Alexandria Engineering Journal 109 (2024) 299–306

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Insight mixed deep neural network architectures for molecular representation

Tianze Zhao [a], Zhenyu Yin [b], Yong Lu [a,*], Shaocong Cheng [c], Chunyan Li [d]

[a] School of Information Engineering, Minzu University of China, Beijing 100081, China
[b] School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China
[c] College of Robotics, Beijing Union University, Beijing 100101, China
[d] School of Informatics, Yunnan Normal University, Kunming 650500, China

ARTICLE INFO

Keywords:
Deep neural network
Graph neural network
Molecular representation
Molecular property prediction

ABSTRACT

Learning molecular representation is a crucial task in the field of drug discovery, particularly for various specific applications such as predicting molecular properties. Current methods are mainly based on deep neural network models, such as convolutional neural networks (CNN), recurrent neural networks (RNN), graph neural networks (GNN) and their mixed models. However, these neural network models mentioned above do not provide detailed explanations for the ability to learn molecular representations and why such experimental results occurred. In this paper, we aim to compare the performance of these models in predicting molecular properties based on molecular representation, and give more insight into deep neural network architectures for specific molecular tasks. Our experimental results demonstrate that the graph neural network can obtain superior performance on the regression tasks, while the mixed deep neural network models show better performance on the classification tasks. Ablation study also gives more explanation and analysis to the experimental results.

1. Introduction

Molecular representation is a critical step in computational drug design, serving as the foundation for various predictive models and simulations [1–4]. It involves encoding the structural and chemical properties of molecules into a format that can be processed by machine learning algorithms and other computational tools. Accurate molecular representation is crucial for the effectiveness of computational models in drug discovery. In drug design [5], the goal is to identify and optimize molecules that can interact effectively with biological targets to produce a therapeutic effect. Accurate molecular representation allows for prediction of biological activity [2], property prediction [3] and optimization [6]. Traditionally, drug molecular representation can be obtained via the experience of a chemist or pharmacist. Meanwhile, many simulations and experiments are essential for understanding the properties of a compound, which are time-consuming and inefficient. The exponential expansion of experiment data provides support for deep learning and makes it possible.

The rapid development of deep learning has been applied in drug discovery and compound screening [7–11]. Existing research [9,12] shows that deep neural networks-based deep learning has outperformed conventional machine learning methods. Deep neural network models can learn the latent knowledge in a large amount of data to obtain an effective representation of the target object. Then the effective representation is always used to accurately predict various downstream tasks. Therefore, the data-driven deep learning method is expected to accelerate the drug discovery. Recently, many molecular representation methods for properties prediction tasks have emerged. The neural network architectures include convolutional neural network (CNN) [12], recurrent neural networks (RNN) [12], graph convolutional neural network (GNN) [13] and graph attention network (GAT) [7]. On the other hand, the molecular representation includes 1D SMILES sequence manner [12,14], 2D fingerprint and graph [12,15,16], 3D molecular structure [9,13] and mixed representation learning methods of different forms of molecules [12]. SMILES [14] is the abbreviation for *Simplified Molecular-Input Line-Entry System*, which is in form of one-dimensional (1D) sequence of chemical structure. Extended connectivity fingerprint (ECFP) [15] is a widely used molecular fingerprint to construct quantitative structure-activity relationship (QSAR) model in order to predict molecular activity for 2D representation of molecule.

* Corresponding author.
E-mail addresses: 2424037767@qq.com (T. Zhao), 1328006854@qq.com (Z. Yin), 2006153@muc.edu.cn (Y. Lu), 2756847004@qq.com (S. Cheng), lchy0316@gmail.com (C. Li).

https://doi.org/10.1016/j.aej.2024.08.113
Received 8 June 2024; Received in revised form 13 August 2024; Accepted 30 August 2024
Available online 9 September 2024

MoleculeNet [17] is a large-scale benchmark for molecular machine learning, which establishes metrics for evaluation based on multiple public datasets. Molecules can be represented as one-dimensional SMILES sequences, enabling the application of deep neural networks, such as CNN, RNN and MLP, to learn molecular representation by effectively extracting features from these sequence structures. Additionally, molecules can also be represented as two-dimensional graph structures, where nodes denote atoms and edges represent the bonds between these atoms. Consequently, GNN are employed to extract features from these molecular graph structures and facilitate molecular representation learning. In this paper, we aim to design a neural network model to evaluate the effectiveness of molecular representation. The evaluation is based on the model's accuracy in predicting various downstream molecular properties. we compare the performance of various deep neural networks and graph neural networks in predicting molecular properties based on molecular representation, and give more insight into deep neural network architectures for specific molecular tasks. We evaluate molecular properties in four public benchmark datasets (ESOL, Lipophilicity, HIV, BACE) based on conventional machine learning methods, graph-based methods and mixed deep neural network methods. In summary, the key contributions to this paper are as follows:

- We propose a novel neural network architecture aimed at providing insights and comparing the performance of mixed deep neural network models to the other models (conventional machine learning and graph convolutional models) for molecular property prediction tasks.
- Different neural network models learn distinct molecular representations, resulting in varying predictive performance across different tasks. We find that the graph neural network can obtain superior performance on the regression tasks, while the mixed deep neural network models show better performance on the classification tasks.
- We provide detailed explanations for the ability to learn molecular representations and why such experimental results occurred. We give more insight into deep neural network architectures for specific molecular tasks.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents our approaches. Section 4 analyzes and discusses the experimental results. Section 5 concludes this paper.

2. Related work

In this section, we will briefly describe some neural networks related to molecular representation, that can summarized as follows:

2.1. Graph-based molecular representation

Molecular representation is crucial in the field of drug discovery, which is the basis of downstream tasks. A spatial graph embedding method (C-SGEN) [18] is proposed based on GNN for learning molecular feature to predict molecular properties. Message passing neural network (MPNN) [19] consists of two phases, a message passing phase and a readout phase. It is a common framework for learning molecular representation to propagate the node embeddings and update features via the embeddings of their neighbors. A Bayesian graph convolutional neural network [20] is proposed to solve the problem of model training in the absence of data, which is important for implementing downstream tasks of the molecule in the case of data-deficient. Graph convolutional (GC) model [21] learns molecular graph fingerprints using neural network. Weave [22] is proposed to use undirected graph to learn molecular structure information.

2.2. Recurrent networks-based molecular representation

A mixed deep neural network architecture CheMixNet [12] is proposed for learning molecular representation via two parallel neural networks, that learn molecular SMILES sequences and molecular ACCess System (MACCS) fingerprint, respectively. MACCS fingerprint consists of 166-dimensional features, which is a 2D representation of molecular structure. In the characterization of SMILES sequences, CheMixNet uses RNN and CNN neural network architectures to extract molecular sequence features respectively. The RNN is based on long short-term memory (LSTM) [23] or gated recurrent unit (GRU) [23] unit. Finally, CheMixNet is used in the tasks of molecular attribute predictions to detect the ability of molecular representation and obtains superior predictive performance. Lin et al. proposed sequence-based manner to learn molecular representation by using bidirectional gated recurrent unit (BiGRU) neural network [24], that is designed for solving the single- and multi-task classification in the field of drug discovery.

2.3. Conventional machine learning-based molecular representation

Conventional machine learning-based molecular representation has been a significant area of research, focusing on various methods to encode molecular structures and properties into formats suitable for machine learning algorithms. Conventional machine learning methods, such as XGBoost [25], KernelSVM [26] and Random Forests [27], are frequently employed for molecular representation learning. All three algorithms are capable of handling non-linear relationships between features and labels, especially KernelSVM, which maps data to higher dimensions through kernel techniques to find non-linear decision boundaries. XGBoost and Random Forests are both ensemble methods that improve model stability and predictive performance by combining multiple base learners, such as decision trees. These conventional machine learning-based approaches to molecular representation have laid the foundation for more advanced methods, such as deep learning-based representations, and continue to play a critical role in cheminformatics and related fields.

3. Methods

In this section, we propose a novel neural network architecture aimed at providing insights and comparing the performance of mixed deep neural network models to the other models (conventional machine learning and graph convolutional models) for molecular property prediction tasks. CheMixNet points out the mixed deep neural network architecture models outperforms 1D CNN, RNN models and other graph convolutional models. However, we argue that this is not always the case. In this paper, we more comprehensively demonstrate the performance of mixed deep neural network models proposed in [12], conventional machine learning models and graph convolutional neural network models in predicting molecular attributes based on molecular representation. Meanwhile, we give more insight into deep neural network architectures for specific molecular tasks. Generally, for the mixed deep neural network architecture, use RNN or CNN model or a mixture of RNN and CNN to learn and characterize the molecular SMILES sequence. This kind of deep neural network can learn the potential knowledge of SMILES sequence and obtain accurate feature expression of SMILES sequence. In addition, use multilayer perceptron (MLP) extracts the feature of molecular fingerprint for representing the fingerprint information. CheMixNet [12] uses ACCess System (MACCS) fingerprint with 166-dimensional features, which is input to MLP neural network. It is different from ours in that we use ECFP molecular fingerprints [15] with 1024-dimensional features. Fig. 1 shows our proposed neural network architecture. Above is a mixed deep neural network architecture proposed in CheMixNet, and below is a graph convolutional neural network. As shown in Fig. 1, the two parallel neural networks RNN or CNN and MLP are finally concatenated

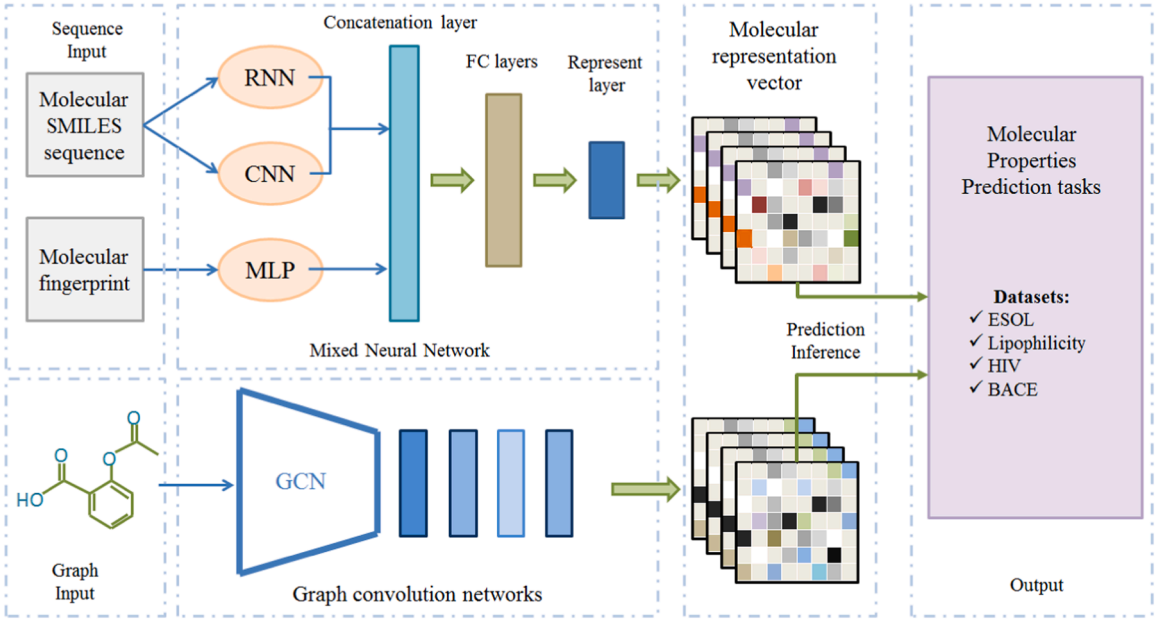

Fig. 1. Our proposed neural network architecture. Above is a mixed deep neural network architecture, which consists of two parallel neural networks, RNN or CNN and MLP. Below is a graph convolutional neural network with the aggregation process of graph convolutional layers shown in Fig. 2.

together, then are input to several fully connected (FC) layers. Use the molecular representation output from the fully connected layers to predict specific tasks, such as molecular properties predictions, including water solubility, lipophilicity, activity and inhibition. It is worth noting that in the research of this article, we use LSTM [23] recurrent neural network unit as the RNN network. For graph convolutional neural network, we use graph convolutional method proposed in [28] as our method of molecular graph feature extraction. Graph convolutional network (GCN) [28] obtains the neighbors information of a node as the feature of this node via a first-order approximation of spectral graph convolutions. As shown in Fig. 2, Node X3 aggregates the features of its 4 neighbor nodes to update itself. Other nodes also perform similar aggregation operations. After several graph convolutional layers, each node has a new aggregation feature, here are Y1, Y2, Y3, Y4, Y5, respectively. For conventional machine learning, we use XGBoost [25], KernelSVM [26] and Random forests (RF) [27] as our baselines for performance comparison.

In the following, we will describe the details of LSTM [23] and GCN [28]. We first present LSTM [23] (Section 3.1), then describe GCN [28] (Section 3.2).

3.1. Long short-term memory (LSTM)

Long short-term memory (LSTM) [23] is a special recurrent neural network, aimed to solve the problem of gradient disappearance and gradient explosion in the training process of long sequences. Compared to ordinary recurrent neural networks [29], LSTM can capture the relationships of long sequences. Therefore, LSTM perform better in longer sequences. LSTM uses memory cell as traditional neuron in neural network. The memory cell is controlled using input gates, output gates and forget gates. Fig. 3 shows how LSTM works. The sigmoid neural layer functions as a gate, which allows data to pass or prohibits data to pass. As shown in Fig. 3, we will give the working mechanism of each data stream in LSTM, such as f_t, i_t, G_t, j_t, C_t, which can be defined as follows:

$$f_t = \sigma(W_f \bullet [h_{t-1}, \quad X_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \bullet [h_{t-1}, \quad X_t] + b_i) \tag{2}$$

$$G_t = \tanh(W_G \bullet [h_{t-1}, \quad X_t] + b_G) \tag{3}$$

$$C_t = \quad f_t * C_{t-1} + \quad i_t * G_t \tag{4}$$

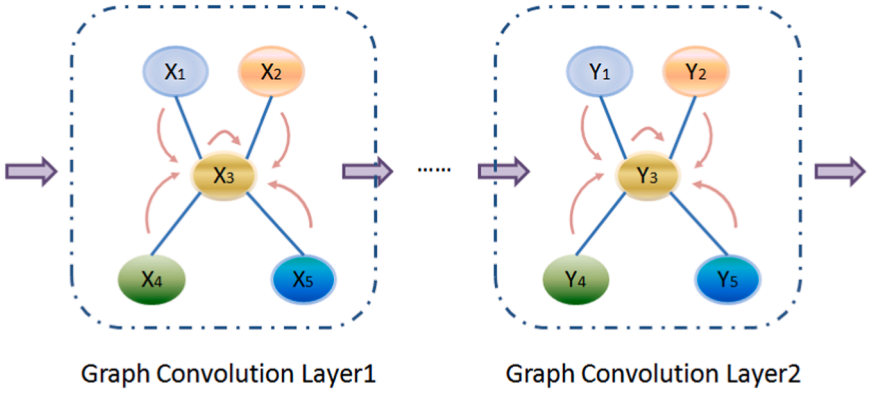

Fig. 2. The graph convolutional neural network, which aggregates neighbor nodes via a localized first-order approximation in the process of convolutional operation.

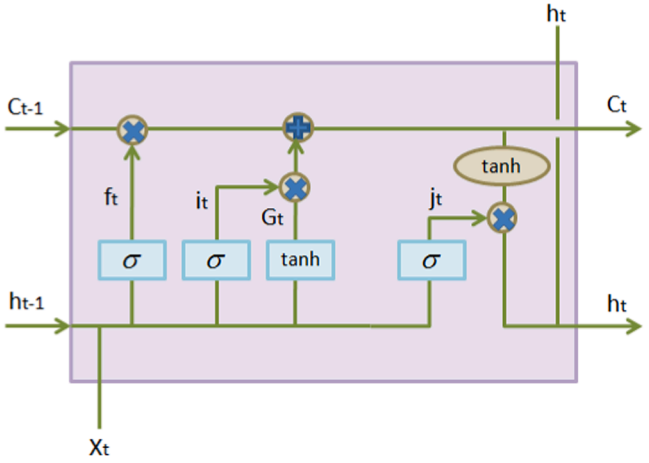

Fig. 3. LSTM work flow chart with gates, which can optionally let data through.

$$j_t = \sigma(W_j \bullet [h_{t-1}, \quad X_t] + b_j) \tag{5}$$

$$h_t = \quad j_t \quad * \tanh(C_t) \tag{6}$$

3.2. Graph convolutional network (GCN)

Graph convolutional network (GCN) [28] can aggregate neighbor nodes information by combining local graph structure and node features. Define a graph $\mathcal{G}=(\mathcal{V}, \mathcal{E})$ with N nodes $v_i \in \mathcal{V}$ and edges $(v_i \quad , v_j) \in \mathcal{E}$, $A \in \mathbb{R}^{N\times N}$ is an adjacency matrix. Graph convolution operation can be calculated as

$$F^{(t+1)} = \quad f(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}F^{(t)}W^{(t)}) \tag{7}$$

where $\widetilde{A} = A + I$, I is an identity matrix, $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$ is a degree matrix of $\widetilde{A}$, $W^{(t)}$ is a weight matrix. $f(\cdot)$ is an activation function. $F^{(t)}$ is a feature matrix in the t^{th} layer; $F^{(0)}$ is initial value. The normalized Laplacian $L = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. The spectral graph convolution is defined as the multiplication of a signal $x \in \mathbb{R}^N$ with a filter $g_\theta = \mathrm{diag}(\theta)$, that is parameterized by θ, which can be defined as:

$$g_\theta * x = U g_\theta U^T x \tag{8}$$

where U is the matrix of eigenvectors of L, Λ is its eigenvalues, and $U^T x$ is the Fourier transform of input features x to a set of basis. We can approximate the value using Chebyshev polynomials $\mathrm{T_k}(x)$ up to K-order [30,31]:

$$g_{\theta'}(\Lambda) \quad \approx \sum_{k=0}^{K} \theta'_k T_k(\widetilde{\Lambda}) \tag{9}$$

$$g_{\theta'} * x \approx \sum_{k=0}^{K} \theta'_k T_k(\widetilde{L}) x \tag{10}$$

where θ'_k is a Chebyshev coefficient,

$\widetilde{L} = \quad \frac{2}{\lambda_{\max}} L - I$, $\lambda_{\max}$ is the max eigenvalue of Laplacian matrix L. $\mathrm{T}_0(x) = 1$,

$\mathrm{T}_1(x) = x, \mathrm{T}_k(x) = \quad 2x\mathrm{T}_{k-1}(x) - \quad \mathrm{T}_{k-2}(x)$. So Eq.10 can be written as follows:

$$g_{\theta'} * x \approx \theta'_0 x + \theta'_1 (L - I)x = \theta'_0 x - \theta'_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \tag{11}$$

$$g_{\theta'} * x \approx \theta \left(I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) x \tag{12}$$

where $\theta = \theta'_0 = \ -\theta'_1$, and D is the degree matrix of A.

The graph is a kind of important structure in deep neural network. Graph convolutional network is a strong baseline for this article. Graph neural network can capture the characteristics of data like graph and learn the representation. It is a powerful model for processing non-Euclidean data with graph structure. The molecule can be regarded as a graph, the atoms in the molecule are regarded as the nodes of the graph, and the bonds in the molecule are regarded as the edges of the graph. Therefore, molecules can be learned and represented by graphs.

4. Experiments

Extensive experiments have been conducted to evaluate the performance of molecular representation using various mixed deep neural network architectures based on the tasks of molecular property predictions. In this section, we describe the datasets, baselines, metrics and the experimental results.

4.1. Datasets

We evaluate the models of mixed deep neural network architectures by using four public molecular datasets, including classification tasks and regression tasks, that is recommended by MoleculeNet [17]. The predictive properties include water solubility, lipophilicity, active and inhibitory. Table 1 shows the information of four datasets. The details of used datasets are shown as follows:

- ESOL. ESOL dataset [32] contains the water solubility data of 1128 molecules, which gives the SMILES structures of the compounds, and is often used to train models to estimate water solubility from chemical structures as regression task in deep neural networks.
- Lipophilicity. Lipophilicity [33] contains 4200 experimental results of octanol/water distribution coefficient. Lipophilicity is important in membrane permeability and solubility.
- HIV. HIV dataset [34] contains 41,127 compounds to show their capability of inhibiting HIV replication. The compounds are divided as active compound (label = 1) and inactive compound (label = 0). This is a typical dual-classification task in deep neural networks.
- BACE. BACE [35] contains 1513 compounds and always is built as a classification task, which provides quantitative and qualitative binding results for a set of inhibitors.

4.2. Baselines

In this paper, we focus on conventional machine learning-based models, graph-based models and the models proposed in CheMixNet [12]. We aimed to compare the performance of these models in predicting molecular properties based on molecular representation, which give related researchers an insight into deep neural network

Table 1
The description of used datasets, including regression tasks and classification tasks.

Dataset	Data Format	Task	Number of molecules	Metric
ESOL	SMILES	Regression	1128	MAE, RMSE
Lipophilicity	SMILES	Regression	4200	MAE, RMSE
HIV	SMILES	Classification	41,127	AUROC, AUPRC
BACE	SMILES	Classification	1513	AUROC, AUPRC

architectures for specific molecular tasks. We consider the following representative models as our baselines.

- RNN & MLP. RNN & MLP is proposed in CheMixNet [12]. It consists of two parallel neural networks RNN (RNN extracts molecular feature using SMILES sequences) and MLP (MLP learns molecular feature using ECFP fingerprint).
- CNN & MLP. CNN & MLP is proposed in CheMixNet [12]. It consists of two parallel neural networks CNN (CNN extracts molecular feature using SMILES sequences) and MLP (MLP learns molecular feature using ECFP fingerprint).
- CNN_RNN. CNN_RNN is proposed in CheMixNet [12]. It is stacked neural networks by CNN and RNN in SMILES sequence.
- MLP. Multilayer perceptron (MLP) model is a feed forward neural network with multiple layers. Each layer is fully connected to the next, which always consists of several dense layers.
- XGBoost. XGBoost [25] is an ensemble method to implement gradient boosting. XGBoost improves model performance by constructing multiple weak learners and gradually reducing prediction errors, which adds a new learner at each step to correct the error of the previous step, and ultimately combines the results of all learners.
- RF. Random forests (RF) [27] is another ensemble learning method based on multiple decision trees, which constructs multiple decision trees through Bagging and feature random selection. Each tree is trained on randomly selected data and feature subsets, and the final prediction result is obtained by averaging or voting on the prediction results of all trees.
- KernelSVM. KernelSVM [26] is a machine learning model based on support vector machine (SVM). KernelSVM maps data to higher dimensions through kernel techniques to identify non-linear decision boundaries.
- GCN. Graph convolutional networks [28] is proposed to learn hidden layer representations and encode node features via first-order approximation of spectral graph convolutions. The entire graph embeddings can be obtained by pooling the node embeddings, that is used to predict various tasks.
- GC. GC is a kind of graph convolutional models, proposed in [21]. GC generalizes standard molecular feature extraction methods based on circular fingerprints, which shows more interpretability on a variety of tasks.
- MPNN. Message Passing Neural Network (MPNN) is a generalized architecture [19] based on graph structure model. MPNN models learn a message passing algorithm and aggregation procedure to compute a function of the entire input graph.
- Weave. Weave [22] uses graph convolutional operation to obtain molecular representations. Weave utilizes atoms, bonds and distances in molecular graph to encode molecules, thereby leveraging the structural information inherent in graphs more effectively.

4.3. Evaluation metrics

In this experiment, the average area under the receiver operating characteristics curve (AUROC) and the area under precision-recall curve (AUPRC) obtained from test sets are used to evaluate the overall performance of the classification tasks. For regression tasks, such as ESOL and Lipophilicity, we use Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) as metric in test set when training and predicting. Receiver operating characteristics (ROC) is a curve, in which the horizontal axis is the false positive rate (FPR), and the vertical axis is the true positive rate (TPR), which are computed as follows:

$$FPR = \frac{FP}{FP + TN} \tag{13}$$

$$TPR = \frac{TP}{TP + FN} \tag{14}$$

Where FP and TN denote the false positive samples and the true negative samples, respectively. Similarly, TP and FN denote the true positive samples and the false negative samples, respectively. Precision Recall Curve (PRC) uses Precision as vertical axis and Recall as horizontal axis. It is an evaluation index for comprehensive evaluation of the overall result. Precision and Recall can be computed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + TN} \tag{16}$$

Where *Precision* and *Recall* denote the Precision and Recall, respectively.

4.4. Implementation details

The experiment is based on Keras framework with Tensorflow [36] as the backend. The methods in Keras such as *EarlyStopping*, *ModelCheckpoint*, *ReduceLROnPlateau*, and *Tensorboard* are used as the callback function when the model is being trained. The hidden layer dimension in graph convolutional network is set to 128, with 2 graph convolutional layers, 0.001 learning rate and 0.2 dropout rate. In CNN, the number of filters and kernel size are set to 32 and 3, respectively. In RNN, LSTM is used to extract features of the sequence. In MLP, the hidden layers dimensions are set to 1024, 512, 256, 64 for performing a fully connected operation. We use ECFP molecular fingerprint to learn structure-based features of the molecules. In addition, all used datasets are split by using random pattern. The training set and test set account for $\frac{3}{4}$ and $\frac{1}{4}$ of the total dataset, respectively. The validation set accounts for $\frac{1}{10}$ of the training set.

4.5. Performance comparison

In this section, we demonstrate the predictive performances for regression task datasets (ESOL, Lipophilicity) and classification task datasets (HIV, BACE) by using mixed deep neural network methods, conventional methods and graph-based methods. Table 2 shows the comparison results of various different models, which shows that the graph-based models achieve the best performance for the datasets of the regression task. MPNN model obtained 0.58 root mean squared error for ESOL, and GCN model obtained 0.5323 mean absolute error for ESOL, which outperforms mixed deep neural network and conventional methods with a large margin. In Lipophilicity dataset, GCN model shows the best sores with 0.5906 root mean squared error and 0.4663 mean absolute error, compared with mixed deep neural network and conventional methods, there is a big gap. Fig. 4 shows the comparison of performance based on mixed deep neural network and graph models

Table 2
The predictive performances for ESOL and Lipophilicity with RMSE and MAE in test set (the smaller the value, the better). The Best results are highlighted in bold.

Model		ESOL		Lipophilicity	
		RMSE	MAE	RMSE	MAE
Mixed deep neural network	RNN & MLP	0.9106	0.6874	0.8553	0.6440
	CNN & MLP	0.8158	0.6141	0.8668	0.6560
	CNN_RNN	1.0419	0.8010	1.0513	0.8282
	MLP	1.1617	0.8517	0.9125	0.6986
Conventional methods	XGBoost	0.99	-	0.799	-
	RF	1.07	-	0.876	-
Graph-based methods	GCN	0.6806	**0.5323**	**0.5906**	**0.4663**
	GC	0.97	-	0.655	-
	MPNN	**0.58**	-	0.719	-
	Weave	0.61	-	0.715	-

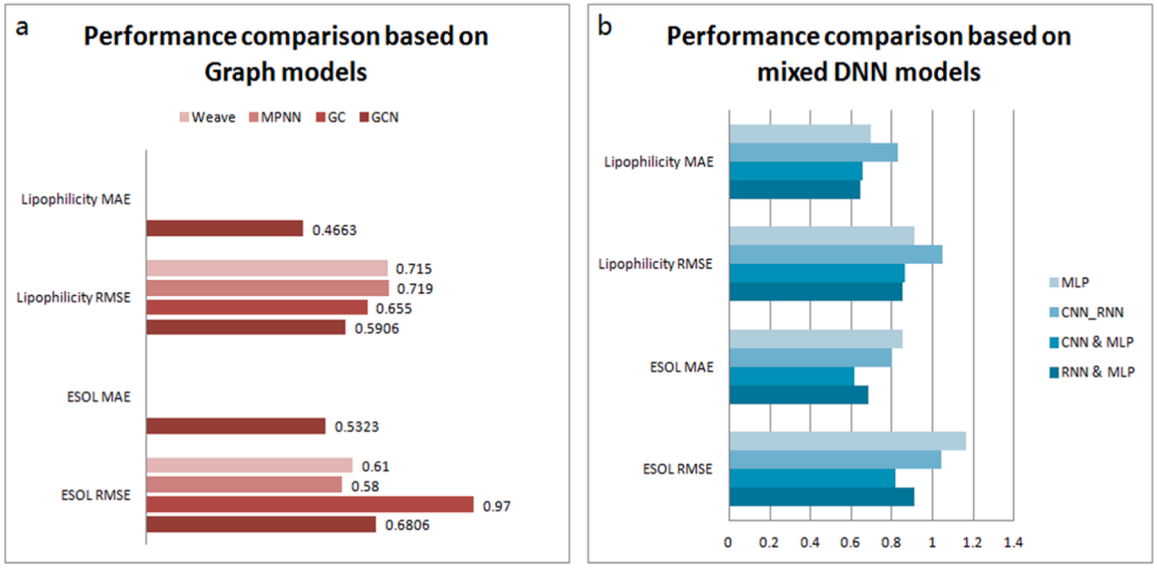

Fig. 4. The performance comparison in ESOL and Lipophilicity based on (a) graph models and (b) mixed deep neural network models. The graph-based model achieves superior performance with the performance scores in (a).

more intuitively in the form of a histogram in ESOL and Lipophilicity datasets.

On the other hand, Table 3 shows the predictive performances in classification tasks such as HIV and BACE. The sample distributions in datasets of the classification task (such as BACE and HIV) can be demonstrated as Fig. 5. Since there is seriously unbalanced data of the positive and negative samples in HIV dataset, we randomly sample negative samples so that the number of negative samples is the same as the number of positive samples. From Table 3, the mixed deep neural network architecture CNN & MLP model obtained the best performance with 0.9763 AUROC and 0.8961 AUPRC for HIV, 0.9670 AUROC for BACE. The AUPRC score in BACE dataset is 0.8249 by using CNN & MLP model, that is slightly lower than the 0.8510 AUPRC score predicted by the GCN model. In general, on the classification datasets, the mixed deep neural network model has achieved superior performance on the task of molecular properties prediction. Fig. 6 shows the comparison of performance based on mixed deep neural network and graph models more intuitively in the form of a histogram in HIV and BACE datasets.

It is worth noting that part of the data in the Table 2 and Table 3 are taken from the reference [17]. Therefore, different neural network models have different performances on different tasks. It is difficult to point out that a certain model is optimal for all tasks. In the research of this paper, the graph neural network shows superior performance on the regression tasks, while the mixed deep neural network models show better performance on the classification tasks.

Table 3
The predictive performances for HIV and BACE with AUROC and AUPRC in test set (the larger the value, the better). The Best results are highlighted in bold.

Model		HIV		BACE	
		AUROC	AUPRC	AUROC	AUPRC
Mixed deep neural network	RNN & MLP	0.9424	0.8833	0.9116	0.8101
	CNN & MLP	**0.9763**	**0.8961**	**0.9670**	0.8249
	CNN_RNN	0.8204	0.8614	0.7429	0.7162
	MLP	0.9725	0.8937	0.9547	0.8213
Conventional methods	XGBoost	0.756	-	0.850	-
	KernelSVM	0.792	-	0.862	-
Graph-based methods	GCN	0.8180	0.8574	0.8751	**0.8510**
	GC	0.763	-	0.783	-
	Weave	0.703	-	0.806	-

4.6. Ablation study

For the mixed deep neural network architectures, we carried out ablation studies on the four datasets. As shown in Table 2 and Table 3, CNN & MLP obtains the best performance among all mixed deep neural network models in all datasets. CNN & MLP consists of two parallel neural networks CNN and MLP. In Table 2, the performances of CNN_RNN and MLP are poor, however, their combination, namely CNN & MLP has achieved superior performance, which is different from Table 3. In Table 3, compared with CNN & MLP, MLP has achieved almost the same performance. Therefore, MLP has the greatest contribution to the CNN & MLP model in the classification task datasets. We use a histogram (Fig. 7) to show the ablation experiments more intuitively.

5. Conclusion

In this study, we explored and gave some insight into mixed deep neural network architectures for molecular representation in the context of molecular properties prediction tasks. Different neural network models learn distinct molecular representations, resulting in varying predictive performance across different tasks. Given that the structure of a molecule can be represented as a graph, which is composed of nodes and edges. Therefore, graph neural networks exhibit competitive performance in processing graph-like data. In this paper, we proposed a novel deep neural network architecture, conducted a lot of experiments and came to the following conclusions: the graph neural network demonstrated superior performance on regression task datasets, which highlights that the graph neural network has a better enhanced ability to fit the data effectively. Conversely, the neural networks within the mixed deep neural network architectures exhibited better performance in the classification task, which indicates that mixed deep neural network possess a notable advantage in data fitting for classification datasets. As demonstrated, graph neural networks excel in regression tasks due to their ability to capture complex relationships within graph-structured data. It shows that GNNs are particularly effective because they can naturally incorporate and process the connectivity and relational information inherent in graphs. On the other hand, mixed deep neural network architectures may be better suited for classification tasks, where they can leverage diverse features effectively and handle high-dimensional data efficiently. Our analysis highlights the necessity

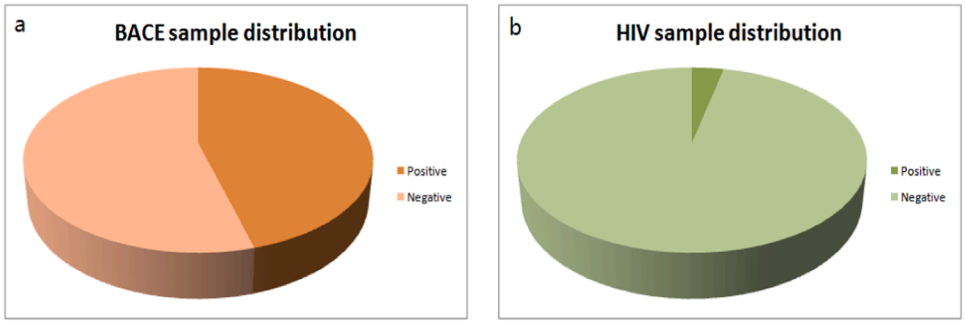

Fig. 5. The sample distributions in datasets of the classification tasks (a) BACE sample distribution and (b) HIV sample distribution.

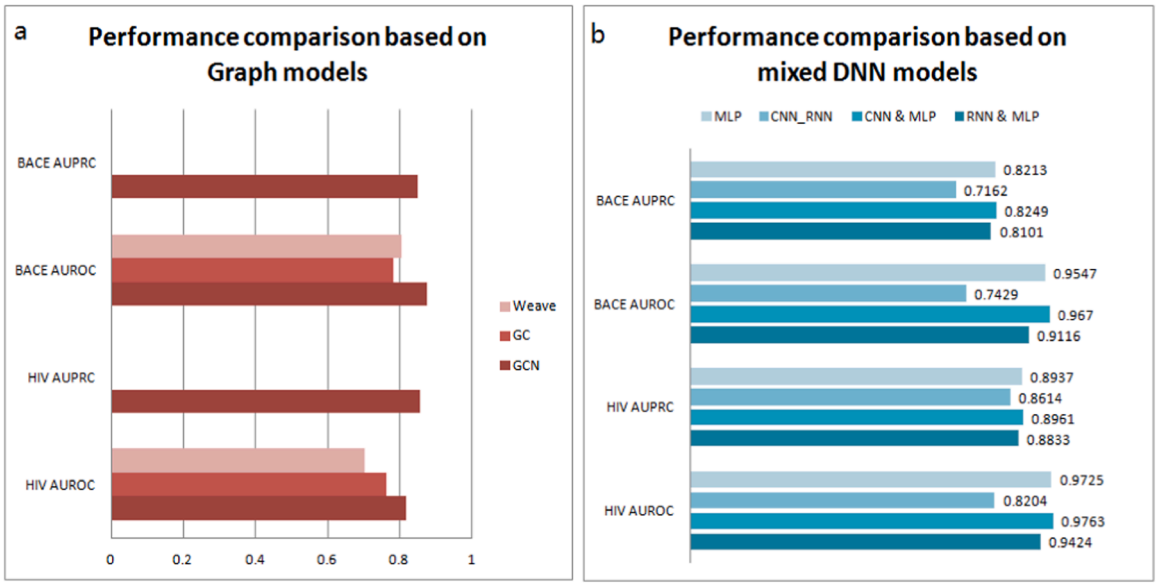

Fig. 6. The performance comparison in BACE and HIV based on (a) graph models and (b) mixed deep neural network models. The mixed deep neural network-based model achieves superior performance with the performance scores in (b).

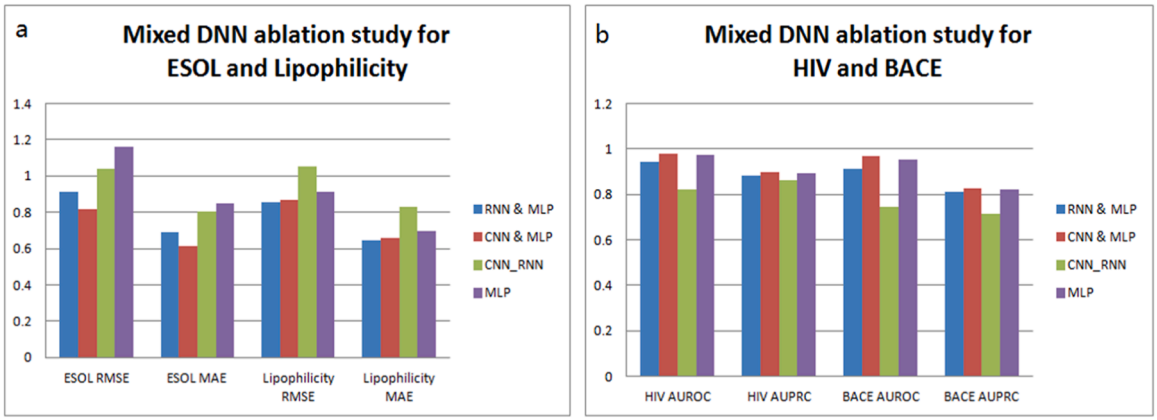

Fig. 7. The ablation results for the mixed deep neural network architectures in (a) ESOL and Lipophilicity datasets, (b) HIV and BACE datasets. The deep neural network models include RNN & MLP, CNN & MLP, CNN_RNN and MLP.

of selecting distinct deep neural network models tailored to specific tasks.

The implementation environment of the project is deliberately kept uncomplicated, as overly complex models tend to exacerbate overfitting issues and compromise generalizability. The proposed model demonstrates potential feasibility for drug applications such as property prediction. Additionally, it is well-suited for feature extraction and learning of sequence structures using deep neural networks. However, this study does not take into account the 3D structural features of molecules. Existing researches [37–40] indicate that these 3D structural features are crucial for understanding molecular interactions in three-dimensional space and their interactions with target proteins.

Subsequent investigations will persist in the domain of molecular representation and molecular properties prediction. Our forthcoming endeavors will center on the development of diverse molecular representation algorithms customized for individual tasks as well as related algorithm optimization. Concurrently, we will prioritize solving the problem of feature extraction and representation learning for molecules in the 3D field, as well as interpretability challenges.

CRediT authorship contribution statement

Shaocong Cheng: Validation, Data curation. **Chunyan Li:** Writing – review & editing, Resources, Methodology, Formal analysis. **Zhenyu Yin:** Writing – review & editing, Software. **Yong Lu:** Supervision, Resources, Methodology, Conceptualization. **Tianze Zhao:** Writing – original draft, Investigation.

Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in the manuscript entitled “Insight Mixed Deep Neural Network Architectures for Molecular Representation”.

References

[1] D. Wigh, J. Goodman, A. Lapkin, A review of molecular representation in the age of machine learning, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 12 (5) (2022) e1603, https://doi.org/10.1002/wcms.1603.

[2] Y. Harnik, A. Milo, A focus on molecular representation learning for the prediction of chemical properties, Chem. Sci. 15 (2024) 5052–5055.

[3] S. Han, H. Fu, Y. Wu, G. Zhao, Z. Song, F. Huang, Z. Zhang, S. Liu, W. Zhang, HimGNN: a novel hierarchical molecular graph representation learning framework for property prediction, Brief. Bioinforma. 24 (5) (2024).

[4] H. Li, R. Zhang, Y. Min, D. Ma, D. Zhao, J. Zeng, A knowledge-guided pre-training framework for improving molecular representation learning, Nat. Commun. 14 (1) (2023) 1–13.

[5] Z. Zhang, Q. Liu, C. Lee, C. Hsieh, E. Chen, An equivariant generative framework for molecular graph-structure Co-design, Chem. Sci. 14 (2023) 8380–8392.

[6] J. Fromer, C. Coley, Computer-aided multi-objective optimization in small molecule discovery, Patterns 4 (2023).

[7] Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang, M. Zheng, Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism, J. Med. Chem. 63 (16) (2020) 8749–8760.

[8] Li Y., Pei J. and Lai L. Learning to Design Drug-like Molecules in Three-dimensional Space Using Deep Generative Models. ArXiv, 2021. Available: ⟨https://arxiv.org/abs/2104.08474⟩.

[9] C. Li, J. Wang, Z. Niu, J. Yao, X. Zeng, A spatial-temporal gated attention module for molecular property prediction based on molecular geometry, Brief. Bioinforma. 22 (5) (2021).

[10] L. Huang, H. Luo, S. Li, F. Wu, J. Wang, Drug-drug similarity measure and its applications, Brief. Bioinforma. 22 (4) (2021) bbaa265, https://doi.org/10.1093/bib/bbaa265.

[11] Xu M., Luo S., Bengio Y., Peng J. and Tang J. Learning Neural Generative Dynamics for Molecular Conformation Generation. International Conference on Learning Representations (ICLR). 2021.

[12] A. Paul, D. Jha, R. Bahrani, W. Liao, A. Choudhary, A. Agrawal, CheMixNet: mixed DNN architectures for predicting chemical properties using multiple molecular representations, Proc. Nips Workshop Mach. Learn. Mol. Mater. (2018).

[13] C. Li, W. Wei, J. Li, J. Yao, X. Zeng, Z. Lv, 3DMol-Net: Learn 3D Molecular Representation using Adaptive Graph Convolutional Network Based on Rotation Invariance, IEEE J. Biomed. Health Inform. 26 (10) (2022).

[14] W. David, SMILES: a chemical language and information system, J. Chem. Inf. Comput. Sci. 28 (1) (1988) 31–36.

[15] D. Rogers, M. Hahn, Extended-connectivity fingerprints, J. Chem. Inf. Model. 50 (5) (2010) 742–754.

[16] S. Kearnes, K. Mccloskey, M. Berndl, V. Pande, P. Riley, Molecular graph convolutions: moving beyond fingerprints, J. Comput. Aided Mol. Des. 30 (8) (2016) 595–608.

[17] Z. Wu, B. Ramsundar, E. Feinberg, J. Gomes, C. Geniesse, A. Pappu, K. Leswing, V. Pande, MoleculeNet: A benchmark for molecular machine learning ([Online]. Available), Chem. Sci. vol. 9 (2) (2018) 513–530, https://doi.org/10.1039/C7SC02664A.

[18] X. Wang, Z. Li, M. Jiang, S. Wang, S. Zhang, Z. Wei, Molecule Property Prediction Based on Spatial Graph Embedding, J. Chem. Inf. 59 (9) (2019) 3817–3828.

[19] Gilmer J., Schoenholz S., Riley P., Vinyals O. and Dahl G. Neural Message Passing for Quantum Chemistry. In Proceedings of the 34th International Conference on Machine Learning (ICML), 2017, Volume 70: 1263–1272.

[20] S. Ryu, Y. Kwon, W. Kim, A Bayesian Graph Convolutional Network for Reliable Prediction of Molecular Properties with Uncertainty Quantification, Chem. Sci. 10 (36) (2019) 8438–8446.

[21] D. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarelli, T. Hirzel, A. Guzik, R. Adams, Convolutional networks on graphs for learning molecular fingerprints, Adv. Neural Inf. Process. Syst. (NIPS) (2015) 2224–2232.

[22] S. Kearnes, K. Mccloskey, M. Berndl, V. Pande, P. Riley, Molecular graph convolutions: moving beyond fingerprints, J. Comput. -Aided Mol. Des. 30 (8) (2016) 595–608.

[23] S. Hochreite, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[24] X. Lin, Z. Quan, Z. Wang, H. Huang, X. Zeng, A novel molecular representation with BiGRU neural networks for learning atom, Brief. Bioinforma. 21 (6) (2019) 2099–2111.

[25] Chen T. and Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 785–794.

[26] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.

[27] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[28] Kipf T. and Welling M. Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations (ICLR), 2017.

[29] D. Rumelhart, G. Hinton, R. Williams, Learning representations by back-propagating errors, Nature 323 (1986) 533–536.

[30] Bruna J., Zaremba W., Szlam A. and LeCun Y. Spectral networks and locally connected networks on graphs. International Conference on Learning Representations (ICLR), 2014. [Online]. Available: ⟨http://arxiv.org/abs/1312.6203⟩.

[31] Defferrard M., Bresson X., and Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. Advances in Neural Information Processing Systems (NIPS), 2016.

[32] J. Delaney, ESOL: estimating aqueous solubility directly from molecular structure, Chem. Inf. Comput. Sci. 44 (3) (2004) 1000.

[33] Wenlock M. and Tomkinson N. Experimental in vitro DMPK and physicochemical data on a set of publicly disclosed compounds. [Online]. Available: https://doi.org/10.6019/chembl3301361.

[34] Zaharevd. Aids antiviral screen data. [Online]. 2004, Available: ⟨https://wiki.nci.nih.gov/display/NCIDTPdata/AIDS+Antiviral+Screen+Data⟩.

[35] G. Subramanian, B. Ramsundar, V. Pande, R. Denny, Computational modeling of beta-ecretase 1 (bace-1) inhibitors using ligand based approaches. Chemical Information and Modeling 56 (10) (2016) 1936–1949.

[36] Abadi M., Barham P., Chen J., Chen Z., Davis A., Dean J., Devin M., Ghemawat S., Irving G., Isard M., Kudlur M., Levenberg J., Monga R., Moore S., Murray D., Steiner B., Tucker P., Vasudevan V., Warden P., Wicke M., Yu Y. and Zheng X. TensorFlow: A system for large-scale machine learning. in Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, ser. OSDI 2016. USA: USENIX Association, 2016: 265–283.

[37] Wang X., Zhao H., Tu W. and Yao Q. Automated 3D Pre-Training for Molecular Property Prediction. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023: 2419-2430.

[38] C. Pang, H. Tong, L. Wei, Advanced deep learning methods for molecular property prediction, Quant. Biol. 11 (4) (2023).

[39] L. Cai, Y. He, X. Fu, L. Zhuo, Q. Zou, X. Yao, AEGNN-M: a 3D graph-spatial co-representation model for molecular property prediction, IEEE J. Biomed. Health Inform. (2024), https://doi.org/10.1109/JBHI.2024.3368608.

[40] R. Zhang, R. Yuan, B. Tian, PointGAT: a quantum chemical property prediction model integrating graph attention and 3D ceometry, J. Chem. Theory Comput. 20 (10) (2024) 4115–4128.