FP-GNN++: Towards Accurate Molecule Property Classification via Leveraging Versatile Features

Zhangzhi Xiong
ShanghaiTech University
2023533146
xiongzhzh2023@shanghaitech.edu.cn

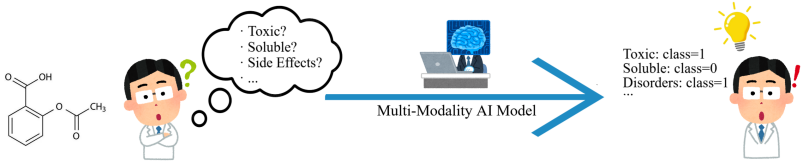
Yixuan Chen
ShanghaiTech University
2023533082
chenyx2023@shanghaitech.edu.cn

Tianni Yang
ShanghaiTech University
2023533107
yangtn2023@shanghaitech.edu.cn

Abstract

Prediction of molecular properties, such as toxicity, solubility, etc., plays a critical role in drug discovery. Traditional experimental approaches to assess these properties are often costly and time-consuming, making deep learning, with classification as one of its most fundamental applications, a powerful tool in this domain. In this paper we propose FP-GNN++, a powerful classification model leveraging versatile features from multi-modality. We utilize cross attention mechanism and graph multi-head attention for better feature fusion. Furthermore, we introduce in chemical bond information in graph processing component as strong inductive bias in the field of chemistry. Our model shows better performance in various dataset and split type experiments.

1 Introduction



In molecular property prediction, molecular fingerprints and molecular graphs are two of the most commonly used formats. To make use of complementary strengths of them, researchers have proposed multi-modal models, among which FP-GNN (Fingerprint-enhanced Graph Neural Network)[1] is a representative approach. However, there remains room for improvement in how the two representations are interpreted and combined in the model. Building upon the FP-GNN framework, we propose our enhanced model called FP-GNN++.

To summarize, the key contributions of this project are:

- Zhangzhi Xiong: Chemical Bond Information, Graph multi-head and Cross-Attention mechanism, code implementation, report, experiment, presentation
- Yixuan Chen: Self-attention and GraphSAGE exploration, code implementation, report, presentation
- Tianni Yang: Experiment, Dataset and Baseline Reproduce, report

2 Related Work

In recent developments in deep learning for molecule property prediction, two primary forms of molecular representation have gained prominence: fingerprint-based and graph-based representations.

Fingerprint-based models represent molecules as vectors that encode specific substructures. Commonly used fingerprints include MACCS keys[2], which capture predefined chemical patterns; Pharmacophore ErG fingerprints[3], which reflect pharmacologically relevant features; and PubChem fingerprints[4], which encode a wider range of substructures. Models that well leverage these features include XGBoost[5] and AttentiveFP[6].

Graph-based models, in contrast, treat molecules as graphs, where atoms are nodes and chemical bonds are edges, and learn molecular properties by aggregating information from neighboring atoms. Representative models include Graph Convolutional models (GC, Weave[7]), Graph Attention Networks (GAT)[8]. Note that these models only exploit the neighboring atom information in graph.

MoleculeNet[9] includes datasets for a wide range of classification tasks, such as toxicity, bioactivity, etc., and provides standard splits (including random splits and scaffold splits, the latter one ensures molecules with the same core chemical scaffold are placed in the same subset so as to test generative capability rigorously) and baseline results for various models, including GC, Weave[7], XGBoost[5] and AttentiveFP[6]. These benchmarks will be the standard testbed for us to evaluate our models.

3 Methodology

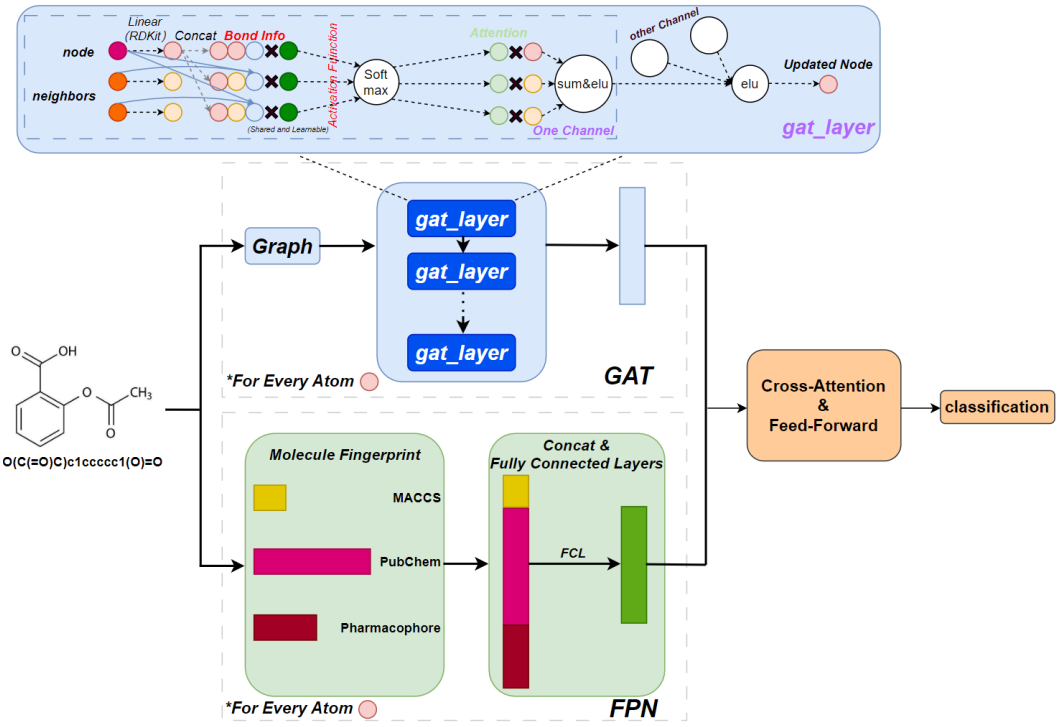


Figure 1: Pipeline of FP-GNN++

3.1 Overall Architecture

Overall, our FP-GNN++ model consists of three main components as shown in Figure 1:

· A Graph Attention Network (GAT) module that learns representations from molecular graphs.
· A Fingerprint Network (FPN) module that processes molecular fingerprints.
· A Cross-Attention Fusion mechanism that effectively integrates information from both modality.

3.2 Graph Attention Network Module

The GAT module processes the molecular graph structure to capture atomic interactions and structural patterns. GAT consists of several gat_layer. Each molecule is represented as a graph where atoms are nodes and chemical bonds are edges. Our GAT also supports multi-head mechanism for robustness.

Using RDkit, we initialize the feature vector for every atom. For each node (atom) in the graph, the gat_layer computes attention coefficients that determine the importance of neighboring nodes and then update the node feature vector accordingly. For a specific node atom i in one specific channel, we collect the index set that represent atom's neighbors or itself:

$$\mathcal{N}_i = \{j \mid Atom_j \; is \; Atom_i's \; neighbor \; or \; itself\} \tag{1}$$

Then un-softmaxed scores are computed as follow:

$$e_{ij} = \textit{Activation Function}\left(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j \| \mathbf{E}_{ij}]\right), j \in \mathcal{N}_i \tag{2}$$

where $\mathbf{h}_i$ is the feature vector of atom i, $\mathbf{W}$ is a learnable weight matrix for linear operation, $\mathbf{a}$ is a shared and learnable matrix, $\mathbf{E}_{ij}$ represents the edge (chemical bond) information between atoms i and j, and $\|$ denotes concatenation. Inspired by the research on the important role chemical bond plays in machine learning applied in chemistry[10], we assign discrete label (represented in number) for $\mathbf{E}$ to represent Single, Double, Triple or Aromatic chemical bond. Note that for the bond between the atom and itself is marked as label 'None'. The attention coefficients are normalized using softmax. The updated node representation in one channel is then computed as:

$$\mathbf{h}_i'^c = elu\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}\mathbf{W}\mathbf{h}_j\right), \; where \; \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i}\exp(e_{ik})} \tag{3}$$

where $\mathbf{h}_i'^c$ stands for the updated atom feature vector in the c_{th} channel. To enhance the expressiveness of the network, we employ multi-head attention heads. In Graph Attention Networks [8], it is mentioned that both concatenation or averaging style is acceptable for multi-head attention in graphs. For expressive robustness, we choose to concatenate along the feature dimension:

$$\mathbf{h}_i' = elu(concat(\mathbf{h}_i'^c)) \tag{4}$$

In the multi-head of the last gat_layer, we choose to eventually alter the dimension to *hidden_dim*. Through stacking gat_layers, we set the result after the last layer as the input $\mathbf{z}_{GAT}$ of cross-attention stage.

3.3 Fingerprint Network Module

The FPN module processes traditional molecular fingerprints to capture chemical patterns. We utilize a combination of three types of fingerprints:

· MACCS keys[2] (167 dim):
· ErG Pharmacophore fingerprints[3] (881 dim):
· PubChem[4] fingerprints (441 dim):

With given SMILES, fingerprint toolkit will give every atom a feature vector so that the molecule tensor can encode the corresponding information. These fingerprints are concatenated to form a comprehensive feature vector of 1489 dimensions:

$$\mathbf{f} = [\mathbf{f}_{\text{MACCS}} \| \mathbf{f}_{\text{PubChem}} \| \mathbf{f}_{\text{Pharmacophere}}] \tag{5}$$

Hence, we choose to process fingerprint vector through a simple two-layer neural network to alter dimension into *hidden_dim*.

3.4 Cross-Attention Fusion Mechanism

To effectively integrate information from both the GAT and FPN modules, instead of simply concatenating features and using MLP for regression, we use a cross-attention fusion mechanism. This mechanism is more robust and can fuse the two modality features more deeply.

More specifically, we choose GAT feature atom sequence as $\mathbf{q}$ and FPN feature atom sequence as $\mathbf{k}, \mathbf{v}$. After calculating the softmaxed attention score and use $\mathbf{v}$ to update the tensor accordingly, a residual connection[11] is used empirically. Then we use a two-layer feed-forward network for regression down-streaming task head.

4 Experiments

4.1 Implementation Details

In our implementation, GAT comprises 8 gat_layer and supports multi-head attention with 8 heads, processing atom features of dimension 133 (derived from RDKit) and outputting 60-dimensional features per head, which are concatenated and mapped to a hidden dimension of 300. The FPN processes mixed molecular fingerprints with an initial dimension of 1489, reduced to 512 via a fully connected layer, and further mapped to 300. The fused representation, with a hidden dimension of 300, is fed into a feed-forward network with two linear layers and ReLU activation, producing a single output for classification. No dropout is applied. All experiments are run on a single GeForce RTX2080Ti GPU. All experiments are run with three different seeds.

4.2 Results and Discussion

Table 1: Model Comparison Results

Dataset Name	Split type	GC	WEAVE	AttentiveFP	XGBoost	FP-GNN	Ours
BACE	Random			0.876	0.889	0.881	**0.903**±0.007
	Scaffold	0.783	0.806			**0.860**	0.828±0.009
HIV	Random			0.824	0.816	0.825	**0.835**±0.016
	Scaffold	0.763	0.703			0.824	**0.825**±0.008
Tox21	Random	0.829	0.820	**0.852**	0.836	0.815	0.823±0.005
BBBP	Random			0.887	0.926	**0.935**	0.907±0.008
	Scaffold	0.690	0.671			0.916	**0.931**±0.012
Clintox	Random	0.807	0.832	0.904	**0.911**	0.840	0.870±0.011
SIDER	Random	0.638	0.581	0.623	0.642	0.661	**0.674**±0.005

As presented in Table 1, we compare our model with graph-based model(GC, Weave), Fingerprint-based model(AttentiveFP, XGBoost) and baseline FP-GNN which is a simple network trying to integrate graph and fingerprint information. Our model's results show better performance.

Table 2: Ablation Study Results

Dataset Name	Split type	FP-GNN	Ours	Bond$^-$	CrossAttn$^-$	SelfAttn	GraphSAGE
BACE	Random	0.881	0.903±0.007	0.889±0.031	0.886±0.027	0.640±0.033	0.871±0.017
BBBP	Scaffold	0.916	0.931±0.012	0.901±0.023	0.911±0.025	0.816±0.085	0.871±0.011
SIDER	Random	0.661	0.674±0.005	0.672±0.013	0.679±0.008	0.584±0.014	0.627±0.021

We conducted an ablation study to test the benefit effect of chemical bond information and cross-attention mechanism in Table 2. We've also tried self-attention on the atom sequence with concatenated fingerprint features and replace Graph Multi-head Attention with GraphSAGE which is an advanced model based on GC. We replace the corresponding network component based on the baseline of FP-GNN and result in poor performance as presented in Table 2. We believe this is because, empirically, Graph Convolution mechanism is less robust and powerful in expressiveness compared to Graph Attention, and self-attention might underperform with lengthy concatenated fingerprints since these features often lack the inherent sequential or relational structure that self-attention mechanisms are designed to exploit. All ablation studies are run on three representative experiments.

5 Conclusion

In this paper, we introduce FP-GNN++, a more powerful molecule property classifier via leveraging versatile features, which features chemical bond information, multi-head graph attention, and a cross-attention mechanism for the fusion of multi-modality features. Our model shows better performance on several dataset and split type experiments.

Limitation Our approach didn't leverage more recently and more developed feature extractors which utilize more advanced chemical inductive bias. Our implementation of self-attention may also possibly account for its poor performance besides the reasons we analyzed before. We haven't systematically tested the potential benefits of the number of cross-attention blocks and gat_layers.

References

[1] Hanxuan Cai, Huimin Zhang, Duancheng Zhao, Jingxing Wu, and Ling Wang. Fp-gnn: a versatile deep learning architecture for enhanced molecular property prediction. *Briefings in Bioinformatics*, 23(6):bbac408, 2022.

[2] Joseph L. Durant, Burton A. Leland, Douglas R. Henry, and James G. Nourse. Reoptimization of mdl keys for use in drug discovery. *Journal of Chemical Information and Computer Sciences*, 42(6):1273–1280, 2002. PMID: 12444722.

[3] Nikolaus Stiefl, Ian A. Watson, Knut Baumann, and Andrea Zaliani. Erg: 2d pharmacophore descriptions for scaffold hopping. *Journal of chemical information and modeling*, 46 1:208–20, 2006.

[4] Yanli Wang, Jewen Xiao, Tugba Suzek, Jian Zhang, Jiyao Wang, and Stephen Bryant. Pubchem: A public information system for analyzing bioactivities of small molecules. nucleic acids research, 37, w623-w633. *Nucleic acids research*, 37:W623–33, 07 2009.

[5] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794. ACM, August 2016.

[6] Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, and Mingyue Zheng. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of Medicinal Chemistry*, 63(16):8749–8760, 2020. PMID: 31408336.

[7] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of Computer-Aided Molecular Design*, 30(8):595–608, 2016.

[8] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Fp-gnn: a versatile deep learning architecture for enhanced molecular property prediction. *arXiv preprint arXiv:1710.10903*, 2017.

[9] Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. Moleculenet: A benchmark for molecular machine learning. 2018.

[10] Liu Yidi, Yang Qi, Li Yao, Zhang Long, and Luo Sanzhong. Application of machine learning in organic chemistry. *Chinese Journal of Organic Chemistry*, 40(11):3812, 2020.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.