molecular characteristics, and a multi-channel grid-based CNN was constructed for toxicity prediction. The result shows that the proposed method outperforms other convolution machine learning methods and deep learning methods.

However, our method is to extract the properties of the 2D grid created by the molecule. The study of the properties of a grid-based molecule can continue to be studied through other perspectives.



In future research, the authors intend to continue to expand the depth and breadth of research based on the current direction, such as introducing more accurate molecular descriptors and the grid generation method. On the other hand, other effective and fast methods of deep learning can be found which could improve the accuracy of the prediction.

**Author Contributions:** Q.Y. started the project and performed all computational analyses; Z.W. prepared the data; Z.L. planned the experiment and supervised the process of the project; X.G. analyzed the data; M.J. performed the experiments, S.W. and S.Z. explained the experimental results.

**Funding:** This work is supported by the National Nature Science Foundation of China (No. 61602430), and Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No.2018SDKJ0402).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Table A1.** Hyperparameter settings of 12 properties.

	Learning Rate	Dropout Rate	L2 Weight Decay
AhR	0.00001	25% input layer	0.0001
AR	0.000001	20% input layer	0.0001
AR-LBD	0.000001	20% input layer	0.001
ARE	0.0001	20% input layer	0.0001
Aromatase	0.00001	20% input layer	0.0001
ATAD5	0.00001	–	0.0001
ER	0.00001	–	0.0001
ER-LBD	0.00001	–	0.001
PPAR.g	0.0001	–	0.0001
HSE	0.00001	–	0.001
MMP	0.0001	25% input layer	0.0001
p53	0.0001	–	0.001

## References

- He, B.; Luo, Y.; Li, B.-K.; Xue, Y.; Yu, L.-T.; Qiu, X.-L.; Yang, T.-K. Predicting and virtually screening breast cancer targeting protein hec1 inhibitors by molecular descriptors and machine learning methods. *Acta Phys. Chim. Sin.* **2015**, *31*, 1795–1802.
- Hessler, G.; Baringhaus, K.-H. Artificial Intelligence in Drug Design. *Molecules* **2018**, *23*, 2520. [[CrossRef](#)] [[PubMed](#)]
- Laurie, A.T.R.; Jackson, R.M. Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **2005**, *21*, 1908–1916. [[CrossRef](#)] [[PubMed](#)]
- Lee, M.S.; Olson, M.A. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys. J.* **2006**, *90*, 864–877. [[CrossRef](#)] [[PubMed](#)]
- Dahl, G.E.; Jaitly, N.; Salakhutdinov, R. Multi-task neural networks for QSAR predictions. *arXiv*, 2014; arXiv:1406.1231.
- Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36. [[CrossRef](#)]
- Barnard, J.M.; Downs, G.M. Chemical Fragment Generation and Clustering Software. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 141–142. [[CrossRef](#)]
- Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [[CrossRef](#)] [[PubMed](#)]
- Morgan, H.L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113. [[CrossRef](#)]
- Awale, M.; Reymond, J.-L. Atom Pair 2D-Fingerprints Perceive 3D-Molecular Shape and Pharmacophores for Very Fast Virtual Screening of ZINC and GDB-17. *J. Chem. Inf. Model.* **2014**, *54*, 1892–1907. [[CrossRef](#)]

11. Zhao, C.; Zhang, H.; Zhang, X.; Liu, M.; Hu, Z.; Fan, B. Application of support vector machine (SVM) for prediction toxic activity of different data sets. *Toxicology* **2006**, *217*, 105–119. [[CrossRef](#)]
12. Zhang, H.; Chen, Q.-Y.; Xiang, M.-L.; Ma, C.-Y.; Huang, Q.; Yang, S.-Y. In silico prediction of mitochondrial toxicity by using GA-CG-SVM approach. *Toxicol. Vitro* **2009**, *23*, 134–140. [[CrossRef](#)] [[PubMed](#)]
13. Como, F.; Carnesecchi, E.; Volani, S.; Dorne, J.; Richardson, J.; Bassan, A.; Pavan, M.; Benfenati, E. Predicting acute contact toxicity of pesticides in honeybees (*Apis mellifera*) through a k-nearest neighbor model. *Chemosphere* **2017**, *166*, 438–444. [[CrossRef](#)] [[PubMed](#)]
14. Gupta, S.; Basant, N.; Singh, K.P. Estimating sensory irritation potency of volatile organic chemicals using QSARs based on decision tree methods for regulatory purpose. *Ecotoxicology* **2015**, *24*, 873–886. [[CrossRef](#)]
15. 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. [[CrossRef](#)] [[PubMed](#)]
16. Solimeo, R.; Zhang, J.; Kim, M.; Sedykh, A.; Zhu, H. Predicting Chemical Ocular Toxicity Using a Combinatorial QSAR Approach. *Chem. Res. Toxicol.* **2012**, *25*, 2763–2769. [[CrossRef](#)]
17. Zhang, H.; Cao, Z.-X.; Li, M.; Li, Y.-Z.; Peng, C. Novel naïve Bayes classification models for predicting the carcinogenicity of chemicals. *Food Chem. Toxicol.* **2016**, *97*, 141–149. [[CrossRef](#)] [[PubMed](#)]
18. Cruz-Monteagudo, M.; Cordeiro, M.N.D.S.; Borges, F. Computational chemistry approach for the early detection of drug-induced idiosyncratic liver toxicity. *J. Comput. Chem.* **2008**, *29*, 533–549. [[CrossRef](#)]
19. Segler, M.H.S.; Kogej, T.; Tyrchan, C.; Waller, M.P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* **2017**, *4*, 120–131. [[CrossRef](#)]
20. Qin, Q.; Feng, J. Imputation for transcription factor binding predictions based on deep learning. *PLoS Comput. Biol.* **2017**, *13*, e1005403. [[CrossRef](#)]
21. AlQuraishi, M. End-to-End Differentiable Learning of Protein Structure. *Cell Syst.* **2019**, *8*, 292–301. [[CrossRef](#)]
22. Wallach, I.; Dzamba, M.; Heifets, A. AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv*, 2015; arXiv:1510.02855.
23. 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](#)]
24. Wu, K.; Wei, G.-W. Quantitative Toxicity Prediction Using Topology Based Multitask Deep Neural Networks. *J. Chem. Inf. Model.* **2018**, *58*, 520–531. [[CrossRef](#)] [[PubMed](#)]
25. Hughes, T.B.; Miller, G.P.; Swamidass, S.J. Modeling Epoxidation of Drug-like Molecules with a Deep Machine Learning Network. *ACS Cent. Sci.* **2015**, *1*, 168–180. [[CrossRef](#)] [[PubMed](#)]
26. Karim, A.; Singh, J.; Mishra, A.; Dehzingi, A.; Newton, M.A.H.; Sattar, A. Toxicity Prediction by Multimodal Deep Learning. In *Information Technology in Bio- and Medical Informatics*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2019; pp. 142–152.
27. Liew, C.Y.; Lim, Y.C.; Yap, C.W. Mixed learning algorithms and features ensemble in hepatotoxicity prediction. *J. Comput. Mol. Des.* **2011**, *25*, 855–871. [[CrossRef](#)] [[PubMed](#)]
28. Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L. Deep Learning for Drug-Induced Liver Injury. *J. Chem. Inf. Model.* **2015**, *55*, 2085–2093. [[CrossRef](#)] [[PubMed](#)]
29. 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. [[CrossRef](#)] [[PubMed](#)]
30. Huang, R.; Sakamuru, S.; Martin, M.T.; Reif, D.M.; Judson, R.S.; Houck, K.A.; Casey, W.; Hsieh, J.-H.; Shockley, K.R.; Ceger, P.; et al. Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. *Sci. Rep.* **2014**, *4*, 5664. [[CrossRef](#)]
31. 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*, 2017; arXiv:1706.06689.
32. Galushka, M.; Browne, F.; Mulvenna, M.; Bond, R.; Lightbody, G. Toxicity Prediction Using Pre-trained Autoencoder. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; pp. 299–304.

33. 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. [[CrossRef](#)]
34. 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*, 395. [[CrossRef](#)]
35. Jiménez, J.; Skalic, M.; Martinez-Rosell, G.; De Fabritiis, G. K DEEP: Protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J. Chem. Inf. Model.* **2018**, *58*, 287–296. [[CrossRef](#)]
36. Asako, Y.; Uesawa, Y. High-Performance Prediction of Human Estrogen Receptor Agonists Based on Chemical Structures. *Molecules* **2017**, *22*, 675. [[CrossRef](#)] [[PubMed](#)]
37. Chawla, A. Nuclear Receptors and Lipid Physiology: Opening the X-Files. *Science* **2001**, *294*, 1866–1870. [[CrossRef](#)] [[PubMed](#)]
38. Huang, R.; Xia, M.; Nguyen, D.-T.; Zhao, T.; Sakamuru, S.; Zhao, J.; Shahane, S.A.; Rossoshek, A.; Simeonov, A. Tox21Challenge 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. [[CrossRef](#)]
39. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [[CrossRef](#)] [[PubMed](#)]
40. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
41. Deng, B.-C.; Yun, Y.-H.; Liang, Y.-Z.; Cao, D.-S.; Xu, Q.-S.; Yi, L.-Z.; Huang, X. A new strategy to prevent over-fitting in partial least squares models based on model population analysis. *Anal. Chim. Acta* **2015**, *880*, 32–41. [[CrossRef](#)] [[PubMed](#)]
42. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
43. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*, 2015; arXiv:1502.03167.
44. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
45. 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. [[CrossRef](#)]
46. Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V.; Leswing, K.; Wu, Z. *Deep Learning for the Life Sciences*; O'Reilly Media: Newton, MA, USA, 2019.

**Sample Availability:** Samples of the compounds are not available from the authors.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).