A spatial-temporal gated attention module for molecular property prediction based on molecular geometry

Article *in* Briefings in Bioinformatics · April 2021

DOI: 10.1093/bib/bbab078

CITATIONS
4

READS
972

5 authors, including:

li Chunyan
yunnan minzu university, china
12 PUBLICATIONS **32** CITATIONS
SEE PROFILE

Jianmin Wang
Yonsei University
11 PUBLICATIONS **57** CITATIONS
SEE PROFILE

Zhangming Niu
MindRank AI
26 PUBLICATIONS **574** CITATIONS
SEE PROFILE

Junfeng Yao
Xiamen University
91 PUBLICATIONS **599** CITATIONS
SEE PROFILE

Briefings in Bioinformatics, 00(0), 2021, 1–11

doi: 10.1093/bib/bbab078
Problem Solving Protocol

A spatial-temporal gated attention module for molecular property prediction based on molecular geometry

Chunyan Li, Jianmin Wang, Zhangming Niu, Junfeng Yao and Xiangxiang Zeng

Corresponding authors: Junfeng Yao, C613, Administrative Building, School of information, Xiamen University, Xiamen, Fujian Province, China, 361005. Tel.: 86-13799289292; E-mail: yao0010@xmu.edu.cn; Xiangxiang Zeng, Department of Computer Science, Hunan University, Hunan, China, 410086. Tel.: 86-18959216855; E-mail: xzeng@hnu.edu.cn

Abstract

Motivation: Geometry-based properties and characteristics of drug molecules play an important role in drug development for virtual screening in computational chemistry. The 3D characteristics of molecules largely determine the properties of the drug and the binding characteristics of the target. However, most of the previous studies focused on 1D or 2D molecular descriptors while ignoring the 3D topological structure, thereby degrading the performance of molecule-related prediction. Because it is very time-consuming to use dynamics to simulate molecular 3D conformer, we aim to use machine learning to represent 3D molecules by using the generated 3D molecular coordinates from the 2D structure. **Results:** We proposed Drug3D-Net, a novel deep neural network architecture based on the spatial geometric structure of molecules for predicting molecular properties. It is grid-based 3D convolutional neural network with spatial-temporal gated attention module, which can extract the geometric features for molecular prediction tasks in the process of convolution. The effectiveness of Drug3D-Net is verified on the public molecular datasets. Compared with other deep learning methods, Drug3D-Net shows superior performance in predicting molecular properties and biochemical activities. **Availability and implementation:** https://github.com/anny0316/Drug3D-Net **Supplementary Data** Supplementary data are available online at https://academic.oup.com/bib.

Key words: property prediction; geometry; attention; spatial-temporal; gate; molecule.

Chunyan Li is currently pursuing the PhD degree with the School of Informatics, Xiamen University, Xiamen, China. Her current research interests include computational intelligence and bioinformatics.
Jianmin Wang is a research assistant from the Hunan University, Changsha, China. His current research includes bioinformatics, *in silico* drug design and deep learning for drug discovery.
Zhangming Niu is the Founder CEO of MindRank AI ltd. He is ranked in the top 10 of multiple global AI competitions for healthcare and industry applications.
Junfeng Yao is a professor with the School of Informatics, Xiamen University, China. His research interests include computer graphics and intelligent algorithm research.
Xiangxiang Zeng is an Yuelu distinguished Professor with the College of Information Science and Engineering, Hunan University, Changsha, China. His main research interests include computational intelligence, graph neural networks and bioinformatics.
Submitted: 2 January 2021; **Received (in revised form):** 4 February 2021

Introduction

Molecular property prediction is a crucial task in the field of drug discovery, because it can help to determine the drug functionality. Traditionally, drug characterization is based on the experience of a chemist or pharmacist and relies on simulations and experiments to understand the properties of a drug. With the exponential expansion of data volume and the rapid development of deep learning in various fields [1–4], deep learning has also been successfully applied in drug design and virtual screening [5–7], including identification and prediction of drug properties. This data-driven computational method accelerates the development and discovery of drugs, and promotes the prediction of new drug properties.

Accurate representation of molecules is the basis of molecular property prediction. At present, the representation of drug molecules mainly focused on 1D and 2D properties, such as 1D sequence-based manner [3, 8], 2D molecular fingerprint [9, 10] and 2D graph [11–13]. However, these methods ignore the 3D structural features of the molecules, which limit the accuracy of feature learning and extraction. On the other hand, encouraged by recent successes in computer vision and computer graphics in 3D, especially point cloud processing [14–17], 3D research has also appeared in the field of medicinal chemistry. However, in 3D field, the accurate information about the 3D structure of molecules is not easy to obtain. The computer program only uses simple connectivity information to calculate the approximate three-dimensional structure. The lack of three-dimensional location information also means that so far stereochemistry is neglected by many researchers. However, In 3D space, drug molecules can produce infinite conformations because of the influence of rotating bond and other characteristics, how to specify conformer to generate 3D descriptors will be a challenge. If the chosen conformer is not the most relevant conformer of the predictive task, it will lead to poorer model performance than 2D molecular representation methods, whereas drug molecules exist in three-dimensional space, and all interactions with molecules are also carried out in three-dimensional space. Therefore, 3D molecular representation is a keystone of structure-based drug development and design [18–20]. Existing study [21] has shown that 3D structure-based models successfully predict new active molecules or binding sites, which is unpredictable by 2D-based models. However, the 3D molecular representation methods have exposed some flaws, such as under-robustness and low predictive performance.

Inspired by the remarkable achievements mentioned above and the remarkable performance of convolutional neural network (CNN) in the field of computer vision, in this paper, we proposed Drug3D-Net with spatial-temporal gated attention module, a novel deep neural network architecture of 3D drug molecular representation based on molecular geometry for prediction molecular properties. Drug3D-Net is robust because of ensure rotation invariance, and Drug3D-Net improves the accuracy of molecular property prediction. The fast convergence of the proposed model greatly reduces training time and memory usage. Firstly, we proposed spatial-temporal gated attention module for extracting molecular geometric features based on 3D molecular grid descriptor. Then, we proposed Drug3D-Net architecture to obtain 3D molecular representation. Extensive experiments demonstrated that the proposed model achieved superior performance on four molecular datasets compared with current state-of-the-art models. All in all, the main contributions of this paper are as follows:

- We proposed spatial-temporal gated attention module based on 3D grid descriptor. The proposed attention module implements spatial attention and channel attention mechanisms based on gate mechanism, respectively, which can capture the 3D structural features of molecules more effectively in molecular prediction tasks.
- We proposed a novel Drug3D-Net model that performs 3D convolution operation on the discretized molecular grid.
- Extensive experiments are conducted on different datasets to demonstrate the effectiveness of the proposed model. Drug3D-net achieved superior performance compared with the state-of-the-art deep learning models.

The rest of the article is organized as follows. Firstly, second section reviews related work. Third section presents our approaches. Then, fourth section analyzes and explains the experimental results. Finally, fifth section concludes this paper.

Related work

Molecular representation methods

1D and 2D representation methods

Simplified Molecular Input Line Entry System (SMILES) [8] is a one-dimensional representation sequence of chemical structure, which uses linear sequence of letters and numbers to extract the structural features of compounds. Extended connectivity fingerprint (ECFP) [9] is a widely used molecular fingerprint to construct quantitative structure-activity relationship (QSAR) model of chemical compounds, which take molecular 2D substructures into account for getting molecular representation at a some size. CheMixNet [10] proposed a mixed deep learning architecture to simultaneously learn molecular SMILES sequences and Molecular ACCess System (MACCS) keys for molecular properties prediction. MACCS keys is a 2D representation of molecular structure based on 166 features. FP2VEC [22] built a QSAR model to represent a compound as a set of embedding vectors using a CNN architecture, which also supports the modeling of multitask learning. C-SGEN [23] proposed a composite model, a convolutional spatial graph embedding combined with molecular fingerprints for predicting molecular properties. Its convolutional process is implemented with 1D CNN. In addition, there are many excellent works based on graph neural network to predict drug molecular properties and interactions between biomacromolecules and drug molecules. The method of attributed molecular graphs [11] based on convolutional embedding is proposed for predicting molecular property relations, such as aqueous solubility, melting point and toxicity, which also demonstrates the limitations and expansibility. Weave [12], a molecular graph convolutional model, encodes the molecule to obtain expressive molecular representation for various predictive tasks. An interpretable deep model based on graph convolutional neural network extracts molecular pharmacophores to express molecule for prediction of molecular properties [24]. It shows an interpretable method and unravels the black boxes model of neural network. The above methods are based on 1D and 2D to learn molecular representation and then used for various downstream tasks. The methods all neglect that molecules exist in 3D space, and the conformational information of molecules is quite abundant, which cannot be learned by 1D or 2D-based methods.

3D representation methods

The extended 3D fingerprint (E3FP) [21] is three-dimensional representation of molecular conformers that extend ECFP to 3D space. E3FP aggregates neighbor information of nearby atoms, even if they do not have bonds connected to each

other. Most importantly, E3FP predicts novel drug-target binding that is unpredictable by two-dimensional fingerprints. A deep learning framework 3DGCN [18] was proposed to obtain 3D molecular feature from the augmented information of bond directions based on original GCN model. MaSIF [25] proposed a conceptual framework based on a geometric deep learning method to capture fingerprints for protein-related tasks, such as protein pocket-ligand prediction, protein-protein interaction prediction. Another voxel-based methods [26–28] learn 3D representation of biological macromolecules like proteins for various downstream tasks. KDEEP [27] predicts protein-ligand binding affinity using 3D CNN. In virtual screening prediction, the validity of this method is verified. DeepDrug3D [28] characterizes and classifies binding pockets of proteins with 3D convolutional neural network by using voxelized biomolecular structures. Nevertheless, voxelization methods have not yet been applied to small molecule drugs.

Attention mechanism

Attention mechanisms are currently used in a variety of tasks, such as neural machine translation, image captioning and so on. Vaswani et al. [29] proposed transformer attention without needing any convolutional operations entirely and showed superior performance. The success of attention modeling in NLP has inspired its application in the field of computer vision in which different variants of transformer are applied to recognition tasks such as object detection and semantic segmentation. Spatial attention mechanism is a variant of transformer [30]. It explores and insights into the application of spatial attention mechanisms in deep neural networks. Convolutional block attention module (CBAM) [31] performs attention operations from two different dimensions and can be integrated into CNN architectures. CBAM shows superior performance on the ImageNet dataset. The difference between our proposed attention and CBAM is that our spatial attention module and channel attention module are in a parallel relationship, based on which, we use a gate mechanism to better integrate the contributions of different attention modules. Graph attention networks (GAT) [32] can be widely used in graph-structured data, which can assign different weights for each node without performing any expensive matrix operations or relying on prior knowledge of the graph structure. The attention coefficient represents the importance ratio between adjacent vertex. Another graph-based work about attention-augmented graph convolutional network is proposed in [33].

Methods

The key idea behind Drug3D-Net is that we focus on the spatial-temporal gated attention and 3D grid structure based on molecular geometry. Figure 1 illustrates the design of our framework, which consists of three parts: (i) the original 3D grid descriptor X_{grid} as input, (ii) several stacked 3D CNN and spatial-temporal gated attention layers and (iii) the 3D grid feature for the final molecular representation to predict various tasks. Given a preprocessed training dataset $\{\hat{\mathcal{D}}_i\}_{i=1}^{M}$ of M molecules with different number of atoms, which have been transformed from the original SMILES dataset $\mathcal{D}$ to 3D grid descriptor X_{grid} by discretizing 3D molecular conformer into a grid based on interactional energy between atoms. Then, we feed X_{grid} to 3D CNN and attention layers to learn molecular geometry feature and achieve 3D molecular representation. We further take the 3D molecular representation as the graph-level feature for predicting various tasks. Figure 2 demonstrates the comparison of 2D structure, 3D structure and voxelized grid structure for compound Amigdalin.

We first give the mathematical definition of the problem (Section Problem definition). Then, present spatial-temporal gated attention module (Section Constructing spatial-temporal gated attention module). We finally propose the 3D convolutional Drug3D-Net architecture to extract global 3D molecular features for property prediction tasks (Section Building Drug3D-Net architecture with 3D grid attention module). We will present each part of our proposed approach in detail.

Problem definition

A 3D feed-forward convolutional neural network can be defined as $\mathcal{CNN}_{3D} = (Kernel_{3D}, Channel_{3D})$, where $Kernel_{3D}$ is the 3D convolution kernel and $Channel_{3D}$ is the convolution channel. Let X_{grid} is molecular 3D grid feature, D_{grid} is 3D grid descriptor, and $\mathcal{ATT}_{grid}$ is denoted as 3D grid attention mechanism. Our problem is to learn the function: f_{grid} (map raw 3D grid features to final 3D molecular characterization) and further obtain the 3D molecular representation $X_{mol\_feature}$. The mapping relations is represented as follows:

$$\left[X_{grid};\ D_{grid}, \mathcal{CNN}_{3D}, \mathcal{ATT}_{grid}\right] \xrightarrow{f_{grid}} X_{mol\_feature} \tag{1}$$

Where $X_{grid} \in \mathcal{R}^{N_{batch} \times N_{grid} \times N_{grid} \times N_{grid} \times N_{channel}}$. The N_{batch} is the number of each batch, N_{grid} is defined grid size and $N_{channel}$ is the number of channel in 3D CNN.

Constructing spatial-temporal gated attention module

Inspired by the attention mechanism in computer vision [31], we proposed spatial-temporal gated attention mechanism for 3D CNN architecture and abbreviated it as the 3D grid attention module. Let $\mathcal{ATT}_{grid}$ be a 3D grid attention map, $\mathcal{ATT}_{xyz}$ is defined as the spatial attention map, $\mathcal{F}_{space}$ is the result of spatial attention map. Similarly, define channel attention as $\mathcal{ATT}_{channel}$, $\mathcal{F}_{channel}$ is the result of channel attention map. Pooling operation is aimed to maintain a certain invariance of features, the result of which can reduce the dimensions of features and thus reduce the number of parameters. The mean pooling (variance-pooing) computes the average (variance) of the elements existing in the region of feature map covered by a filter. While the max-pooling chooses the maximum element, which is the most prominent feature in the patch of the feature map. The fusion of these pooling methods not only allows the model to learn general features, but also highlights the obvious features of local changes. The pooling process can be summarized as follows:

$$\tilde{X}_{mean} = MeanPooling(\mathcal{NN}(X_{grid})) \tag{2}$$

$$\tilde{X}_{variance} = VariancePooling(\mathcal{NN}(X_{grid})) \tag{3}$$

$$\tilde{X}_{max} = MaxPooling(\mathcal{NN}(X_{grid})) \tag{4}$$

where $X_{grid} \in \mathcal{R}^{N_{batch} \times N_{grid} \times N_{grid} \times N_{grid} \times N_{channel}}$ is an original input feature. The $\mathcal{NN}$ is a neural network for transformation and remapping of the original feature X_{grid}. The *MeanPooling*, *VariancePooling* and *MaxPooling* are expressed as mean pooling operator, variance pooling operator and maximum pooling operator, respectively.

$$\tilde{X}_{grid} = Concatenate(\tilde{X}_{mean}, \tilde{X}_{variance}, \tilde{X}_{max}) \tag{5}$$

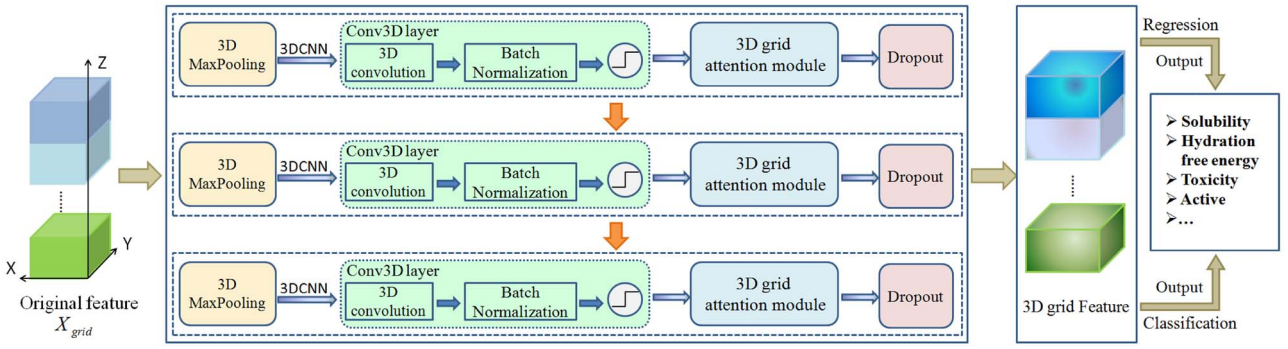

Figure 1. The architecture of 3D convolutional neural network with grid attention module (Drug3D-Net). Each 3D grid attention module is the proposed spatial-temporal gated attention. The 3D grid feature obtained from 3D grid attention module would need to reduce down to a single output node via neural network operations for regression or classification predictions.

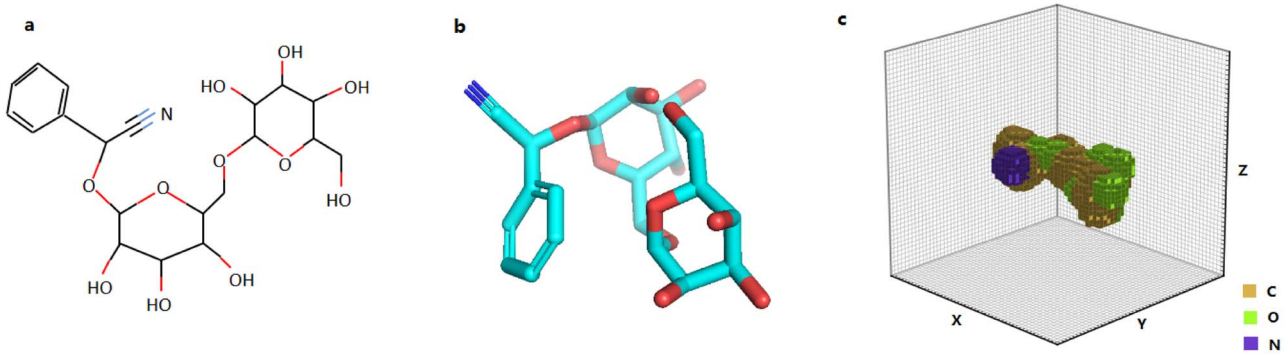

Figure 2. Three different expressions for the same compound, Amigdalin. (**A**) is the 2D molecular structure. (**B**) is the 3D molecular structure. (**C**) is the 3D voxelized expression for Amigdalin, different colors denote different atoms.

$$\mathcal{ATT}_{xyz} = \mathcal{CNN}_{3D}(\tilde{X}_{grid}) \tag{6}$$

The *Concatenate* is generally used to combine or merge features, which is manifested as an increase in dimensionality. For example, after performing the *Concatenate* operation with two 128-dimensional vectors, a 256-dimensional vector can be obtained. The $\otimes$ denotes bitwise multiplication. Thus the final spatial attention can be represented as follows:

$$\mathcal{F}_{space} = \mathcal{ATT}_{xyz} \otimes \mathcal{NN}(X_{grid}) \tag{7}$$

Then, the channel attention process can be computed:

$$\tilde{X}_{fc-mean} = \mathcal{FC}(MeanPooling(\mathcal{NN}(X_{grid}))) \tag{8}$$

$$\tilde{X}_{fc-variance} = \mathcal{FC}(VariancePooling(\mathcal{NN}(X_{grid}))) \tag{9}$$

$$\tilde{X}_{fc-max} = \mathcal{FC}(MaxPooling(\mathcal{NN}(X_{grid}))), \tag{10}$$

where '$\mathcal{FC}$' denotes full connection mapping layer, let $\tilde{X}_{mean}$ is the input of $\mathcal{FC}$ layer, then:

$$\tilde{X}_{fc-mean} = ReLU(\Sigma_i \omega^{(i)}(\tilde{X}_{mean})^{(i)} + b^{(i)}) \tag{11}$$

where $\omega^{(i)}$ is a set of weights and $b^{(i)}$ is bias, similarly, $\tilde{X}_{fc-variance}$ and $\tilde{X}_{fc-max}$ can be defined. Let '*Fusion*' represent fusion operator, such as add, subtract and multiply (element-wise), that takes the sequence of tensor as the operation object and returns a single tensor. *Fusion* is different from *Concatenate*. The 'σ; denotes sigmoid activation function, then

$$\tilde{X}_{grid} = Fusion(\tilde{X}_{fc-mean}, \tilde{X}_{fc-variance}, \tilde{X}_{fc-max}) \tag{12}$$

$$\mathcal{ATT}_{channel} = \sigma(\tilde{X}_{grid}) \tag{13}$$

The final channel attention can be represented as follows:

$$\mathcal{F}_{channel} = \mathcal{ATT}_{channel} \otimes \mathcal{NN}(X_{grid}) \tag{14}$$

The functions of spatial attention and channel attention are expressed by Equations (7) and (14), respectively. Gate mechanism can further improve the ability of feature extraction at an appropriate rate. Finally, the overall 3D grid attention mechanism can be defined as follows:

$$\mathcal{ATT}_{grid} = \alpha \otimes \mathcal{F}_{space} + (1 - \alpha) \otimes \mathcal{F}_{channel}, \tag{15}$$

where α is trainable parameter matrix. It determines the proportion of spatial attention and channel attention relative

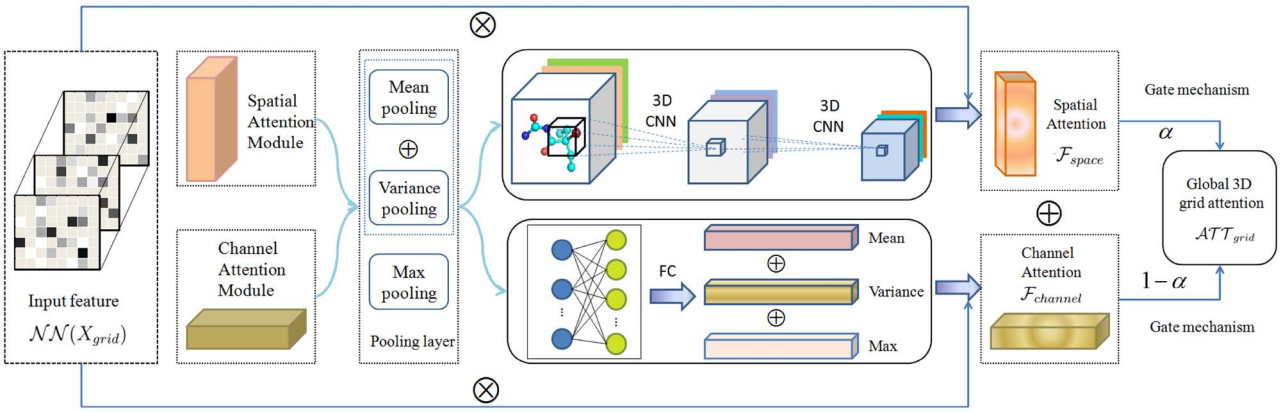

Figure 3. The overview of spatial-temporal gated attention (3D grid attention) module. The pooling operation in the pooling layer consists of mean-pooling, variance-pooling, and max-pooling. 3D grid attention module consists of spatial attention, channel attention and gate mechanism. The spatial attention module extracts 3D molecular feature and obtains spatial attention scores through 3D CNN, and the channel attention module obtains the attention scores of different channels (atomic types) through the fully connected layer. The gate mechanism is adopted to gain global 3D grid attention.

to global 3D grid attention. Figure 3 shows the components of spatial-temporal gated attention module.

Algorithm 1: Molecular geometry-based 3D grid attention algorithm

Input: dateset $\mathcal{D}$, spatial attention score *spatialScore*, channel attention score *channelScore*, 3D descriptor operation *3DGridDescriptor*, pooling *3DPooling*, 3D convolution *3DCNN*, regularization *BN*.
Parameter: layer size *layersize*, 3D grid resolution r, channels c, number of filters f, kernel size k.
Output: 3D molecular representation.

function 3DGridAttention(X_{grid})
  $\mathcal{ATT}_{xyz} \leftarrow spatialScore(X_{grid})$ ▷ Equ (6)
  $\mathcal{F}_{space} \leftarrow \mathcal{ATT}_{xyz} \otimes \mathcal{NN}(X_{grid})$
  $\mathcal{ATT}_{channel} \leftarrow channelScore(X_{grid})$ ▷ Equ (13)
  $\mathcal{F}_{channel} \leftarrow \mathcal{ATT}_{channel} \otimes \mathcal{NN}(X_{grid})$
  $\mathcal{ATT}_{grid} \leftarrow \alpha \otimes \mathcal{F}_{space} + (1-\alpha) \otimes \mathcal{F}_{channel}$
  return $\mathcal{ATT}_{grid}$
end function
$X_{ori} \leftarrow 3DGridDescriptor(\mathcal{D}, r, c)$
while $l = 1 : layersize$ **do**
  $X^{(l)} \leftarrow 3DPooling(X^{(l)}_{ori})$ ▷ Equ (2, 3, 4)
  $\tilde{X}^{(l)} \leftarrow 3DCNN(X^{(l)}, f, k)$
  $X^{(l)}_{grid} \leftarrow BN(\tilde{X}^{(l)})$
  $X^{(l)}_{result} \leftarrow$ 3DGridAttention$(X^{(l)}_{grid})$
  $X^{(l)}_{ori} \leftarrow X^{(l)}_{result}$
end while
$X_{mol\_feature} \leftarrow X^{(l)}_{result}$
return $X_{mol\_feature}$.

Building Drug3D-Net architecture with 3D grid attention module

Drug3D-Net is a grid-based 3D CNN with the spatial-temporal gated attention module to learn 3D molecular representation for prediction molecular properties. It is an end-to-end model, the input of which is the SMILES sequence of the drug molecule. Then preprocess datasets, establishing 3D molecular grid descriptor, which is fed into deep neural network model. Here, we implement 3 layers of 3D convolutional neural network with 3D grid attention module for each layer. First, the 3D max pooling operation is performed on the input features, and then the 3D convolution operation and batch normalization. Each 3D convolution is followed by the 3D grid attention module. Figure1 shows the architecture of 3D CNN with gird attention mechanism. Take a layer of 3D convolution operation as an example, regardless of the batch size, the input feature dimension is (48, 48, 48, 10), and the maximum padding operation is performed with a 'valid' padding setting and strides size of (2, 2, 2). Then, get a transition feature vector with dimensions (24, 24, 24, 10). Further, a 3D convolution operation is performed with the 'same' padding setting and strides size as (1, 1, 1), and the convolutional filter is set to 32. In this way, a transitional feature vector with the dimension (24, 24, 24, 32) is obtained. After 3D convolution, the 3D grid attention module is applied to obtain a more accurate representation of topological molecules. Finally, the learned molecular representation can be obtained for predicting molecular properties and biochemical activities. This architecture of proposed Drug3D-Net is shown in Figure 1. Figure 3 shows the implementation details of each 3D grid attention module in Figure 1. The molecular geometry-based 3D grid attention algorithm for 3D molecular representation in Drug3D-Net is summarized as Algorithm 1.

Experiments

Extensive experiments have been conducted using four different property datasets, which contain regression tasks and classification tasks respectively for fully evaluating the performance of Drug3D-Net. In this section, we describe the datasets, preprocessing, baselines and the experimental results.

Dataset description

We evaluate Drug3D-Net based on several commonly molecular datasets recommended by MoleculeNet [34], following which the split methods and metrics for each dataset are adopted in our experiment (Table 1). The details of these datasets are as follows:

- **ESOL.** ESOL dataset [35] contains SMILES sequence and water solubility (log solubility in moles per liter) data of

Table 1. The description of the datasets, the splits and metrics recommended by MoleculeNet [34]

Dataset	Data type	Number of tasks	Task type	Number of compounds	Split	Metric
ESOL	SMILES	1	Regression	1128	Random	MAE,RMSE
FreeSolv	SMILES	1	Regression	642	Random	MAE,RMSE
Tox21	SMILES	12	Classification	7831	Random	AUROC,AUPRC
HIV	SMILES	1	Classification	41 127	Scaffold	AUROC,AUPRC

1128 drug molecules. The dataset also includes other fields, such as, compound id, minimum degree, molecular weight, number of H-Bond donors, number of rings, number of rotatable bonds and polar surface area. ESOL dataset is often used for regression task in neural network. However, ESOL does not include the geometric structure information of molecule.

- **FreeSolv.** The Free Solvation Database (FreeSolv) provides calculated and experimental hydration free energy of 642 small molecules in water [36]. The values are obtained using molecular dynamics simulations.
- **Tox21.** The 'Toxicology in the 21st Century' (Tox21) is a dataset for measuring the toxicity of compounds, which contains toxicity information for 7831 drug molecules on 12 different targets. Tox21 has been used in the 2014 Tox21 Data Challenge [37].
- **HIV.** The HIV dataset, introduced by the Drug Therapeutics Program AIDS Antiviral Screen, tests the ability to inhibit HIV replication for over 40 000 compounds [38]. Screening results include three categories: confirmed inactive (CI), confirmed active (CA) and confirmed moderately active (CM). When HIV is used as a classification task for prediction, CA and CM are regarded as active, and CI is regarded as inactive. As a result, there is a serious imbalance in the number of positive and negative samples in HIV dataset: a total of 1443 active compounds and 39 684 inactive compounds.

Dataset preprocessing

Cleaning data to exclude invalid molecules and distinguish isomeric SMILES. As to the classification task datasets, such as Tox21 and HIV, in the preprocessing stage, we show the distribution of the number of atoms for all molecules and the difference between positive and negative samples (Figures 4 and 5). The number of atoms in different molecules varies greatly with a minimum of 1 and a maximum of 132 for Tox21, with a minimum of 2 and a maximum of 222 for HIV. A molecule with a large number of atoms has a richer topology structure, which can be more conducive to obtain geometric information for training model. On the other hand, we use RDKit [39] to generate molecular conformers based on the topological structure information. Then, we minimize conformers energy with the Universal Force Field [40] by setting the maximum number of iterations. Finally, choose the conformation with the best energy as the 3D topological expression of this molecule. Furthermore, discretize the selected molecular conformer into a 3D grid, which is defined by $48 \times 48 \times 48$ grid points and 0.5Å resolution. Considering the difference in the contribution of different atoms in the same molecule to molecule-level characterization, we set each channel of the 3D grid to a different atom type, which is similar to RGB channels in images. We define 10 atom types including carbon, nitrogen, oxygen, boron, fluorine, phosphorus, sulfur, chlorine, bromine and iodine. Namely, the count of channels in 3D grid is 10. We randomly rotate and translate molecules to ensure rotation invariance. We use a simple Boolean as the atom density representation. It is a 'hard' discrete boolean representation, which constitutes the primitive 3D molecular gird feature X_{grid}. Figure 6 illustrates different sizes of voxelized grid, which are set to 48 and 64, respectively. Obviously, the larger the grid size, the denser the voxels of the structure. When the grid size is low, the grid is coarse, but, the voxels for the structured expression still have a continuous shape.

Baselines

The Drug3D-Net model is based on 3D molecular spatial structure for property prediction tasks. In this section, we focus on 3D molecular representation methods, including conventional methods and graph-based methods as the baselines for comparison.

- **Random Forests.** Random forests (RF) are ensemble prediction methods [41]. It consists of many decision trees, and the prediction results of these decision trees are averaged as the final prediction result of RF.
- **Multitask network.** Multitask network [42] shares the processing of all tasks in a dataset, and then uses different classifiers or regressors to predict different tasks in the final prediction stage.
- **XGBoost.** XGBoost [43] is an implementation version of gradient boosting [44], which is another ensemble method, combining all of subtrees together into an additive model.
- **KernelSVM.** KernelSVM is based on support vector machine (SVM), which is one of the most famous and widely used machine learning method [45]. Regularization and a radial basis function kernel are incorporated into SVM for increasing performance.
- **CheMixNet-ECFP.** CheMixNet was proposed in [10], which learned molecular properties based on SMILES and molecular fingerprints representations. In this architecture, LSTM is used to learn SMILES sequence, and MLP is used to learn molecular MACCS fingerprints. Then concatenate the results of LSTM and MLP for final properties prediction. In this experiment, we only consider the comparison of molecular ECFP fingerprints by using CheMixNet architecture and ignore the molecular SMILES sequence. This is a variant of CheMixNet, which is labeled as CheMixNet-ECFP.
- **Smi2Vec.** Smi2Vec*-LSTM and Smi2Vec-BiGRU [3] are used to single- and multi-task classification in drug discovery area by transforming SMILES to vectors, that is fed to LSTM and BiGRU neural networks for training. Sim2Vec achieved superior performances on some benchmark datasets.
- **All SMILES VAE.** The All SMILES VAE [46] encodes molecules with multiple SMILES strings to learn a bijective mapping between molecules and latent representations near the high-probability-mass subspace of the prior.

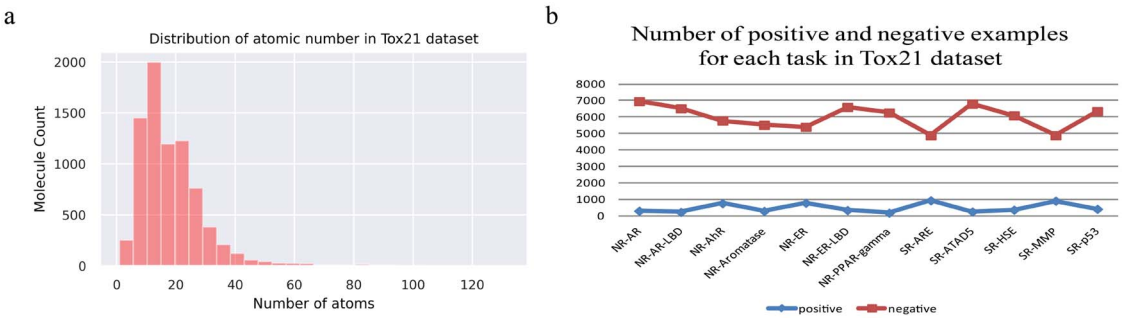

Figure 4. The characterization of Tox21 dataset. (**A**) The distribution of atomic number for all twelve tasks in Tox21 dataset. The minimum number of atoms is 1, and the maximum number of atoms is 132. (**B**) The difference between number of positive and negative samples for each task in Tox21 dataset.

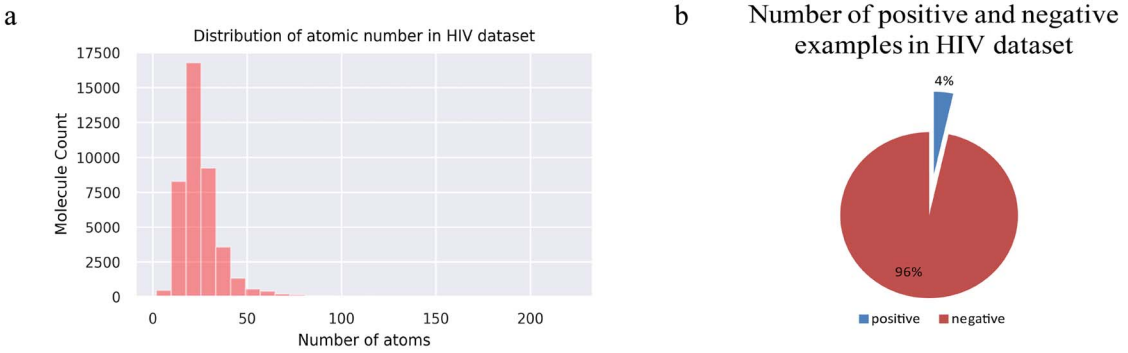

Figure 5. The characterization of HIV dataset. (**A**) The distribution of atomic number in HIV dataset with the minimum number of atoms is 2 and the maximum number is 222. (**B**) The difference between number of positive and negative samples in HIV dataset.

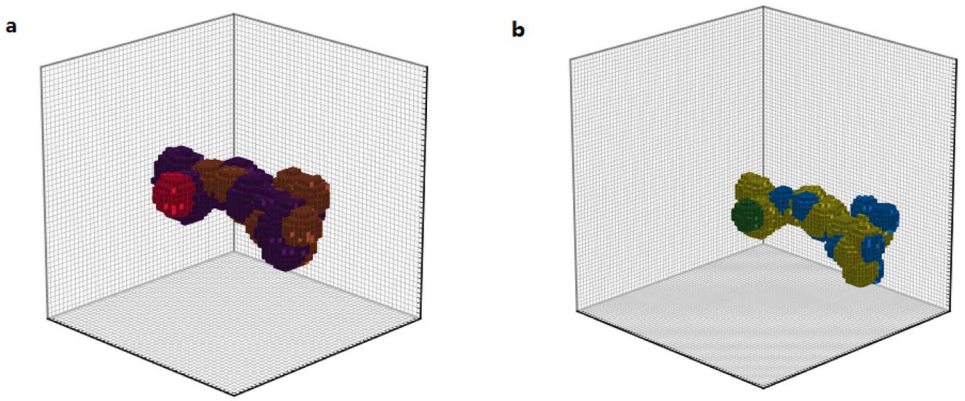

Figure 6. Illustration of Amigdalin compound for different size of grid. (**A**) The grid size is set to 48, the grid is coarse. (**B**) The grid size is set to 64, the grid is denser than the grid with size 48.

- **GC.** Graph convolutional (GC) models [13] is the neural graph fingerprints. It is based on circular fingerprints for obtaining molecular feature. GC is an implementation of circular fingerprints ECFP using graph neural network.
- **Weave.** The Weave [47] is a graph-based model that regards each molecule as an undirected graph for taking greater advantage of information in the graph structure.
- **E3FP.** E3FP was proposed in [21], that obtains 3D molecular fingerprint based on molecular space conformation information. E3FP reports novel drug-target binding predictions inaccessible by 2D fingerprints. We choose the 3D molecular conformation with the best energy to generate molecular fingerprints for comparison.
- **3D voxelized PLI.** 3D protein-ligand interaction (PLI) was proposed in [26]. It described a CNN scoring function, which can automatically learn the key features of protein-ligand interaction. For convenience, a simplified version of PLI is used to evaluate performance. PLI is a voxelized representation learning method, which is abbreviated as 3D voxelized PLI, here.
- **Drug3D-Net (CBAM).** It is the implementation of CBAM attention module [31] using Drug3D-Net architecture with data enhancement. The purpose is to compare the performance difference between CBAM and 3D grid attention proposed in this paper with the same network architecture.

- **Drug3D-Net.** This is the method presented in this paper with 3D grid attention consisted of spatial attention, channel attention and gate mechanism. Drug3D-Net pays more attention to the geometric structure information of molecule in 3D space and attempts to provide an insight into the geometric features of molecules.

Experimental setup

Drug3D-Net is trained based on Keras framework, with Tensorflow [49] as the backend and Adam [50] as the optimizer to apply gradient back-propagation algorithm. The used datasets are divided into training set, validate set and test set with the ratio of 8:1:1 following the splits recommended by MoleculeNet [34] for different datasets (Table 1). The experiments implemented 10-fold cross-validation, batch normalization, L2 regularized method and dropout ($dropout\_ratio = 0.5$) for avoiding overfitting. A learning rate of 0.001 was set with learning rate decay using the ReduceLROnPlateau ($factor = 0.9, patience = 10, min\_lr = 0.0005$) in Keras. A total of 300 epochs were set with early stopping setting. In addition, for regression tasks, we adopt mean squared error to be the training loss, mean absolute error (MAE) and root mean square error (RMSE) to evaluate the performance of Drug3D-Net. And for classification tasks, cross-entropy is used as the loss function and the average area under the receiver operating characteristics curve (AUROC) and area under precision-recall curve (AUPRC) are used to evaluate the overall performance of the proposed model. Because the Tox21 and HIV datasets are highly imbalanced with most samples being negative, in order to ensure the rotation invariance of 3D molecules to increase the robustness of the model, we randomly rotate the positive samples 5, 10, 15 or 20 times for different datasets and tasks. The count of rotation depends on the proportion of the positive examples in all samples. In the end, the ratio of positive and negative samples after data augmentation is about between $\frac{1}{2}$ and $\frac{2}{3}$. The change in the ratio of positive and negative samples due to data enhancement results in a higher AUPRC in Tox 21 dataset and HIV dataset compared with other 1D or 2D deep learning methods. As for the Tox21 dataset, the overall performance is averaged over all 12 tasks. Additionally, the MolGridDescriptorLayer of our model has the ability to randomly rotate and translate the input molecular structures on-the-fly, which is controlled by using hyperparameters to set whether to rotate or translate the input structures, and the count of rotations as well.

Result

The 10-fold cross-validation performances of Drug3D-Net are shown in Tables 2–5 for ESOL, FreeSolv, Tox21 and HIV dataset, respectively. The smaller the value, the better for ESOL and FreeSolv dataset. As to Tox21 and HIV dataset, larger is better. Drug3D-Net achieves the best performance in Table 2 and outperforms 2D molecular fingerprints ECFP, 3D molecular fingerprints E3FP and 3D voxelized PLI model with a large margin. In this experiment, the predictive MAE and RMSE have been multiplied by the normalized variance of the whole dataset. Furthermore, we take into account the prediction implemented by the variant of Drug3D-Net with CBAM attention module. It demonstrates that the performance of 3D grid attention module proposed in this paper is better than that of CBAM. These performance evalutation results are also shown in Table 2. Table 3 shows the prediction performance on the FreeSolv dataset. Compared with other baselines, graph convolutional model (GC) achieves better performance. This may be caused by the lack of abundant molecular structural features, due to the small FreeSolv dataset. However, our model obtains almost equivalent predictive performance compared with the GC model and exceeds conventional machine learning methods such as RF, Multitask and XGBoost. From Tables 4 and 5, Drug3D-Net has achieved competitive performance compared to conventional machine learning methods, graph-based methods and grid-based 3D molecular representation methods. The high performance obtained by our model mainly owes to the proposed 3D grid attention mechanism and data augmentation. The data augmentation by rotating positive molecules can reduce the gap between positive and negative samples, enrich the sample structure, reduce overfitting and greatly improve the predictive ability of our model. We found it only takes a few epochs for our model to converge, which greatly reduces training time and memory usage.

In addition, Table 6 shows the predictive performances for each task in Tox21 dataset with AUROC scores, compared to the RF, Smi2Vec*-LSTM and Smi2Vec-BiGRU models. Our proposed model achieves superior performance in the prediction of each task and exceeds the baselines by a large margin. Furthermore, Table S1 shows AUROC and AUPRC scores for twelve tasks in Tox21 with 3D grid attention and without attention module. It shows that the spatial-temporal gated attention module proposed in this paper plays a key role in the task of molecular properties prediction. Furthermore, Figures S1–S2 more intuitively shows histograms to compare the performance of different models on the ESOL, Tox21, FreeSolv and HIV datasets. Figure S3 (a) shows the AUROC and AUPRC scores in each task of Tox21 with 3D grid attention and without attention, and (b) is the performance comparison of the twelve tasks for different models on the Tox21 dataset.

Discussion and conclusion

In this work, we aimed to construct the Drug3D-Net architecture based on 3D CNN with the spatial-temporal gated attention module. Drug3D-Net is a molecular geometry-based deep neural network architecture to learn 3D drug molecular representation for property prediction tasks. First of all, we proposed spatial-temporal gated attention module based on 3D molecular descriptor. Then, we performed 3D CNN operations for extracting spatial structure features of molecules. Finally, we proposed Drug3D-Net to predict the molecular properties, which further validates the validity and accuracy of 3D molecular representation. Our model has been trained on benchmark datasets in the field of Biochemistry, and the results showed that Drug3D-net has achieved competitive performance compared with the state-of-the-art deep learning models used in chemistry. Consistent with the limitation of voxel-based 3D CNN methods, our method also has high computational overhead due to convolve over large regions of empty space. However, the proposed voxel-based spatial-temporal gated attention mechanism obtains superior performance. Our model is robust and can be applied not only to molecular property prediction, but also to other molecule-related tasks, such as DDI prediction and CPI prediction. It could be also applicable for large molecules like proteins. Especially when the ratio of positive and negative samples is imbalanced, our model performs best.

In general, the research on the 3D molecular structure is rarely compared with the research on the 2D structure. A clear limitation of most 3D structured deep learning in chemistry is

Table 2. Predictive performances for ESOL dataset with MAE mean absolute error and RMSE root mean square error (the smaller the value, the better). The Best results are highlighted in bold. ⋆ The values of the starred models are taken from the ref [34]

Model	MAE		RMSE	
	Validation	Test	Validation	Test
⋆ RF	–	–	1.16	1.07
⋆ Multitask	–	–	1.17	1.12
⋆ XGBoost	–	–	1.05	0.99
CheMixNet-ECFP	0.7866	0.8571	0.9966	1.1719
⋆ GC	-	-	1.05	0.97
E3FP	0.8770	0.9001	1.1168	1.1886
3D voxelized PLI	1.0796	1.1147	1.3310	1.3800
Drug3D-Net (CBAM)	0.9088	0.8127	1.1426	1.0199
Drug3D-Net	**0.7823**	**0.7841**	**0.9746**	**0.9683**

Table 3. Predictive performances for FreeSolv dataset with MAE and RMSE mean absolute error and root mean squard error (the smaller the value, the better). The best results are highlighted in bold

Model	MAE		RMSE	
	Validation	Test	Validation	Test
⋆ RF	-	-	2.12	2.03
⋆ Multitask	-	-	1.95	1.87
⋆ XGBoost	-	-	1.76	1.74
⋆ GC	-	-	**1.35**	**1.40**
Drug3D-Net	1.1158	1.1598	1.4416	1.4709

⋆ The values of the starred models are taken from the ref [34].

Table 4. Predictive performances for Tox21 dataset with AUROC and AUPRC scores (the larger the value, the better). The Best results are highlighted in bold. ⋆ The values of the starred models are taken from the ref [34]. The values marked by ◦ and • are the best performances among conventional methods and graph-based methods in MoleculeNet [34], respectively

Model	AUROC		AUPRC	
	Validation	Test	Validation	Test
⋆ RF	0.7630	0.7690	–	–
⋆ Multitask	0.7950	0.8030	–	–
⋆ XGBoost	0.7750	0.7940	–	–
◦ KernelSVM	0.8180	0.8220	–	–
• GC	0.8250	0.8290	–	–
⋆ Weave	0.8280	0.8200	–	–
All SMILES VAE	–	0.8750	–	–
3D voxelized PLI	0.8723	0.8797	0.8886	0.8872
Drug3D-Net (CBAM)	0.9016	0.8933	0.9104	0.9011
Drug3D-Net	**0.9572**	**0.9525**	**0.9555**	**0.9524**

Table 5. Predictive performances for HIV dataset with AUROC and AUPRC scores (the larger the value, the better). The Best results are highlighted in bold. ⋆ The values of the starred models are taken from the ref [34]. The values marked by ◦ and • are the best performances among conventional methods and graph-based methods in MoleculeNet [34], respectively

Model	AUROC		AUPRC	
	Validation	Test	Validation	Test
⋆ Multitask	0.7110	0.6980	–	–
⋆ XGBoost	0.8410	0.7560	–	–
◦ KernelSVM	0.8370	0.7920	–	–
• GC	0.7920	0.7630	–	–
⋆ Weave	0.7420	0.7030	–	–
3D voxelized PLI	0.9055	0.8987	0.9146	0.9078
Drug3D-Net	**0.9635**	**0.9621**	**0.9628**	**0.9617**

Table 6. Predictive performances for each task in Tox21 dataset with AUROC scores (the larger the value, the better). The Best results are highlighted in bold. ⋆ The values of the starred models are taken from the ref [3]

Task Model	⋆ RF		⋆ Smi2Vec*-LSTM		⋆ Smi2Vec-BiGRU		Drug3D-Net	
	Validation	Test	Validation	Test	Validation	Test	Validation	Test
NR-AR	0.6730	0.6732	0.6909	0.6914	0.7713	0.7114	**0.9731**	**0.9708**
NR-AR-LBD	0.5825	0.6384	0.7228	0.7477	0.8442	0.8243	**0.9716**	**0.9689**
NR-AhR	0.6076	0.5980	0.6698	0.6780	0.8751	0.8793	**0.9415**	**0.9428**
NR-Aromatase	0.5798	0.5500	0.4991	0.4964	0.8241	0.6985	**0.9732**	**0.9647**
NR-ER	0.5433	0.5507	0.5546	0.6231	0.7085	0.7360	**0.9263**	**0.9234**
NR-ER-LBD	0.5931	0.5170	0.5256	0.5308	0.7922	0.8675	**0.9651**	**0.9619**
NR-PPAR-gamma	0.4984	0.5263	0.5000	0.5659	0.7085	0.7494	**0.9764**	**0.9704**
SR-ARE	0.5562	0.5568	0.5901	0.6414	0.7945	0.7611	**0.9202**	**0.9034**
SR-ATAD5	0.5356	0.5348	0.5171	0.5000	0.7909	0.7632	**0.9623**	**0.9578**
SR-HSE	0.5107	0.5124	0.6381	0.6120	0.7540	0.7845	**0.9669**	**0.9601**
SR-MMP	0.6809	0.6862	0.7438	0.7425	0.8846	0.8599	**0.9457**	**0.9422**
SR-p53	0.5310	0.5138	0.5149	0.5180	0.7896	0.7321	**0.9644**	**0.9638**

that they require precise 3D structure or conformation information, which will introduce a certain amount of noise depending on the chosen tools, methods and the structure of conformer. However, how to choose structured information from numerous molecular 3D conformations will be a challenge. We have taken a step in the research of 3D molecules. Our next work will continue to focus on the research of 3D molecular representation, 3D pretrain based on self-supervised methods and other neural system [51, 51].

Key Points

- We proposed spatial-temporal gated attention mechanism based on 3D molecular descriptor, which can capture the 3D structural features of molecules more effectively in molecular prediction tasks.
- We proposed a novel Drug3D-Net model that performs 3D convolution operation on the discretized molecular grid.
- A large number of experiments conducted have verified the effectiveness of our model and Drug3D-Net shows superior performance in predicting molecular properties and biochemical activities.

Supplementary Data

Supplementary data are available online at *Briefings in Bioinformatics*.

Acknowledgments

We thank anonymous reviewers for valuable suggestions.

Funding

This work has been supported by the Scientific Research Fund of Yunnan Provincial Department of Education (No. 2020J0362), the National Natural Science Foundation of China (No. 62072388), the Natural Science Foundation of the Science Technology Bureau of Fujian Province in 2019 (No. 2019J01601, 2019J01001), the creation fund project of Science Technology Bureau of Fujian Province in 2019 (No. 2019C0021), the Science Technology Bureau of Xiamen Municipal Government in 2018(No.3502Z20184058) and the Foreign Cooperation Project of Science Technology Bureau of Fujian Province in 2018(No.2018I0015).

Conflicts of Interest

The authors have no conflicts of interest to declare.

Data Availability

The datasets and source code are publicly available at https://github.com/anny0316/Drug3D-Net.

References

1. Zeng X, Zhu S, Liu X, *et al*. deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics* 2019; **35**(24): 5191–8.
2. Ryu S, Kwon Y, Kim WY. A bayesian graph convolutional network for reliable prediction of molecular properties with uncertainty quantification. *Chem Sci* 2019; **10**: 8438–46.
3. Xuan L, Zhe Q, Zhi-Jie W, *et al*. A novel molecular representation with BiGRU neural networks for learning atom. *Brief Bioinform* 2019;bbz125.
4. Li C, Liu H, Qian H, *et al*. A novel computational model for predicting microRNA-disease associations based on heterogeneous graph convolutional networks. *Cell* 2019; **8**(9): 977.
5. Lim J, Ryu S, Park K, *et al*. Predicting drug-target interaction using a novel graph neural network with 3D structure-embedded graph representation. *J Chem Inf Model* 2019; **59**(9): 3981–8.
6. Deac A, Huang Y-H, Veličković P, *et al*. Drug-drug adverse effect prediction with graph co-attention. *ArXiv* 2019.
7. Xiong Z, Wang D, Liu X, *et al*. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J Med Chem* 2019.
8. David W. Smiles: a chemical language and information system. *J Chem Inf Comput Sci* 1988; **28**(1): 31–6.

9. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010; **50**(5): 742–54.
10. Paul A, Jha D, Al-Bahrani R, *et al*. Chemixnet: mixed DNN architectures for predicting chemical properties using multiple molecular representations. *Proc Nips Workshop on Machine Learning for Molecules and Materials* 2018.
11. Coley CW, Barzilay R, Green WH, *et al*. Convolutional embedding of attributed molecular graphs for physical property prediction. *J Chem Inf Model* 2017; **57**(8): 1757–72.
12. Kearnes S, Mccloskey K, Berndl M, *et al*. Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des* 2016; **30**(8): 595–608.
13. Duvenaud DK, Maclaurin D, Iparraguirre J, *et al*. Convolutional networks on graphs for learning molecular fingerprints. *Proc. NIPS* 2015;2224–32.
14. Qi CR, Yi L, Hao S, *et al*. Pointnet++: deep hierarchical feature learning on point sets in a metric space. *Proc NIPS* 2017.
15. Wang Y, Sun Y, Liu Z, *et al*. Dynamic graph cnn for learning on point clouds. *ArXiv* 2018.
16. Zhang Y, Rabbat M. A graph-cnn for 3d point cloud classification. *Proc ICASSP* 2018.
17. Zhiheng K, Ning L. Pyramnet: point cloud pyramid attention network and graph embedding module for classification and segmentation. *ArXiv* 2019.
18. Cho H, Choi IS. Three-dimensionally embedded graph convolutional network (3DGCN) for molecule interpretation. *ArXiv* 2018.
19. Denis K, Daniil P, Artur K, *et al*. 3D molecular representations based on the wave transform for convolutional neural networks. *Mol Pharm* 2018; **15**(10): 4378–85.
20. Sieg J, Flachsenberg F, Rarey M. In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening. *J Chem Inf Model* 2019; **59**(3): 947–61.
21. Axen SD, Roth, Bryan L, Huang X-P, *et al*. A simple representation of three-dimensional molecular structure. *J Med Chem* 2017; **60**(17): 7393–409.
22. Jeon W, Kim D. FP2VEC: a new molecular featurizer for learning molecular properties. *Bioinformatics* 2019; **35**(23): 4979–85.
23. Wang X, Li Z, Jiang M, *et al*. FP2VEC: a new molecular featurizer for learning molecular properties. *J Chem Inf Model* 2019; **59**(9): 3817–28.
24. Preuer K, Klambauer G, Rippmann F, *et al*. Interpretable deep learning in drug discovery. *ArXiv* 2019.
25. Gainza P, Sverrisson F, Monti F, *et al*. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods* 2020.
26. Ragoza M, Hochuli J, Idrobo E, *et al*. Protein-ligand scoring with convolutional neural networks. *J Chem Inf Model* 2017; **57**(4): 942–57.
27. Jiménez J, Škalič M, Martínez-Rosell G, *et al*. Fabritiis. KDEEP: protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. *Chem Info Model* 2018; **58**(2): 287–96.
28. Pu L, Govindaraj RG, Lemoine JM, *et al*. Deepdrug3d: classification of ligand-binding pockets in proteins with a convolutional neural network. *PLoS Comput Biol* 2019; **15**(2): 1–23.
29. Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. *Proc NIPS* 2017.
30. Zhu X, Cheng D, Zhang Z, *et al*. An empirical study of spatial attention mechanisms in deep networks. *ArXiv* 2019.
31. Woo S, Park J, Lee J-Y, *et al*. CBAM: convolutional block attention module. *Proc ECCV* 2018.
32. Veličković P, Cucurull G, Casanova A, *et al*. Graph attention networks. *ArXiv* 2017.
33. Ryu S, Lim J, Hong SH, *et al*. Deeply learning molecular structure-property relationships using attention- and gate-augmented graph convolutional network. *ArXiv* 2018.
34. Wu Z, Ramsundar B, Feinberg EN, *et al*. Moleculenet: a benchmark for molecular machine learning. *Chem Sci* 2018; **9**(2): 513–30.
35. Delaney JS. Esol: estimating aqueous solubility directly from molecular structure. *J Chem Inf Comput Sci* 2004; **44**(3): 1000.
36. Mobley DL, Guthrie JP. Freesolv: a database of experimental and calculated hydration free energies, with input files. *J Comput Aided Mol Des* 2014; **28**(7): 711–20.
37. Tox21 challenge, 2014.
38. Aids antiviral screen data, accessed 2019-11-25.
39. Landrum G. *Rdkit: Open-source cheminformatics*, 2010.
40. Rappé AK, Casewit CJ, Colwell KS, *et al*. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Am Chem Soc* 1992; **114**(25): 10024–35.
41. Breiman L. Random forests. *Machine Learning* 2001; **45**:5–32.
42. Ramsundar B, Kearnes S, Riley P, *et al*. Massively multitask networks for drug discovery. *ArXiv* 2015.
43. Chen T, Guestrin C. *Xgboost: A scalable tree boosting system*, 2016, 785–94.
44. Greedy function approximation. A gradient boosting machine. *Annals of Statistics* 2001; **29**(5): 1189–232.
45. Cortes C, Vapnik V. Support-vector networks. *Machine Learning* 1995; **20**:273–97.
46. Alperstein Z, Cherkasov A, Rolfe JT. All smiles variational autoencoder. *ArXiv* 2019.
47. Kearnes S, Mccloskey K, Berndl M, Pande V, Riley P. Molecular graph convolutions: moving beyond fingerprints. *Journal of Computer-Aided Molecular Design* 2016; **30**(8):595–608.
48. Abadi M, Barham P, Chen J, *et al*. Tensorflow: A system for large-scale machine learning. In: *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. USA: USENIX Association, 2016, 265–83.
49. Kingma D, Ba J. Adam: a method for stochastic optimization. In: *Proc. ICLR*, 2014.
50. Song T, Pang S, Hao S, *et al*. A parallel image skeletonizing method using spiking neural p systems with weights. *Neural Proc Lett* 2019; **50**(2): 1485–502.
51. Song T, Zeng X, Zheng P, *et al*. A parallel workflow pattern modeling using spiking neural p systems with colored spikes. *IEEE Trans Nanobioscience* 2018; **17**(4): 474–84.