

Generalized Graph Neural Network for prediction of molecular property

Lilapati Waikhom

*Department of Computer Science & Engineering
National Institute of Technology Silchar
Cachar-788010, Assam, India
lilapati_rs@cse.nits.ac.in*

Ripon Patgiri

*Department of Computer Science & Engineering
National Institute of Technology Silchar
Cachar-788010, Assam, India
ripon@cse.nits.ac.in*

Abstract—The blood-brain barrier in humans is a membrane for preventing solutes from entering the central nervous system. The blood-brain barrier is a network of blood vessels and tissue composed of tightly spaced cells that aid in the prevention of hazardous chemicals from reaching the brain. Some substances, including water, oxygen, carbon dioxide, and general anesthetics, can pass through the blood-brain barrier and enter the brain. It also prevents germs and other chemicals entering, such as many anticancer medications. Any new medicine design must consider the barrier's presence and create new molecules that target central nervous system components or find new chemicals that do not permeate it. Several studies have been conducted to identify blood-brain barrier penetration in the literature, but with poor success, if any, relevance to real-world drug development and discovery programs. The fact that only approximately 2% of tiny compounds can cross the blood-brain barrier is part of the reason. This paper proposes a generalized neural network model that incorporates the relationship between molecules and their attributes to predict its blood-brain barrier permeability property. Results show considerable improvement over the traditional machine learning-based models.

Index Terms—Graph, Neural Network, Graph Neural Network, Weave, Graph Convolution, Bypass multitask network, SingleTask, Influence relevance voting, Random Forest, XGBoost, Kernel Support vector classification, Logistic Regression.

I. INTRODUCTION

Graphs can describe many real-world problems and social settings. Several tasks in such settings can be classified as Node Classification, Graph Classification, and Link Prediction in graphs. The most challenging among them is the task of link prediction. For example, citations prediction among articles is frequently translated into link prediction [1] in citations graph. As a result, the predicted citations will benefit future academic studies. Predicting whether two nodes in a graph are likely to have a link is known as link prediction [2]. The most promising technique to solving the link prediction issue is based on recent developments in Graph Neural Network (GNN) [3]. Nonetheless, most GNNs learn node embeddings through a shallow neighborhood that contains information only two or three hops away from each node [4]. The performance of GNNs on the link prediction task is negatively affected due to the shallowness of the neighborhood and structural difficulties present in the citations graph.

As graphs are so common, they have a wide range of applications, including metabolic network reconstruction [5], friend

recommendation [6], knowledge graph completion [7], and movie recommendation [8]. One of the popular applications is to predict the blood-brain barrier permeability property of the molecule. The blood-brain barrier is a membrane that divides the blood from the extracellular fluid in the brain, preventing most medications (molecules) from reaching the brain. As a result, the blood-brain barrier permeability property has been necessary to research in order to develop new medications that target the central nervous system [9]. The molecules in a drug are naturally represented as an undirected graph. The atoms are represented as nodes, and the relationships of the atoms are represented as edges or links. Traditional approaches such as random forests, support vector machines, etc., have been widely used for predicting molecular properties. Unlike GNNs, traditional techniques rely on pre-computed properties of molecules, for example, molecular charge, polarity, weight, and the number of carbon atoms. Although these molecular attributes are good predictors of various molecular properties, it is hypothesized that operating on these more basic and low-level features could be even more effective [10]. Several works such as [11], [12] have been published from the past decades. The overview of the traditional works can be found in [13], and some of the recent review papers that are focused on recent publications are [14], [15]. Several approaches rely on physicochemical descriptors combinations such as polarizability, surface area, ionization, lipophilicity, molecular size, polarity, and hydrogen bonding capacity [16]. Simple rules or ground scores to predict blood-brain barrier-penetrating molecules that are identical to the drug-likeness filters are already defined in [17].

This paper introduces generalized GNN for predicting molecular blood-brain barrier property. Section II introduces some traditional as well as state-of-the-art approaches for human blood-brain barrier prediction. We described the proposed system in Section III and describe the dataset we have used for the experiments in Section IV-A. The experimental setup is elaborated in Section IV. In Section V, we provide the results that we have found from the experiment, and in Section VI, we conclude the paper.

![ht]

TABLE I
SMILES REPRESENTATION AND PERMEABILITY PROPERTY OF SOME MOLECULES

Name	Permeable/ Non-permeable	SMILES
Propranolol	1	[Cl].CC(C)NCC(O)COc1cccc2ccccc12
Terbutylchlorambucil	1	C(=O)(OC(C)(C)C)CCCC1ccc(cc1)N(CCCl)CCCl
Ondansetron	1	Cn1c2CCC(Cn3ccnc3C)C(=O)c2c4cccc14
Tymazoline	0	CC(C)c1ccc(C)cc1OCC2=NCCN2
Vindesine	0	CC[C@]1(O)C[C@H]2CN(CCc3c([nH]c4cccc34)[C@@@](C2)(C(=O)OC)c5cc6c(cc5OC)N(C)[C@H]7[C@](O)([C@H](O)[C@]8(CC)C=CCN9CC[C@]67[C@H]89)C(N)=O)C1
Zorubicin	0	[Cl-].COc1cccc2C(=O)c3c(O)c4C[C@](O)(C[C@H](O[C@H]5C[C@H](N)[C@H](O)[C@H](C)O5)c4c(O)c3C(=O)c12)/C(C)=N/NC(=O)c6cccc6.[H+]

II. RELATED WORKS

Graph neural networks seem to become progressively important and effective in research areas like drug discovery, and several pharmaceutical corporations are enthusiastic about incorporating these techniques into their in-house frameworks [18], [19]. This is very attractive for activities like molecular property prediction, which is frequently one of the most critical tasks in computer-aided drug development processes [20]. Predicting molecular properties is also one of the encouraging applications of GNN, as a molecule can easily be described as a topological graph that treats atoms in the form of nodes and the relationships between atoms as edges. In contrast to alternative molecular representations that represent molecules patterns, however, in these types of representations, there is the risk of structural information loss, even atoms, and bonds between them. The prediction task of molecular property is analogous to a supervised graph classification task in this way. Some common example of molecular property are protein interface prediction [21] and toxicity prediction [22].

Chemistry experts use their domain knowledge to design the molecular fingerprint manually, such as the traditional methods ECFP [23]. Several studies have used deep learning approaches to improve molecular representation. One perspective is to use the SMILES molecular representation [24]. Based on SMILES [25], [26] use RNN-based models to generate the molecular fingerprint. Another promising perspective is the study of the graph structure of a molecule using graphical neural networks (GNN), which has recently elicited much interest [27], [28]. Several of these GNN-based models generate molecular representations. For example, Duvenaud et al., [29] applies convolution networks to molecular graphs to generate molecular fingerprints, and Tang et al., [30] presented low dimensional node embeddings for the huge graphs. Park et al., [31] presented a hyperbolic representation of learning using message passing auto-encoders. In contrast, Kipf and Welling [32] proposed an approach or learning interpretable latent representations for undirected graphs. Park et al., [33] introduced a method for extracting low-dimensional underlying representations from a graph in irregular domains. Further, Cui et al., [34] introduced a graph embedding for learning

vector from the features of the nodes and the topology of the graph. While, Pan et al., [35] introduced a method for graph-structured data representation in a low dimensional space for graph analytics. Hu et al., [36] proposed a generic structural feature extraction by pretraining the GNNs. Peng et al., [37] introduced a subtask to detect meta-paths using node embeddings. Rong et al., [38] working on extracting semantic information, and rich structural from massive volumes of unlabeled molecular data. Cao et al., [39] proposed a method for bipartite graph embedding that maximizes the mutual information between graphs.

Ma et. al., [40] proposed a multi-view message passing architecture that allows for more accurate predictions of molecular characteristics and named it Multi-View Graph Neural Network. Song et al., [41] presented a GNN model called Communicative Message Passing Neural Network to enhance molecular embedding by improving message exchanges between nodes and edges via a communicative kernel. Fu et al., [42] introduces a novel graph neural network technique, Hierarchical Interaction Network for predicting clinical trial outcomes that accommodates complicated interaction patterns from multi-modal data.

III. PROPOSED SYSTEM

A. SMILES Data Representation

Simplified molecular-input line-entry system (SMILES) is a line notation specification for representing the structure of chemical species using short ASCII characters. Most molecular editors can accept SMILES strings and convert them back into two-dimensional drawings or three-dimensional models of the molecules. Table I shows the SMILES of some molecules and their permeability property.

B. Defining Features

We encode the atoms' features and bonds' features from the SMILES representation to streamline the graph representation of the molecules. Here, we consider only the symbol of the element, the total number of bonds between hydrogen, the total number of valence electrons, orbital hybridization for the atom features and bond type, and conjugation for the bond features.

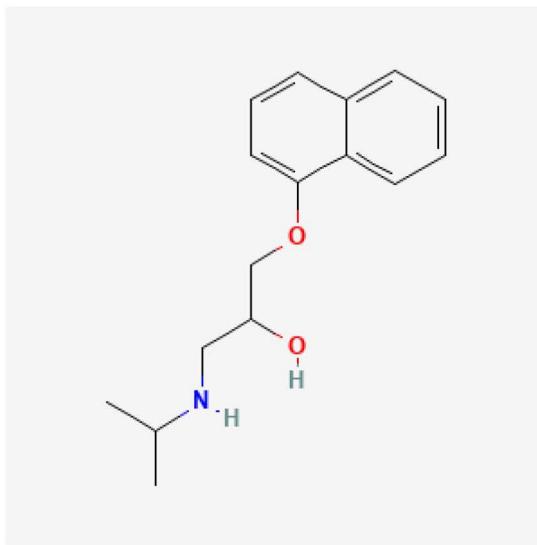


Fig. 1. Molecular representation of Propranolol

C. Building Graphs from SMILES

Before building the graphs from molecules, we need to abstract the generation of molecules from SMILES representation. Further, graph generation from molecules' representation needs to be abstracted. We implement both functionalities to create the graphs from the SMILES by applying both abstractions implemented in the form of callable functions. These functions are subsequently executed on all the SMILES representations of the dataset's training set, testing set, and validation splits. The abstracted functions are described below:

- The first function takes SMILES string as input and provides a molecule object as the output.
- The second function takes molecules objects and gives a graph as the output, which is in the form of three tuples {atom_features, bond_features, pair_indices}.

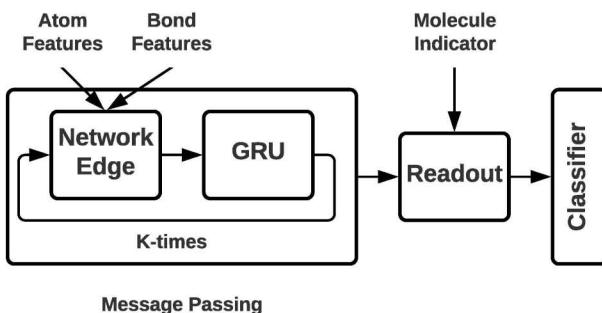


Fig. 2. Architecture of the proposed model.

D. Model Architecture

The proposed model is made up of three steps: (i) Message Passing, (ii) Readout, and (iii) Classification. The architecture is shown in Figure 2.

1) *Message Passing* : The message passing step of the model includes the following two main blocks.

- The *edge network* passes messages to the node v from the 1-hop neighbors u_i based on the edge features e_{vu_i} , and giving an updated state of the node v' . Here u_i stands for the i^{th} neighboring node of v . The ability of the edge network to support non-discrete edge features is an essential feature. However, we are going to use only the discrete edge features.
- The *gated recurrent unit (GRU)* receives the most recent node state as input and updates it based on prior node states. In other words, the most recent node states are fed into the GRU, while the earlier node states are recorded in the GRU's memory. This allows for information to travel from one node state to another.

Both steps are repeated for k times that is the number of hops, where at each iteration of the information aggregated from the source node v grows by one. In our experiment, we considered $k = 4$.

2) *Readout* : When the message passing phase is completed, the k -step aggregated states of the node are divided into subgraphs (one for each molecule in the batch), and then subsequently reduced to graph-level embeddings. Our approach uses a transformer encoder followed by average pooling for readout:

- The subgraphs will be formed from the k -step-aggregated node states (corresponding to each molecule in the batch).
- After that, each subgraph will be padded to match the subgraph with the most nodes, and then stacked.
- To ensure that the paddings do not interfere with training, the stacked and padded encoding of subgraphs (each subgraph comprising a set of node states) are masked.
- Finally, the encoding is sent to the transformer, which then performs average pooling.

IV. EXPERIMENTS

A. Datasets

There are 2,050 molecules in the blood-brain barrier dataset [43]. Each molecule has a name, a label, and a SMILES string. This dataset's labels are binary (1 or 0), indicating the permeability of the compounds. SMILES gives the structure of a given molecule as an ASCII text. The SMILES string is a concise encoding that is somewhat human-readable for smaller molecules. Encoding molecules as a string simplifies and speeds up molecule searches in databases. The dataset also includes features related to atoms and bonds between them. The features collected for atoms are the number of valence electrons, symbol (element), number of hydrogen bonds, and orbital hybridization. The bond features collected in the dataset are (covalent) bond type and conjugation.

B. Experimental Setup

We first constructed graphs from SMILES using RDKit which can convert the SMILES to molecular objects very quickly and easily. After converting, we then extract sets of

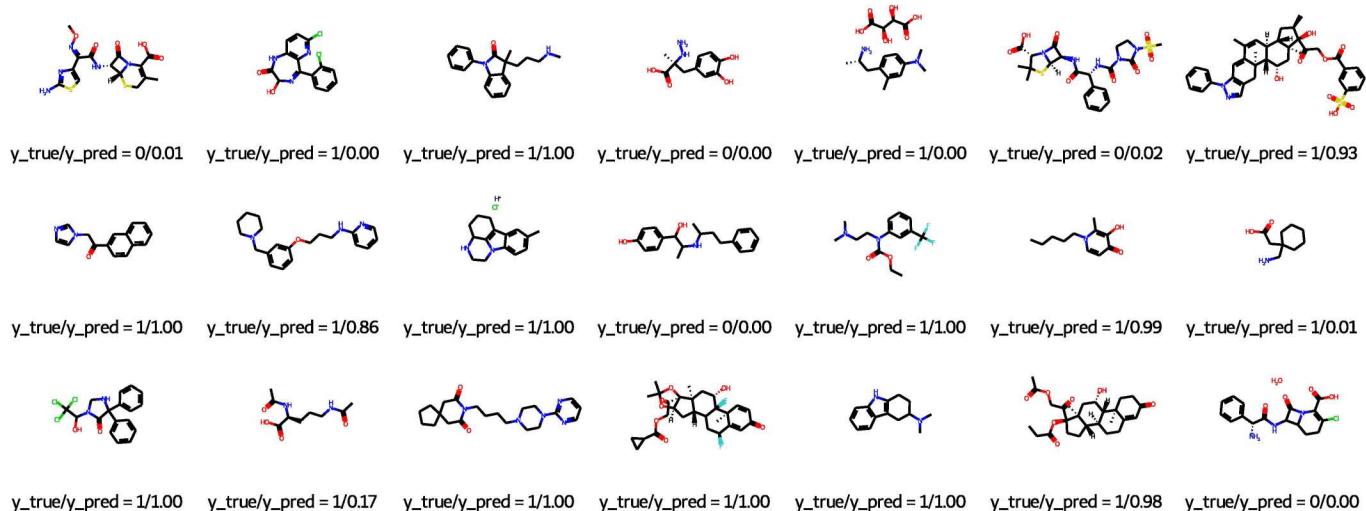


Fig. 3. Inference of the proposed model over few molecules for the blood-brain barrier permeability.

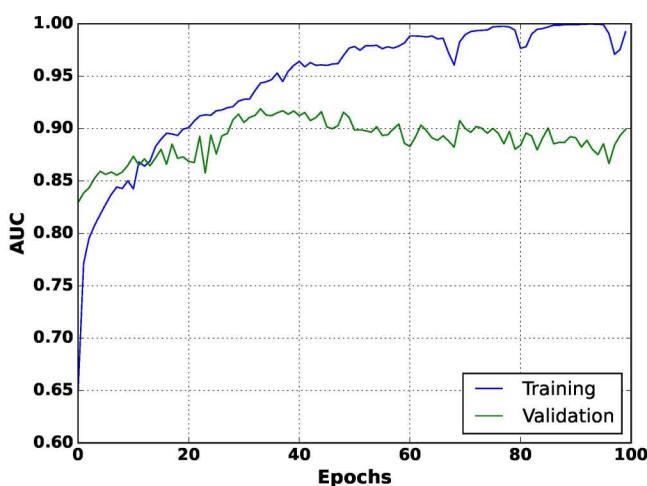


Fig. 4. Performance curves (AUC) of the proposed model reported for training and validation splits of the dataset.

atoms and their bonds from them. RDKit employs techniques to convert a given set of SMILES into a molecule object, that can subsequently be used to compute a wide range of molecular attributes and properties.

The experiments were performed using the Tensorflow framework. We have utilized the GPU acceleration of Tensorflow models. The employed dataset for the experiments is split into three-way sets of 80, 10, and 10 percent. They represent the training set, validation set, and testing set, respectively. The batch size, number of attention heads in transformer Readout are 32 and 8 while 512 neurons dense layer is used before the binary classification using single neuron dense layer with ‘sigmoid’ activation function. The proposed model is trained for 100 epochs, and performance is measured using the *AUC* metric.

V. RESULTS

The results of the experiments are shown in Figure 5. From the result we obtained, we found that our proposed model outperforms the other approaches such as Weave [44], Graph Convolution (GC) [45], Bypass multitask network (Bypass) [46], SingleTask [47], Influence relevance voting (IRV) [48], Random Forest (RF), XGBoost, Kernel Support vector classification (KernelSVM), and Logistic Regression (Logreg) for the node classification task on the blood-brain barrier dataset, and achieves 90.2 % AUC value on the testing dataset and 93 % in the validation split of the dataset. Comparing our proposed model with other approaches over the employed dataset, we observe that the present approaches perform well over the validation split of the dataset but suffer in the inference. The dip in the performance over testing split indicates the overfitting of the given approaches. Contrary to that, our approach performs equally well over both validation and testing splits of the dataset. The training process of the proposed model is shown in Figure 3. Here we observe that the model’s performance increases up to 40 epochs for the given dataset. After that, it stalls, showing the slight overfitting of the model. Further, dropout techniques can be utilized with the transformer readout to improve the generalization and reduce the overfitting of the model. Finally, we have shown inference of the model over a few molecules, as shown in Figure 4, for the brain-barrier permeability.

VI. CONCLUSION

We proposed a generalized neural network that operates over graph structure data in this work. Prediction of molecules’ permeability into the blood-brain barrier is tackled using the proposed model. It has real-world implications for designing new drugs and medicine for brain disorders. We examined the performance of our proposed model with various existing models such as Weave, Graph Convolution, Bypass multitask

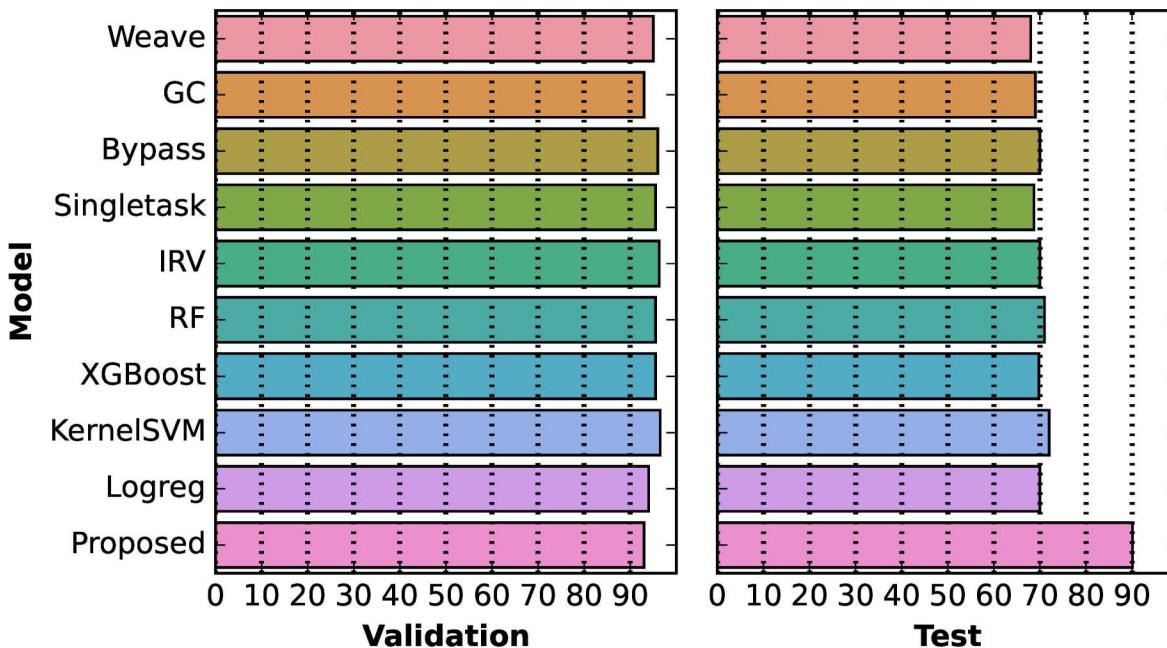


Fig. 5. Performance (AUC) of the proposed model versus other approaches over the training and validation splits of the dataset.

network, SingleTask, Influence relevance voting, Random Forest, XGBoost, Kernel Support vector classification, Logistic Regression. Our proposed model has shown high generalization and inference accuracy over other available approaches.

REFERENCES

- [1] R. Yan, J. Tang, X. Liu, D. Shan, and X. Li, "Citation count prediction: learning to estimate future citations for literature," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 1247–1252.
- [2] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American society for information science and technology*, vol. 58, no. 7, pp. 1019–1031, 2007.
- [3] M. Zhang and Y. Chen, "Link prediction based on graph neural networks," *Advances in Neural Information Processing Systems*, vol. 31, pp. 5165–5175, 2018.
- [4] P. G. Kulkarni, S. Praneet, R. Raghav, and B. Das, "Deep detection of anomalies in static attributed graph," in *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*. Springer, 2020, pp. 627–640.
- [5] T. Oyetunde, M. Zhang, Y. Chen, Y. Tang, and C. Lo, "Boostgapfill: improving the fidelity of metabolic network reconstructions through integrated constraint and pattern-based methods," *Bioinformatics*, vol. 33, no. 4, pp. 608–611, 2017.
- [6] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social networks*, vol. 25, no. 3, pp. 211–230, 2003.
- [7] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 11–33, 2015.
- [8] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [9] Y. I. Wang, H. E. Abaci, and M. L. Shuler, "Microfluidic blood-brain barrier model provides in vivo-like barrier properties for drug permeability screening," *Biotechnology and bioengineering*, vol. 114, no. 1, pp. 184–194, 2017.
- [10] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.
- [11] V. A. Levin, "Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability," *Journal of medicinal chemistry*, vol. 23, no. 6, pp. 682–684, 1980.
- [12] R. C. Young, R. C. Mitchell, T. H. Brown, C. R. Ganellin, R. Griffiths, M. Jones, K. K. Rana, D. Saunders, and I. R. Smith, "Development of a new physicochemical model for brain penetration and its application to the design of centrally acting h2 receptor histamine antagonists," *Journal of medicinal chemistry*, vol. 31, no. 3, pp. 656–671, 1988.
- [13] L. Di, H. Rong, and B. Feng, "Demystifying brain penetration in central nervous system drug discovery: Miniperspective," *Journal of medicinal chemistry*, vol. 56, no. 1, pp. 2–12, 2013.
- [14] J. F. Morales, S. Scioli Montoto, P. Fagiolino, and M. E. Ruiz, "Current state and future perspectives in qsar models to predict blood-brain barrier penetration in central nervous system drug r&d," *Mini reviews in medicinal chemistry*, vol. 17, no. 3, pp. 247–257, 2017.
- [15] H. Liu, K. Dong, W. Zhang, S. G. Summerfield, and G. C. Terstappen, "Prediction of brain: blood unbound concentration ratios in cns drug discovery employing in silico and in vitro model systems," *Drug discovery today*, vol. 23, no. 7, pp. 1357–1372, 2018.
- [16] O. A. Raevsky, V. Y. Grigorev, D. E. Polianczyk, O. E. Raevskaja, and J. C. Dearden, "Contribution assessment of multiparameter optimization descriptors in cns penetration," *SAR and QSAR in Environmental Research*, vol. 29, no. 10, pp. 785–800, 2018.
- [17] M. Gupta, H. J. Lee, C. J. Barden, and D. F. Weaver, "The blood-brain barrier (bbb) score," *Journal of medicinal chemistry*, vol. 62, no. 21, pp. 9824–9836, 2019.
- [18] P. Li, J. Wang, Y. Qiao, H. Chen, Y. Yu, X. Yao, P. Gao, G. Xie, and S. Song, "Learn molecular representations from large-scale unlabeled molecules for drug discovery," *arXiv preprint arXiv:2012.11175*, 2020.
- [19] ——, "An effective self-supervised framework for learning expressive molecular global representations to drug discovery," *Briefings in Bioinformatics*, 2021.
- [20] O. Wieder, S. Kohlbacher, M. Kuenemann, A. Garon, P. Ducrot, T. Seidel, and T. Langer, "A compact review of molecular property prediction with graph neural networks," *Drug Discovery Today: Technologies*, 2020.
- [21] A. M. Fout, "Protein interface prediction using graph convolutional networks," Ph.D. dissertation, Colorado State University, 2017.

- [22] D. E. Pires, T. L. Blundell, and D. B. Ascher, “pkcsm: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures,” *Journal of medicinal chemistry*, vol. 58, no. 9, pp. 4066–4072, 2015.
- [23] G. W. Bemis and M. A. Murcko, “The properties of known drugs. 1. molecular frameworks,” *Journal of medicinal chemistry*, vol. 39, no. 15, pp. 2887–2893, 1996.
- [24] D. Weininger, A. Weininger, and J. L. Weininger, “Smiles. 2. algorithm for generation of unique smiles notation,” *Journal of chemical information and computer sciences*, vol. 29, no. 2, pp. 97–101, 1989.
- [25] S. Jastrzebski, D. Lesniak, and W. M. Czarnecki, “Learning to smile (s),” *arXiv preprint arXiv:1602.06289*, 2016.
- [26] Z. Xu, S. Wang, F. Zhu, and J. Huang, “Seq2seq fingerprint: An unsupervised deep molecular embedding for drug discovery,” in *Proceedings of the 8th ACM international conference on bioinformatics, computational biology, and health informatics*, 2017, pp. 285–294.
- [27] C. Lu, Q. Liu, C. Wang, Z. Huang, P. Lin, and L. He, “Molecular property prediction: A multilevel quantum interactions modeling perspective,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 1052–1060.
- [28] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea *et al.*, “Analyzing learned molecular representations for property prediction,” *Journal of chemical information and modeling*, vol. 59, no. 8, pp. 3370–3388, 2019.
- [29] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, “Convolutional networks on graphs for learning molecular fingerprints,” *arXiv preprint arXiv:1509.09292*, 2015.
- [30] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “Line: Large-scale information network embedding,” in *Proceedings of the 24th international conference on world wide web*. Republic and Canton of Geneva: CHE: International World Wide Web Conferences Steering Committee, 2015, pp. 1067–1077.
- [31] J. Park, J. Cho, H. J. Chang, and J. Y. Choi, “Unsupervised hyperbolic representation learning via message passing auto-encoders,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 5516–5526.
- [32] T. N. Kipf and M. Welling, “Variational graph auto-encoders,” 2016.
- [33] J. Park, M. Lee, H. J. Chang, K. Lee, and J. Y. Choi, “Symmetric graph convolutional autoencoder for unsupervised graph representation learning,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6519–6528.
- [34] G. Cui, J. Zhou, C. Yang, and Z. Liu, “Adaptive graph encoder for attributed graph embedding,” in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD ’20. NY, USA: Association for Computing Machinery, 2020, p. 976–985.
- [35] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, and C. Zhang, “Adversarially regularized graph autoencoder for graph embedding,” in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 7 2018, pp. 2609–2615.
- [36] Z. Hu, C. Fan, T. Chen, K.-W. Chang, and Y. Sun, “Pre-training graph neural networks for generic structural feature extraction,” 2019.
- [37] Z. Peng, Y. Dong, M. Luo, X.-M. Wu, and Q. Zheng, “Self-supervised graph representation learning via global context prediction,” 2020.
- [38] Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, and J. Huang, “Self-supervised graph transformer on large-scale molecular data,” 2020.
- [39] J. Cao, X. Lin, S. Guo, L. Liu, T. Liu, and B. Wang, “Bipartite graph embedding via mutual information maximization,” in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, ser. WSDM ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 635–643.
- [40] H. Ma, Y. Bian, Y. Rong, W. Huang, T. Xu, W. Xie, G. Ye, and J. Huang, “Multi-view graph neural networks for molecular property prediction,” *arXiv preprint arXiv:2005.13607*, 2020.
- [41] Y. Song, S. Zheng, Z. Niu, Z.-H. Fu, Y. Lu, and Y. Yang, “Communicative representation learning on attributed molecular graphs,” in *IJCAI*, vol. 2020, 2020, pp. 2831–2838.
- [42] T. Fu, K. Huang, C. Xiao, L. M. Glass, and J. Sun, “Hint: Hierarchical interaction network for clinical-trial-outcome predictions,” *Patterns*, p. 100445, 2022.
- [43] “Bbbp benchmark (drug discovery) — papers with code,” <https://paperswithcode.com/sota/drug-discovery-on-bbbp>, (Accessed on 10/09/2021).
- [44] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, “Molecular graph convolutions: moving beyond fingerprints,” *Journal of computer-aided molecular design*, vol. 30, no. 8, pp. 595–608, 2016.
- [45] H. Altae-Tran, B. Ramsundar, A. S. Pappu, and V. Pande, “Low data drug discovery with one-shot learning,” *ACS central science*, vol. 3, no. 4, pp. 283–293, 2017.
- [46] B. Ramsundar, B. Liu, Z. Wu, A. Verras, M. Tudor, R. P. Sheridan, and V. Pande, “Is multitask deep learning practical for pharma?” *Journal of chemical information and modeling*, vol. 57, no. 8, pp. 2068–2076, 2017.
- [47] B. Ramsundar, S. Kearnes, P. Riley, D. Webster, D. Konerding, and V. Pande, “Massively multitask networks for drug discovery,” *arXiv preprint arXiv:1502.02072*, 2015.
- [48] S. J. Swamidass, C.-A. Azencott, T.-W. Lin, H. Gramajo, S.-C. Tsai, and P. Baldi, “Influence relevance voting: an accurate and interpretable virtual high throughput screening method,” *Journal of chemical information and modeling*, vol. 49, no. 4, pp. 756–766, 2009.