JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Cite This: J. Chem. Inf. Model. 2019, 59, 3981−3988

Article

pubs.acs.org/jcim

Predicting Drug−Target Interaction Using a Novel Graph Neural Network with 3D Structure-Embedded Graph Representation

Jaechang Lim,[†] Seongok Ryu,[†] Kyubyong Park,[‡] Yo Joong Choe,[§] Jiyeon Ham,[‡] and Woo Youn Kim*[,†,⊥]

[†]Department of Chemistry, KAIST, Daejeon 34141, South Korea

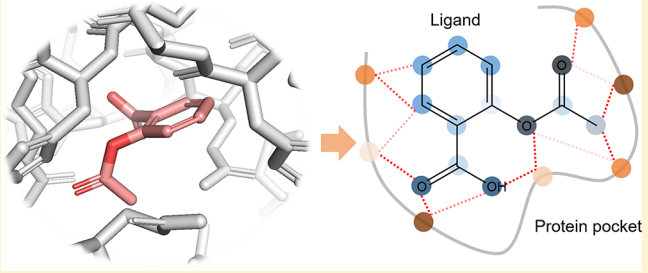
[‡]Kakao Brain, Pangyo, Gyeonggi-do 13494, South Korea

[§]Kakao, Pangyo, Gyeonggi-do 13494, South Korea

[⊥]KI for Artificial Intelligence, KAIST, Daejeon 34141, South Korea

Downloaded via JOHNS HOPKINS UNIV on May 20, 2021 at 15:40:27 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ABSTRACT: We propose a novel deep learning approach for predicting drug−target interaction using a graph neural network. We introduce a distance-aware graph attention algorithm to differentiate various types of intermolecular interactions. Furthermore, we extract the graph feature of intermolecular interactions directly from the 3D structural information on the protein−ligand binding pose. Thus, the model can learn key features for accurate predictions of drug−target interaction rather than just memorize certain patterns of ligand molecules. As a result, our model shows better performance than docking and other deep learning methods for both virtual screening (AUROC of 0.968 for the DUD-E test set) and pose prediction (AUROC of 0.935 for the PDBbind test set). In addition, it can reproduce the natural population distribution of active molecules and inactive molecules.



INTRODUCTION

Accurate prediction of drug−target interaction (DTI) is essential for in silico drug discovery. Thanks to the revolutionary advancement of theories and computing power, computational methods such as molecular dynamics and quantum mechanics/molecular mechanics can be used for reliable prediction of the binding affinity between a ligand and a protein.[1,2] Despite their high accuracy, however, huge computational costs impede their routine usage in high-throughput screening. On this ground, molecular docking has been used to predict the binding affinity with affordable computational costs.[3−10] The molecular docking greatly speeds up computations through principled parameter fitting, but its accuracy is substantially low as a result of trade-off for the high speed.

Recently, deep learning techniques[11] have attracted much attention as a promising alternative to the above principle-based approaches. While various machine learning methods have already been utilized to improve the performance of DTI calculations, deep learning has clear advantages over the other machine learning methods. First of all, it can automatically extract task-related features directly from data without handcrafted features or rules. In addition, the high expressive power of deep neural networks enables efficient training with a large amount of data. Such advantages are particularly important in applications, for instance, drug discovery, where the amount of data is exponentially growing.[12]

Various DTI models based on deep learning have been developed; each uses different deep neural network architectures and representations of protein and ligand structures. Some models do not use explicit binding structures between proteins and ligands. Instead, both proteins and ligands are represented with independent vectors, and the two vectors are integrated within a deep neural network.[13−18] Karimi et al. developed a DTI model, named DeepAffinity, based on a recurrent neural network (RNN) by representing a protein with amino acid sequence and a molecule with the SMILES.[13] Gao et al. augmented protein sequence information with gene ontology data and represented ligands with molecular graphs. They also used a RNN for proteins but adopted a graph convolutional neural network for ligands. Furthermore, they incorporated attention algorithm in integrating the vectors of a protein and a ligand, so they could analyze how their model learns protein−ligand interactions. Although these methods significantly improved the accuracy of DTI predictions, not using explicit protein−ligand binding poses, can hamper the generalization ability of the models. For instance, the performance of DeepAffinity significantly dropped for protein−ligand complexes whose ligands and proteins are not included in the training set.[13]

Explicit binding poses can be considered in deep learning by using the atomic coordinates of protein−ligand complexes. Wallach et al. represented the complex structures around the binding site on a 3D rectangular grid, and then a 3D

Received: May 9, 2019
Published: August 23, 2019

ACS Publications © 2019 American Chemical Society

DOI: 10.1021/acs.jcim.9b00387
J. Chem. Inf. Model. 2019, 59, 3981−3988

convolutional neural network (CNN) was applied to the classification task of activity.[19] The result showed that the 3D CNN model outperformed docking for the DUD-E data set[20] in terms of the area under the receiver operating characteristic (AUROC) and the adjusted LogAUC.[21] Ragoza et al. modified the above approach to classify both activity and binding pose simultaneously.[22] For both classification tasks, the 3D CNN outperformed docking and other scoring functions such as RF-score[23] and NNScore.[24] Similar 3D CNN models were also applied to the regression task of binding affinities[25,26] and showed better performance compared to classical scoring functions such as RF-Score,[23] X-Score,[27] and Cyscore.[28]

The 3D atomic coordinates of molecules apparently contain full structural information. However, the 3D rectangular grid representation entails a lot of redundant grid points corresponding to a void space where no atoms reside, leading to inefficient computations. For more concise and yet effective representation, molecular graphs in which atoms and chemical bonds correspond to nodes and edges, respectively, can be utilized for both proteins and ligands. Then, graph neural networks (GNNs) can be used to make deep learning models for the graph representation. Previous works using sophisticated GNNs reported remarkable performances for predicting various molecular properties.[29−36] Such successful results suggest the GNNs as a promising architecture for improving DTI models. Indeed, Gomes et al. developed an atomic convolutional neural network by defining the atom type convolution and the radial pooling layer.[37] Feinberg et al. proposed a spatial graph convolution, applying different graph convolution filters based on the Euclidean distances between atoms.[38] The two proposed models outperformed conventional machine learning methods for the PDBbind[39] data set. Torng et al. developed a two-step graph convolutional framework using 2D structures of complexes for predicting DTI.[40] In their model, embedding vectors of proteins and ligands are obtained by graph autoencoders, respectively, and then a fully connected layer takes the two embedding vectors as input to predict DTI. As a result, their model considerably improved the accuracy of DTI predictions for the DUD-E data set.

To further improve DTI predictions with a graph representation, it is important to accurately take into account various types of intermolecular interactions because they are the key factors for the binding affinity of a given complex. One way is applying different neural networks to each edge type corresponding to a specific intermolecular interaction, where the edge type can be determined by predefined rules on Euclidean distance between atoms in different molecules.[38] However, predefined rules can cause undesirable biases, for instance, due to sensitive change of interaction types to the rules. A core benefit of deep learning is the ability to extract relevant features directly from raw data. Therefore, the best approach is to enable GNNs to extract DTI-relevant features directly from the 3D structural information embedded in a graph.

In this regard, we propose a novel approach for predicting DTI using a GNN that directly incorporates the 3D structural information on a protein−ligand binding pose. No heuristic chemical rules are used to deal with noncovalent interactions. In particular, we devise distance-aware graph attention mechanism[41] to make the model differentiate the contribution of each interaction to binding affinity. Furthermore, we utilize the graph feature obtained by subtracting each feature of a target protein and a given ligand from the graph feature of their complex for DTI predictions. These strategies allow the model to learn the key factors for accurate DTI predictions by making the model focusing on intermolecular interactions rather than just memorize certain patterns of ligand molecules. Additionally, we improve the performance of our model by adopting a gated skip-connection mechanism.[42] As a result, our method outperformed previous deep learning models as well as docking in terms of both virtual screening and pose prediction. Moreover, our model can reproduce the natural population distribution of active and inactive molecules. Finally, we analyzed our model's generalization ability by applying it to an external molecular library.

METHOD

In terms of methodology, our contribution can be summarized in the following three parts: embedding the 3D structural information of a protein−ligand complex in an adjacency matrix, devising a distance-aware attention algorithm to differentiate various types of intermolecular interactions, and introducing a variant of graph neural networks suitable for learning protein−ligand interaction. Each part is described in the following subsections. Before explaining our contributions, we briefly introduce graph neural networks. Then, we explain which data sets are used and how to preprocess them.

Graph Neural Network. Graphs can be defined by (V, E, $\mathbf{A}$), where V is a set of nodes, E is a set of edges, and $\mathbf{A}$ is an adjacency matrix. In an attributed graph, the attribute of each node is usually represented by a vector. The adjacency matrix, $\mathbf{A}$, is an $N \times N$ matrix, where $\mathbf{A}_{ij} > 0$ if ith and jth nodes are connected and $\mathbf{A}_{ij} = 0$ otherwise. N denotes the number of nodes in the graph. GNNs which operate on graphs have been explored in a diverse range of domains and have shown remarkable performance in various applications.[43] Also, various architectures of GNNs have been developed.[29−33,35]

In general, GNNs are composed of three stages: (i) updating node features, (ii) aggregating the node features and processing graph features, and (iii) predicting a label of the graph.[44] In the first stage, the node feature x_i, representing the attribute of the ith node, is updated over several times of message passing between neighboring nodes. This stage aims to obtain high level representations of node features. Then, the updated node features are aggregated to produce graph features. Here, the result of the aggregation must be invariant over permutations of node ordering. Finally, the graph features are used to predict a label of the entire graph, for instance, molecular properties.

Embedding the Structural Information on a Protein−Ligand Complex. Figure 1 shows a schematic representation of the GNN-based DTI prediction method proposed in this work. We embed the structural information between protein and ligand atoms in two adjacency matrices, $\mathbf{A}^1$ and $\mathbf{A}^2$. $\mathbf{A}^1$ represents purely covalent interactions, and $\mathbf{A}^2$ represents both covalent interactions and noncovalent intermolecular interactions. By constructing the two adjacency matrices, we let our model learn how protein−ligand interactions affect the node feature of each atom. $\mathbf{A}^1$ and $\mathbf{A}^2$ are constructed as follows:

$$\mathbf{A}^1_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected by covalent bond or } i = j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

DOI: 10.1021/acs.jcim.9b00387
J. Chem. Inf. Model. 2019, 59, 3981−3988

$$\mathbf{A}^2_{ij} = \begin{cases} \mathbf{A}^1_{ij} & \text{if } i, j \in \text{protein atoms or } i, j \in \text{ligand atoms} \\ e^{-(d_{ij}-\mu)^2/\sigma} & \text{if } d_{ij} < 5 \text{ Å and } i \in \text{ligand atoms and } j \in \text{protein atoms, or if } d_{ij} < 5\text{Å and } i \in \text{protein atoms and } j \in \text{ligand atoms} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where d_{ij} is the distance between the ith and jth atoms, and μ and σ are learnable parameters. The formula, $e^{-(d_{ij}-\mu)^2/\sigma}$, in eq 2 reflects that intermolecular bonds are weaker than covalent bonds and their strengths are getting weaker as the bond distance increases.

Representing 3D structures in an adjacency matrix has advantages over grid representation. In the grid representation, a large amount of empty grid points can cause unnecessary computation and memory usages. Also, the grid representation can lose distance information between atoms depending on the grid spacing. Furthermore, because it is not rotationally invariant, rotating atomic coordinates changes the prediction value of binding affinity. In contrast, our graph representation is compact and rotationally invariant. In addition, it enables efficient expression of the exact distance between atoms.

Distance-Aware Graph Attention Mechanism and Gate Augmentation Algorithm. The inputs of our graph attention layer are adjacency matrix, $\mathbf{A}$, and the set of node features, $\mathbf{x}^{\text{in}} = \{x_1^{\text{in}}, x_2^{\text{in}}, \cdots, x_N^{\text{in}}\}$ with $x \in \mathbb{R}^F$, where N is the number of nodes (i.e., the number of atoms), and F is the dimension of the node feature. The graph attention layer produces a new set of node features $\mathbf{x}^{\text{out}} = \{x_1^{\text{out}}, x_2^{\text{out}}, \cdots, x_N^{\text{out}}\}$ with $x \in \mathbb{R}^F$. To achieve sufficient expressive power, each node feature is transformed by a learnable weight matrix $\mathbf{W} \in \mathbb{R}^{F\times F}$, such that $x'_i = \mathbf{W}x_i^{\text{in}}$. Then, the attention coefficient, e_{ij}, is obtained as follows:

$$e_{ij} = x'^T_i \mathbf{E} x'_j + x'^T_j \mathbf{E} x'_i \quad (3)$$

where $\mathbf{E} \in \mathbb{R}^{F\times F}$ is also a learnable matrix. The attention coefficient, e_{ij}, represents the importance of the jth node feature to the ith node feature. We forced $e_{ij} = e_{ji}$ by summing $x'^T_i \mathbf{E} x'_j$ and $x'^T_j \mathbf{E} x'_i$. To reflect the graph structure, the attention coefficient, e_{ij}, is computed only for $j \in N_i$, where N_i is the neighborhood of the ith node. We define N_i as the set of nodes, of which $\mathbf{A}_{ij} > 0$ because our adjacency matrix reflects both connectivity and the normalized distance. To manipulate the scale of the attention coefficients across nodes, the attention coefficient is normalized across neighbors. Additionally, we multiply $\mathbf{A}_{ij}$ to the normalized attention coefficients to reflect that a node with a shorter Euclidean distance is more likely to be important than the others. It can be considered as an inductive bias. Consequently, the normalized attention coefficient, a_{ij}, is given by

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{j\in N_i} \exp(e_{ij})} \mathbf{A}_{ij} \quad (4)$$

After the normalized attention coefficient, a_{ij}, is obtained, each node feature is updated as a linear combination of the node features of the neighboring nodes with the normalized attention coefficient:

$$x''_i = \sum_{j\in N_i} a_{ij} x'_j \quad (5)$$

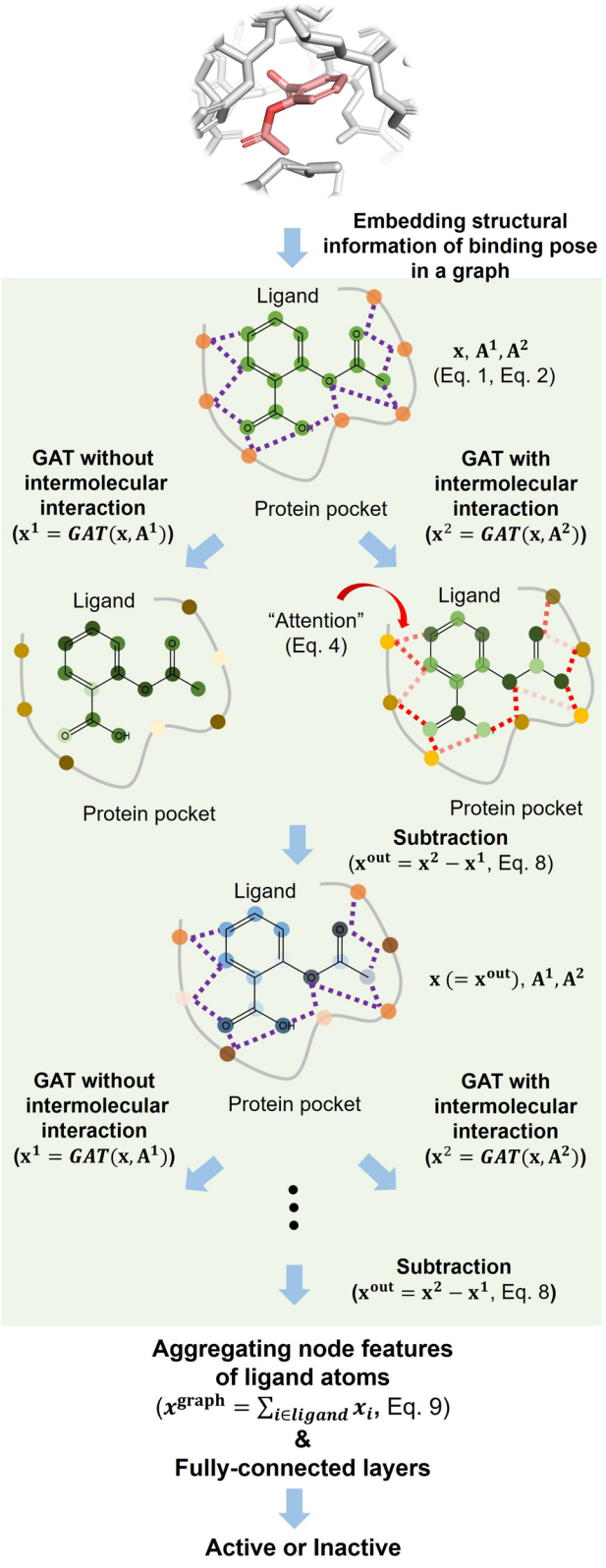


Figure 1. Schematic representation of the proposed DTI prediction method.

We also introduce a gate mechanism to directly deliver information on the previous node features to the next layer. It is found that a gate augmentation algorithm significantly improves the performance of a model.[42] We implement the output of our

DOI: 10.1021/acs.jcim.9b00387
J. Chem. Inf. Model. 2019, 59, 3981−3988

gated graph attention layer as a linear combination of x' and x'' as follows:

$$x_i^{\text{out}} = z_i x_i + (1 - z_i) x''_i \tag{6}$$

with

$$z_i = \sigma(\mathbf{U}(x_i \| x''_i) + b) \tag{7}$$

where $\mathbf{U} \in \mathbb{R}^{2F\times 1}$ is a learnable vector, b is a learnable scalar value, σ denotes a sigmoid activation function, and $(\cdot\|\cdot)$ is a concatenation of two vectors. z_i can be interpreted as how much information on input node features will be directly delivered to the next layer.

Neural Network Architecture. The inputs of our neural network are $\mathbf{x}$, $\mathbf{A}^1$, and $\mathbf{A}^2$. Two new node features, $\mathbf{x}^1$ and $\mathbf{x}^2$, are produced by the gate augmented graph attention layer, respectively, with $\mathbf{A}^1$ and $\mathbf{A}^2$, i.e., $\mathbf{x}^1 = \text{GAT}(\mathbf{x},\mathbf{A}^1)$ and $\mathbf{x}^2 = \text{GAT}(\mathbf{x},\mathbf{A}^2)$, where GAT stands for the gate augmented graph attention layer. It should be noted that one gate augmented graph attention layer is shared when computing both $\mathbf{x}^1$ and $\mathbf{x}^2$. The output node feature, $\mathbf{x}^{\text{out}}$, is obtained by subtracting $\mathbf{x}^1$ from $\mathbf{x}^2$:

$$\mathbf{x}^{\text{out}} = \mathbf{x}^2 - \mathbf{x}^1 \tag{8}$$

By subtracting the two node features, $\mathbf{x}^2$ and $\mathbf{x}^1$, we let our model learn the difference between the structure in a binding pose and the structure as separated. After the feature vectors are updated by several gate augmented graph attention layers, the feature vectors of ligand atoms are summed into one vector representing the graph of the protein−ligand complex:

$$x^{\text{graph}} = \sum_{i\in\text{ligand}} x_i \tag{9}$$

Finally, multilayer perceptrons are applied to x^{graph} to classify whether the protein−ligand complex or the binding pose is active or not. ReLU activation functions are used between the layers, and a sigmoid function is used after the last layer.

Data Set Preparation. We used the DUD-E[20] and PDBbind[39] v.2018 data sets to train and test our model. The 72 proteins and 25 proteins in the DUD-E set were used to train and test the model, respectively. To remove undesirable redundancy, we divided the data set in a way that no protein is present both in the training set and in the test set simultaneously. The 3D binding structures of protein−ligand complexes were obtained from docking calculations.

The PDBbind data set, which provides the experimentally verified binding structures of protein−ligand complexes, was used to train our model to distinguish the most favored binding pose of a given set of protein and ligand. For each sample in the PDBbind data set, we performed docking calculations to generate possible binding poses of the protein−ligand complex. A generated pose was labeled as a positive sample if the root-mean-square deviation (RMSD) from its experimentally verified binding structure is less than 2 Å or labeled as a negative sample if the RMSD is larger than 4 Å. The samples whose RMSD is between 2 and 4 Å were omitted. Then, the PDBbind data set was split into a training set and a test set depending on proteins so that the training and the test sets do not share same proteins. Additionally, the PDBbind samples with proteins included in the DUD-E data set were removed from both the training and the test sets.

The statistics of our training and test sets is summarized in Table 1. It shows that the inactive samples and the DUD-E

Table 1. Numbers of the Training Samples and the Test Samples for DUD-E Active, DUD-E Inactive, PDBbind Positive, and PDBbind Negative

	DUD-E active	DUD-E inactive	PDBbind positive	PDBbind negative
training	15864	973260	1598	9511
test	5841	364149	496	2735

samples are much more abundant in the training set and the test set compared to the active samples and the PDBbind samples. To deal with such imbalance, we sampled DUD-E active, DUD-E inactive, PDBbind positive, and PDBbind negative samples with the fixed ratio of 1:1:1:1 in preparing each training batch.

All the 3D binding structures of protein−ligand complexes were obtained using Smina,[45] the fork of AutoDock Vina,[10] even when experimental 3D structures are available, to maintain consistency. For the DUD-E data set, the default setting of Smina was used. Exhaustiveness = 50 and num modes = 20 were used for the docking calculations of the PDBbind data set. After the docking calculations were completed, the protein atoms whose minimum distance to the ligand atoms is larger than 8 Å were excluded to remove unnecessary atoms in the graph representation of protein−ligand complexes. We capped unsaturated valences using hydrogen atoms. It should be noted that we only consider heavy atoms, so the hydrogen atoms used for capping are implicitly included in the graph of protein−ligand complexes.

We represent initial atom features as a vector of size 56. The 1st to 28th entities represent ligand atoms and the 29th to 56th entities represent protein atoms. The list of the initial atom features is summarized in Table 2. Our model consists of four

Table 2. List of Atom Features

atom type	C, N, O, S, F, P, Cl, Br, B, H (onehot)
degree of atom	0, 1, 2, 3, 4, 5 (onehot)
number of hydrogen atoms attached	0, 1, 2, 3, 4 (onehot)
implicit valence electrons	0, 1, 2, 3, 4, 5 (onehot)
aromatic	0 or 1

gate augmented graph attention layers and three fully connected layers. The dimension of the graph attention layers was 140, and that of the fully connected layers was 128. We trained our model for 150000 iterations with the batch size of 32. To avoid overfitting, we applied dropout with the rate of 0.3 to every layer except the last of the fully connected layers. The data set and source code are available at https://github.com/jaechanglim/GNN_DTI

RESULTS AND DISCUSSION

Performance on the DUD-E and PDBbind Test Set. The performance of structure-based virtual screening can be assessed by measuring the ability to classify active and inactive compounds. We compared the performance of our model with those of docking and other deep learning models in terms of AUROC, adjusted LogAUC, PRAUC, sensitivity, specificity, and balanced accuracy. For calculations of sensitivity, specificity, and balanced accuracy, we considered samples with DTI prediction values above 0.5 as positive and the others as negative. We calculated those metric values of our model by averaging their values of each protein to balance data imbalance between the proteins.

DOI: 10.1021/acs.jcim.9b00387
J. Chem. Inf. Model. 2019, 59, 3981−3988

Table 3. AUROC, Adjusted LogAUC, PRAUC, Sensitivity, Specificity, and Balanced Accuracy of Our Model, Docking, and Other Deep Learning Models[a]

	AUROC	adjusted LogAUC	PRAUC	sensitivity	specificity	balanced accuracy
ours	**0.968**	**0.633**	**0.697**	**0.826**	0.967	**0.909**
ours w/o attention	0.936	0.577	0.623	0.758	**0.970**	0.888
docking	0.689	0.153	0.016			
Atomnet[19]	0.855	0.321				
Ragoza et al.[22]	0.868					
Torng et al.[40]	0.886					
Gonczarek et al.[17]	0.904					

[a]We note that the division of the training and test sets may be different for each model.

Table 3 summarizes the AUROC, adjusted LogAUC, PRAUC, sensitivity, specificity, and balanced accuracy values of our model, docking, and other deep learning methods for the DUD-E test set. Among various deep learning models, we chose the deep learning models trained using the DUD-E data set for fair comparison. Although all the models were trained and tested with the DUD-E data set, it should be noted that division of the training and test sets may be different for each model. Table 3 clearly shows that our GNN-based method outperformed the other deep learning models as well as the docking. Our model achieved AUROC of 0.968 compared to 0.689 of docking and 0.85−0.9 of other deep learning models. In addition, our model achieved high PRAUC value of 0.697, sensitivity of 0.826, specificity of 0.967, and balanced accuracy of 0.909 for the DUD-E test set, where the decoy molecules are much dominant than the active molecules.

We analyzed ROC enrichment (RE)[46,47] score and summarized the result in Table 4. The RE score indicates the ratio of the true positive rate (TPR) to the false positive rate (FPR) at a certain FPR value. In terms of the RE score, our method shows 9−10 times better performance than that of the docking and 2−3 times better performance than those of the other deep learning models. Also, the distance-aware attention algorithm clearly improved the performance in the virtual screening for all metrics. Generally, in the process of hit discovery, only the top hundreds of molecules are subjected to experimental verification. Therefore, such high performance on the LogAUC and the RE score indicates a practical advantage of our model in the hit discovery.

Table 4. ROC Enrichment (RE) Score of Our Model, Docking, and Other Deep Learning Methods[a]

	0.5%	1.0%	2.0%	5.0%
ours	**124.031**	**69.037**	**38.027**	**16.910**
ours w/o attention	107.734	61.346	34.326	16.029
docking	11.538	9.749	6.153	3.789
Ragoza et al.[22]	42.559	29.654	19.363	10.710
Torng et al.[40]	44.406	29.748	19.408	10.735

[a]The RE score indicates the ratio of the true positive rate (TPR) to the false positive rate (FPR) at a certain FPR value.

Selection of the most favored binding pose of a given complex is important to understand binding affinity in terms of intermolecular interactions. Such understanding helps human experts rationally modify the ligand to further improve its efficacy. Our model can be used for pose prediction as well because the 3D conformational information is directly included in the graph representation of a protein−ligand complex. Table 5 summarizes the performance of our model and docking in terms of AUROC and PRAUC for the PDBbind test set. Our model improved AUROC and PRAUC by about 0.11 and 0.26 from the results of docking, respectively. As in the virtual screening results, the attention algorithm clearly improved the performance in the pose prediction.

Table 5. AUROC and PRAUC of Our Model and Docking for Classification between Favored and Unfavored Binding Poses

	AUROC	PRAUC
ours	**0.935**	**0.772**
ours w/o attention	0.927	0.698
docking	0.825	0.509

Figure 2 shows the percentage of the complexes with RMSD smaller than 2 Å with respect to experiments in top-N poses identified by the docking and our model. Our model performed 5−7% better than the docking. However, the performance gap between our model and the docking is relatively small compared to the performance gaps in the virtual screening. This means that the docking is relatively more accurate for ranking binding poses than for predicting binding affinities.

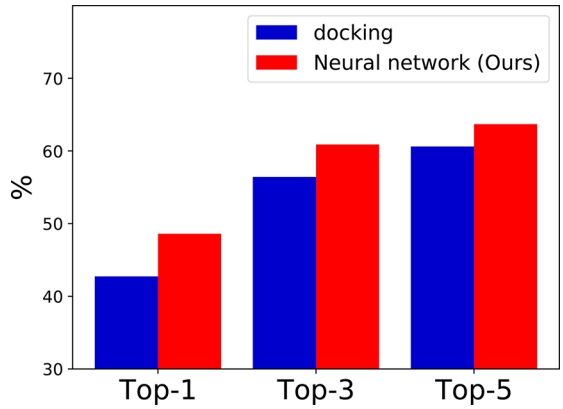


Figure 2. Percentage of pretein−ligand complexes whose RMSD with respect to experimental structures is smaller than 2 Å in top-N poses identified by docking and our model.

Distribution of Predicted Activity for a Molecular Library. The number of potential drug candidates is estimated to be about 10^{23}−10^{60}.[48] Among such large number of molecules, it is expected that most of them are inactive to a given protein. We tested whether our model can reproduce such a naturally expected distribution. To do so, 470094 synthetic compounds were obtained from https://www.ibscreen.com. We preprocessed those synthetic compounds (IBS molecules) as done for the DUD-E data set. The distribution of predicted

DOI: 10.1021/acs.jcim.9b00387
J. Chem. Inf. Model. 2019, 59, 3981−3988

activities for the IBS molecules to epidermal growth factor receptor (EGFR) is plotted in Figure 3. It should be noted that EGFR was excluded in the DUD-E training set.

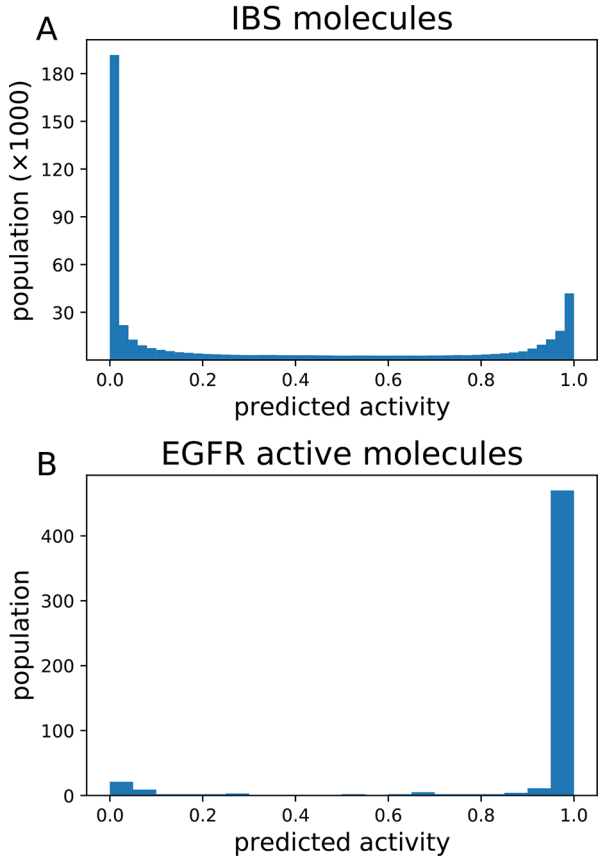


Figure 3. Activity distributions predicted by our model for the IBS molecules and EGFR active molecules in the DUD-E data set.

Figure 3A shows that inactive probabilities are dominant, which seems close to a natural population. For comparison, we also tested our model for the known EGFR active molecules in the DUD-E data set. Figure 3B shows that our model predicted activities close to 1.0 for most active molecules. In the predicted activity distributions, a small peak is observed around 1.0 in Figure 3A and 0.0 in Figure 3B. These unnatural peaks may come from the overconfidence of our model, indicating a possibility of slight overfitting.

Performance on External Libraries: ChEMBL and MUV. In the DUD-E data set, the decoy molecules have considerable structural differences with the active molecules, so that classifying the DUD-E active and decoy molecules might be relatively easy. On the other hand, experimentally verified inactive compounds share more common structures with active compounds than the decoys do. Therefore, we further validated whether our model, trained on the DUD-E data set, can classify active and inactive compounds which were experimentally verified. The active and inactive molecules with respect to the DUD-E test proteins were collected from the ChEMBL[49] database and preprocessed in the same way as done for the DUD-E data set. We labeled the ChEMBL molecules whose IC_{50} value is smaller than 1.0 μM as active or inactive otherwise. As a result, 27389 active and 26939 inactive molecules were obtained. Table 6 shows the AUROC, sensitivity, specificity, and balanced accuracy values of our model and the docking for the

Table 6. AUROC, Sensitivity, Specificity, and Balanced Accuracy of Our Model, Docking, and Other Deep Learning Models on Experimentally Verified Active and Inactive Molecules, and MUV Set[a]

	ChEMBL			
	AUROC	sensitivity	specificity	balanced accuracy
ours	**0.633**	0.813	0.325	0.569
docking	0.572			
	MUV			
	AUROC	sensitivity	specificity	balanced accuracy
ours	0.536	0.286	0.752	0.519
docking	0.533			
Ragoza et al.[22]	0.518			
Torng et al.[40]	**0.563**			

[a]The experimentally verified molecules were obtained from the ChEMBL database.

ChEMBL molecules. Although our model is still better than the docking, the AUROC was dropped about 0.3 from that of the DUD-E test set. The ChEMBL molecules and the DUD-E test set share common proteins, so the difference in their AUROC values only comes from the difference between the ligand sets.

Additionally, we validated our model on the MUV[50] data set. The MUV data set is designed to remove undesirable bias between active molecules and decoys by optimally spreading active molecules in the chemical space of decoys while maintaining the molecular similarities between active-active molecules and active-decoy molecules. In Table 6, our model, as with the other deep learning models and the docking, shows AUROC close to 0.5 for the MUV data set. That a model gives AUROC close to 0.5 means that it randomly classifies active and inactive molecules. The notable performance drop for the ChEMBL and MUV data sets indicates that the deep learning models including ours have common problems in generalization. We hypothesize that the reason for such common performance drop is because molecules available in the DUD-E data set are simply not enough to cover the vast chemical space of natural molecules, so deep neural networks find a hidden pattern within the DUD-E data set.

CONCLUSION

In this work, we proposed a novel approach based on a graph neural network specialized for predicting drug−target interaction. We directly incorporated the 3D structural information of protein−ligand binding poses into an adjacency matrix. We also applied a distance-aware graph attention algorithm with gate mechanism to increase the performance of the model. Our model was trained using the DUD-E data set for virtual screening and the PDBbind data set for binding pose prediction. As a result, our model outperformed docking and other deep learning models in terms of both virtual screening and pose prediction. Our model showed AUROC of 0.968 for virtual screening and 0.935 for pose prediction. It could also reproduce a natural population distribution of active and inactive molecules.

Apart from such high performance, our model has a similar limitation in generalization with other deep learning models. That is, its performance in classification of experimentally verified active and inactive compounds significantly dropped from that of the DUD-E test set. Also, our model, like the docking and the other deep learning models, failed to correctly

DOI: 10.1021/acs.jcim.9b00387
J. Chem. Inf. Model. 2019, 59, 3981−3988

classify active and decoy molecules in the MUV data set. This may be because the DUD-E training set cannot effectively span a huge chemical space of natural molecules. Therefore, it is possible that there is a biased structural pattern to classify active and inactive molecules which can be readily captured by deep neural networks.

Along with the improvement of generalization ability, deep learning techniques such as Bayesian neural networks[51−54] would be useful to quantify the uncertainty of DTI predictions. The uncertainty quantification is particularly advantageous in developing DTI models in that acquiring high quality and sufficient quantity of relevant data is expensive. For instance, quantifying such uncertainty arisen from insufficient data quality and quantity enables to estimate the scope of protein targets where the model can provide reliable predictions. In conclusion, the generalization ability issue with quantification of uncertainty should be addressed to develop a versatile deep learning model applicable to various data sets for virtual screening of drug candidates.

AUTHOR INFORMATION

Corresponding Author

*E-mail: wooyoun@kaist.ac.kr.

ORCID

Woo Youn Kim: 0000-0001-7152-2111

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2017R1E1A1A01078109).

REFERENCES

(1) Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. *J. Chem. Theory Comput.* **2017**, *13*, 42−54.

(2) Beierlein, F. R.; Michel, J.; Essex, J. W. A Simple QM/MM Approach for Capturing Polarization Effects in ProteinLigand Binding Free Energy Calculations. *J. Phys. Chem. B* **2011**, *115*, 4911−4926.

(3) Venkatachalam, C.; Jiang, X.; Oldfield, T.; Waldman, M. Ligandfit: A Novel Method for the Shape-directed Rapid Docking of Ligands to Protein Active Sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289−307.

(4) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36*, 1132−1156.

(5) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **2014**, *10*, No. e1003571.

(6) Zhao, H.; Caflisch, A. Discovery of Zap70 Inhibitors by High-throughput Docking into a Conformation of Its Kinase Domain Generated by Molecular Dynamics. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5721−5726.

(7) Jain, A. N. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.* **2003**, *46*, 499−511.

(8) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(9) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(10) Trott, O.; Olson, A. J. Autodock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, *8*, 455−461.

(11) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436−444.

(12) Park, S.; Kwon, Y.; Jung, H.; Jang, S.; Lee, H.; Kim, W. CSgator: An Integrated Web Platform for Compound Set Analysis. *J. Cheminf.* **2019**, *11*, 17.

(13) Karimi, M.; Wu, D.; Wang, Z.; Shen, Y. DeepAffinity: Interpretable Deep Learning of Compound-Protein Affinity through Unified Recurrent and Convolutional Neural Networks. *Bioinformatics* **2019**, btz111.

(14) Gao, K. Y.; Fokoue, A.; Luo, H.; Iyengar, A.; Dey, S.; Zhang, P. Interpretable Drug Target Prediction Using Deep Neural Representation. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), Stockholm, 13−19 July 2018*; International Joint Conferences on Artificial Intelligence, 2018.

(15) Lee, I.; Keum, J.; Nam, H. DeepConv-DTI: Prediction of Drug-target Interactions via Deep Learning with Convolution on Protein Sequences. *PLoS Comput. Biol.* **2019**, *15*, e1007129.

(16) Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: Deep Drug-Target Binding Affinity Prediction. *Bioinformatics* **2018**, *34*, i821−i829.

(17) Gonczarek, A.; Tomczak, J. M.; Zaręba, S.; Kaczmar, J.; Dąbrowski, P.; Walczak, M. J. Learning Deep Architectures for Interaction Prediction in Structure-based Virtual Screening. *arXiv* **2016**, arXiv:1610.07187v3.

(18) Öztürk, H.; Özgür, A.; Ozkirimli, E. A Chemical Language Based Approach for Protein−Ligand Interaction Prediction. *arXiv* **2018**, arXiv:1811.00761.

(19) 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.

(20) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.

(21) Mysinger, M. M.; Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model.* **2010**, *50*, 1561−1573.

(22) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein−Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model.* **2017**, *57*, 942−957.

(23) Ballester, P. J.; Mitchell, J. B. O. A Machine Learning Approach to Predicting Protein−ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* **2010**, *26*, 1169−1175.

(24) Durrant, J. D.; McCammon, J. A. NNScore 2.0: A Neural-Network Receptor−Ligand Scoring Function. *J. Chem. Inf. Model.* **2011**, *51*, 2897−2903.

(25) Jiménez, J.; Škalič, M.; Martínez-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.

(26) Stepniewska-Dziubinska, M. M.; Zielenkiewicz, P.; Siedlecki, P. Development and Evaluation of a Deep Learning Model for Protein−ligand Binding Affinity Prediction. *Bioinformatics* **2018**, *34*, 3666−3674.

(27) Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-based Binding Affinity Prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11−26.

(28) Cao, Y.; Li, L. Improved Protein−ligand Binding Affinity Prediction by Using a Curvature-dependent Surface-area Model. *Bioinformatics* **2014**, *30*, 1674−1680.

(29) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *arXiv* **2017**, arXiv:1704.01212.

(30) Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R. Gated Graph Sequence Neural Networks. *arXiv* **2015**, arXiv:1511.05493.

DOI: 10.1021/acs.jcim.9b00387
J. Chem. Inf. Model. 2019, 59, 3981−3988

(31) Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *arXiv* **2015**, arXiv:1509.09292v2.
(32) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular Graph Convolutions: Moving Beyond Fingerprints. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 595−608.
(33) Battaglia, P. W.; Pascanu, R.; Lai, M.; Rezende, D.; Kavukcuoglu, K. Interaction Networks for Learning about Objects, Relations and Physics. *arXiv* **2016**, arXiv:1612.00222v1.
(34) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with Dft Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192−3203.
(35) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890.
(36) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecule Neural Network. *ChemRxiv* **2018**, 7151435.
(37) Gomes, J.; Ramsundar, B.; Feinberg, E. N.; Pande, V. S. Atomic Convolutional Networks for Predicting Protein−Ligand Binding Affinity. *arXiv* **2017**, arXiv:1703.10603v1.
(38) Feinberg, E. N.; Sur, D.; Wu, Z.; Husic, B. E.; Mai, H.; Li, Y.; Sun, S.; Yang, J.; Ramsundar, B.; Pande, V. S. PotentialNet for Molecular Property Prediction. *ACS Cent. Sci.* **2018**, *4*, 1520.
(39) Liu, Z.; Su, M.; Han, L.; Liu, J.; Yang, Q.; Li, Y.; Wang, R. Forging the Basis for Developing Protein-Ligand Interaction Scoring Functions. *Acc. Chem. Res.* **2017**, *50*, 302−309.
(40) Torng, W.; Altman, R. B. Graph Convolutional Neural Networks for Predicting Drug−Target Interactions. *bioRxiv* **2018**, 473074.
(41) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, arXiv:1710.10903v3.
(42) Ryu, S.; Lim, J.; Hong, S. H.; Kim, W. Y. Deeply Learning Molecular Structure−Property Relationships Using Attention- and Gate-Augmented Graph Convolutional Network. *arXiv* **2018**, arXiv:1805.10988v3.
(43) Zhou, J.; Cui, G.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph Neural Networks: A Review of Methods and Applications. *arXiv* **2018**, arXiv:1812.08434v4.
(44) Battaglia, P. W.; Hamrick, J. B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R.; Gulcehre, C.; Song, F.; Ballard, A.; Gilmer, J.; Dahl, G.; Vaswani, A.; Allen, K.; Nash, C.; Langston, V.; Dyer, C.; Heess, N.; Wierstra, D.; Kohli, P.; Botvinick, M.; Vinyals, O.; Li, Y.; Pascanu, R. Relational Inductive Biases, Deep Learning, and Graph Networks. *arXiv* **2018**, arXiv:1806.01261v3.
(45) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893−1904.
(46) Jain, A. N.; Nicholls, A. Recommendations for Evaluation of Computational Methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133−139.
(47) Nicholls, A. What Do We Know and When Do We Know It? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239−255.
(48) Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the Size of Drug-like Chemical Space Based on GDB-17 Data. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 675−679.
(49) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magariños, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R. The ChEMBL Database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945−D954.
(50) Rohrer, S. G.; Baumann, K. Maximum Unbiased Validation (MUV) Data Sets for Virtual Screening Based on PubChem Bioactivity Data. *J. Chem. Inf. Model.* **2009**, *49*, 169−184.
(51) Gal, Y.; Ghahramani, Z. Dropout As a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *Proceedings of the 33rd International Conference on International Conference on Machine Learning, 19−24 June, 2016*; ICML, 2016; Vol. *48*, pp 1050−1059.
(52) Kendall, A.; Gal, Y. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc., 2017; pp 5574−5584.
(53) Ryu, S.; Kwon, Y.; Kim, W. Y. A Bayesian Graph Convolutional Network for Reliable Prediction of Molecular Properties with Uncertainty Quantification. *Chem. Sci.* **2019**, 3192−3203.
(54) Kwon, Y.; Won, J.; Kim, B.; Paik, M. Uncertainty Quantification Using Bayesian Neural Networks in Classification: Application to Ischemic Stroke Lesion Segmentation. *International Conference on Medical Imaging with Deep Learning, Amsterdam, 4−6 July, 2018*; MIDL, 2018.

DOI: 10.1021/acs.jcim.9b00387
J. Chem. Inf. Model. 2019, 59, 3981−3988