Communication Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 105–110 DOI: 10.15439/2024F3810 ISSN 2300-5963 ACSIS, Vol. 41

Toxic Molecule Classification Using Graph Neural Networks and Few Shot Learning.

Bhavya Mehta*, Kush Kothari†, Reshmika Nambiar‡ and Seema Shrawne¶
Department of Computer Engineering and Information Technology, Veermata Jijabai Technology Institute
Mumbai, India.
Email: *bdmehta_b19@ce.vjti.ac.in, †kmkothari_b19@ce.vjti.ac.in, ‡rsnambiar_b19@ce.vjti.ac.in, ¶scshrawne@ce.vjti.ac.in

***Abstract*—Traditional methods like Graph Convolutional Networks (GCNs) face challenges with limited data and class imbalance, leading to suboptimal performance in graph classification tasks during toxicity prediction of molecules as a whole. To address these issues, we harness the power of Graph Isomorphic Networks, Multi Headed Attention and Free Large-scale Adversarial Augmentation separately on Graphs for precisely capturing the structural data of molecules and their toxicological properties. Additionally, we incorporate Few-Shot Learning to improve the model's generalization with limited annotated samples. Extensive experiments on a diverse toxicology dataset demonstrate that our method achieves an impressive state-of-art AUC-ROC value of 0.816, surpassing the baseline GCN model by 11.4%. This highlights the significance of our proposed methodology and Few-Shot Learning in advancing Toxic Molecular Classification, with the potential to enhance drug discovery and environmental risk assessment processes.**

***Index Terms*—Graph Neural Networks, Graph Isomorphic Network, Multi Headed Attention, Graph Data Augmentation, Few Shot Learning, Toxicity Prediction.**

I. Introduction

Toxicological assessment of molecular compounds plays a pivotal role in drug discovery, environmental risk assessment, and chemical safety evaluation. Accurate prediction of a molecule's toxicity is crucial in ensuring the development of safe and effective drugs while minimizing potential harm to both human health and the environment.

Traditional methods of toxic molecule detection [[1],[2]] possess some inherent limitations. This is because conducting experiments to synthesize a compound and then analyzing its toxicity is time-consuming and often very expensive. It consumes a lot of resources and is not feasible for large-scale testing of molecules.

A number of approaches based on machine learning have also recently been proposed. The methods described above use several molecular characteristics, such as their physical and chemical properties, to predict their toxicity. However, a present problem in the field is lack of sufficient labelled data, due to the difficulties faced in synthesizing and testing new molecules, as explained above. Moreover, often these machine learning techniques only look at certain numerical properties of the molecules and fail to take into consideration the structural aspects of the molecule.

In recent literature, a lot of research is being done in representing molecules as graphs and processing them through Graph Neural Networks (GNNs). While this method does not tackle the low-data scenario we often face in toxicity prediction, newer methods have integrated few-shot learning into GNNs, like the Adaptive Step Model-Agnostic Meta-Learner (AS-MAML). We believe that this intersection of graph-embedding algorithms and few-shot learning is key to creating effective models for molecular toxicity prediction.

The research problem addressed in this paper is to investigate and propose enhancements to the GNN-specific few-shot learning technique in order to achieve favorable results in the toxicity prediction task on the Tox21 data set under the few shot learning scenario.

II. Background

Before delving into the specifics of the architecture, it is essential to provide some background information that will be helpful for better understanding.

A. Few Shot Learning (FSL)

As suggested by Vinyals et al. in [3], FSL is the ability of an algorithm to generalize well from limited data points with supervised information available for every class. To achieve this, we employ Model-Agnostic Meta-Learning (MAML) given by Finn et al. which aims to find a good initialization for the model parameters θ, for rapid adaptation to novel classes with only a few labeled examples. This is done by optimizing the model's performance on a set of meta-training experiments, where each task simulates a few-shot learning scenario.

B. Adaptive Step Model Agnostic Meta Learning (AS-MAML)

Introduced by Ma et al. in [5], it is a meta-learning technique that builds upon the MAML[4] algorithm by introducing an Adaptation Controller that employs reinforcement learning techniques to determine the optimal step size and when to stop the adaptation process. A StopController model, incorporating LSTM [6] layers and a sigmoid function, estimates the probability of stopping the adaptation process based on the training loss and embedding quality. This addresses the challenge of finding the optimal learning rate and step size in MAML-based meta-learning approaches.

C. Graph Convolution Network (GCN)

GCNs were introduced as a way to extend convolutional neural networks (CNNs) to handle irregular and non-Euclidean data. Kipf and Welling in [7] mentions that the key challenge

in processing graph data is that the number of nodes and their connectivity can vary widely from one graph to another. GCNs address this challenge by learning to exploit the local neighborhood information of each node in the graph to make predictions. The core idea behind GCNs is to perform node feature aggregation through a series of graph convolutions, enabling nodes to gather information from their neighbors and incorporate it into their own representations.

D. Graph Isomorphic Network (GIN)

Graph Isomorphic Networks by Xu et al. in [8] are a class of deep learning models designed for graph classification tasks. Unlike traditional GCNs, GINs do not rely on graph structure during message passing, making them more flexible and suitable for various graph types. The core idea behind GINs is to employ an aggregation function that is permutation-invariant to the node ordering, ensuring that the model produces the same output regardless of how the nodes are arranged. This property allows GINs to capture the global graph information effectively and provide more robust representations for graph classification tasks.

E. Free Large-scale Adversarial Augmentation on Graphs (FLAG)

It is a technique for enhancing graph data to imprive GNNs' performance. FLAG by Kong et al. in [9] suggests augmenting node properties rather than modifying graph topological structures, which is where the majority of existing graph regularizers concentrate their efforts. It improves generalization to out-of-distribution samples by iteratively enhancing node characteristics with gradient-based adversarial perturbations during training. This makes the model invariant to tiny fluctuations in input data. Adversarial data points are created and then inserted into the training data as part of the adversarial training process. The objective of this min-max optimization problem is to minimize the objective function while keeping the perturbation within a predetermined bound.

III. Related Works

Some of the earliest works in toxicity prediction include DeepTox by Mayr et al., who used chemical properties of these compounds fed into a Deep Neural Network to predict their toxicity. By using this method and ample of labelled data Mayr et al., manages to achieve an 0.92 AUC value. Alperstein et al. introduced All SMILES VAE [11], a generative model which uses variational autoencoders (VAEs) for generating SMILES strings using stacked RNNs. The model surpassed state-of-the-art methods and achieved an ROC-AUC score of 0.871 on the dataset. Censnet by [12] learns node and edge features through the use of novel propagation rules while switching the roles of nodes and edges. The method attains about 0.79 AUC score at most on the Tox21 dataset when tested under various splitting scenarios. Zhou et al. proposed Uni-Mol [13], a framework that incorporates the pretraining of transformers in order to use 3D information. It was evaluated on Tox21 as a downstream task and outperformed several methods, achieving an ROC-AUC score of 0.796.

Graph Multiset Transformer (GMT) [14] adopts a novel pooling method wherein multi-head attention is used for learning node interaction based on task relevance. An AUC score of about 0.773 was obtained on Tox21. Meta-MGNN by Guo et al. employs meta-learning to learn molecular representations under few-shot settings. It uses pretrained GNNs and leverages additional tasks to be optimised. When tested on Tox21, an AUC score of 0.769 was obtained under the one-shot setting and about 0.78 under the five-shot setting, outperforming several baseline models. However very few works have obtained significant results in few shot domain with graphs.Chen et al. in [16] achieves an average ROC-AUC score of 0.757 employing the Mean Teacher Semi-Supervised ML Algorithm, which is a 6% increase over GCN models trained using supervised and conventional ML techniques. However for low data scenarios, very few works have been able to get significant results.

IV. Tox21 Dataset

Tox21 is a dataset containing measurements of toxicity of 12 thousand molecules against 12 target proteins. It aims to help analyse the performance of models in predicting the biochemical activity of compounds using their chemical structure. We use the AhR sub-dataset from Tox21 that focuses on chemicals' interactions with this Aryl hydrocarbon Receptor, a ligand-activated transcription factor that is essential for the toxic response to toxins and medications. The dataset is open source and can be downloaded from Tox21 AHR[1].

Each chemical compound in this collection is represented as a graph, with atoms serving as nodes and chemical bonds between atoms serving as edges. Molecules' structural information is preserved in the graph representation, making it ideal for GNN-based approaches that can efficiently handle graph-structured data. By learning from the graph structure and associated node features, GNNs can discern complex relationships and identify key structural characteristics associated with toxic and non-toxic compounds.

V. Baseline Model

The initial configuration we are evaluating serves as the baseline, which is the standard GCN + AS-MAML model utilizing the few-shot learning setup detailed earlier. While we remain consistent with the framework described in the paper, there is one notable difference: we do not employ distinct classes for training and validation. This configuration consists of three successive layers: a GCN convolution layer, followed by a TopK Pooling Layer [17], each with a hidden layer dimension of 128. The Baseline Architecture is shown in Fig.1, has a validation accuracy of 65.02% and an AUC-ROC value of 0.732 on Tox21 AhR data.

[1]http://bioinf.jku.at/research/DeepTox/tox21.html

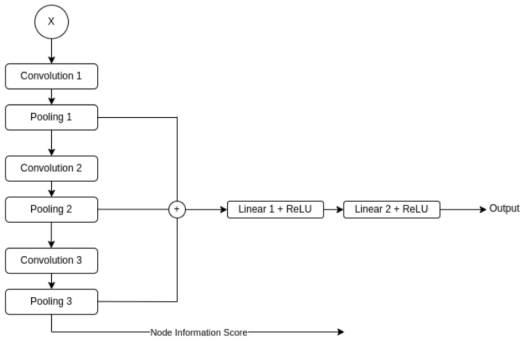

Fig. 1: The image illustrates the baseline sub-architecture.The resulting output vector is subsequently fed into a binary sigmoid classifier. The obtained Node Information Score is utilized by the FSL Reinforcement Learning Agent to optimize gradients and weights, thereby achieving a faster convergence rate.

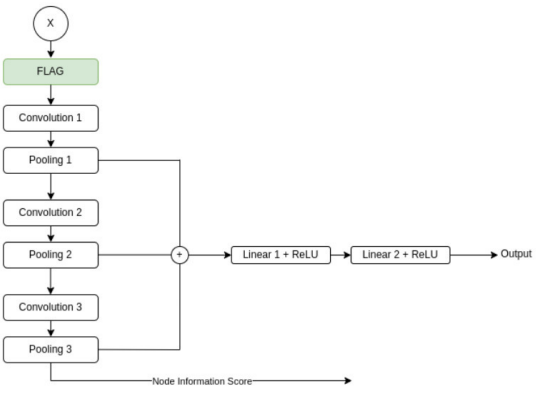

Fig. 2: Proposed sub-architecture of FLAG+GCN based classification model.

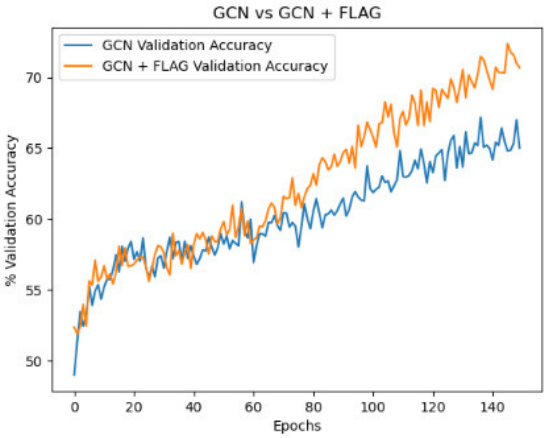

Fig. 3: A comparision of validation accuracy for GCN vs GCN+Flag method.

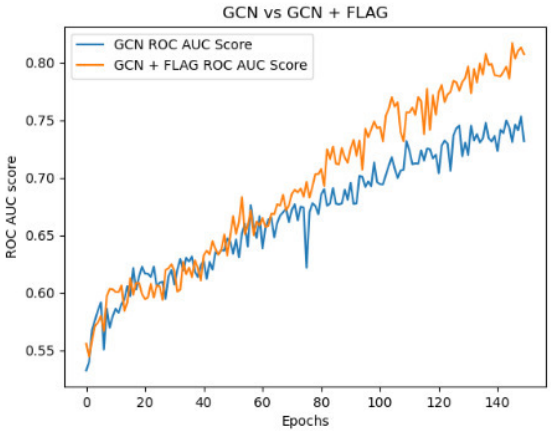

Fig. 4: A plot of ROC score for GCN vs the proposed GCN+FLAG sub-architecture.

VI. Proposed Architectures

In this research paper, we introduce and empirically evaluate three distinct architectural frameworks, each of which outperforms the baseline model in terms of achieved results. These three novel architectures systematically introduce variations across distinct components of the baseline model's structure the body, the input and the output, enhancing the model's capacity to capture intricate patterns and further enriching its learning capabilities.

A. Augmenting Input data using FLAG

1) Architecture: The first suggested setting adds a preprocessing step of FLAG in order to augment the data being fed into the model as shown in Fig.2. This adds perturbations to node features and provides greater variations in novel tasks available for few-shot learning. The aim of FLAG is to generate additional realistic graph instances that maintain the underlying distribution of the original data, effectively expanding the dataset and boosting model generalization.

2) Experimental Results: As shown in Fig.3 and Fig.4, we observe significant improvement in performance on the use of FLAG. This could be because it preserves the structural integrity of the graphs since only node features are modified. Tox21 is a molecular dataset where random changes in structure may not be realistic. Also, certain constraints are imposed on the perturbations, further improving reliability. FLAG has been found to be effective for discrete features which are commonly encountered in molecular data. In addition to this, FLAG improves generalization, robustness and data diversity, and is computationally efficient with validation accuracy of **70.68%** and validation AUC-ROC score of **0.806,** both of them greater than the baseline GCN model.

B. Replacing GCNs with GINs

1) Architecture: The proposed novel architecture (Refer Fig.5) introduces a modification to the AS-MAML algorithm[5] by replacing the three Graph Convolutional Network (GCN)[7] components with three Graph Isomorphism

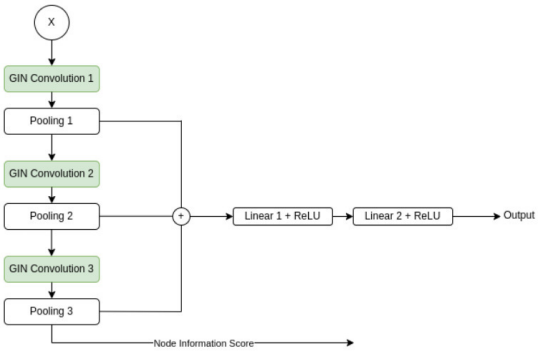

Fig. 5: Proposed sub-architecture of GIN based classification model.

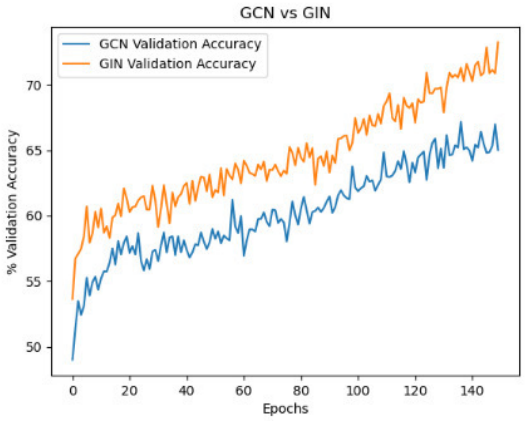

Fig. 6: A plot of validation accuracies of the baseline and the proposed GIN sub-architecture.

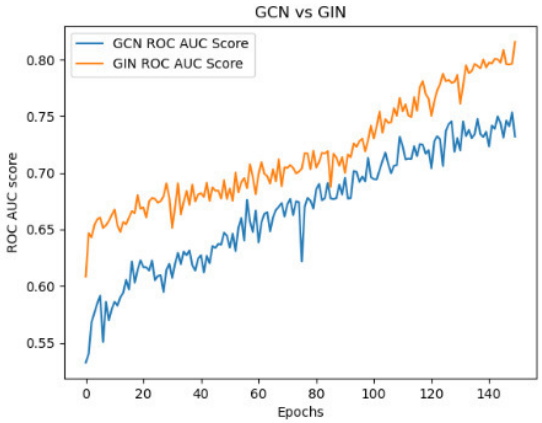

Fig. 7: A plot of ROC score for GCN vs the proposed GIN sub-architecture.

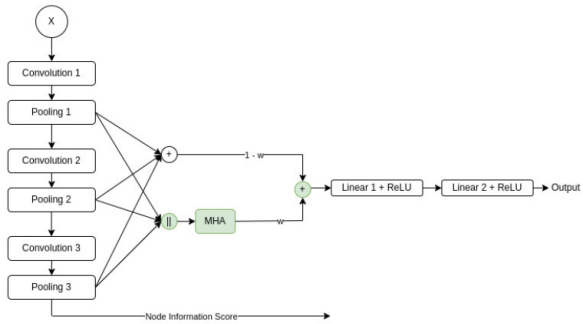

Fig. 8: Proposed sub-architecture of GCN+MHA based classification model.

Network (GIN) convolution operators. The GIN operator is anticipated to offer enhanced expressibility, resulting in improved hidden layer embeddings compared to the original GCN-based approach. In the forward() operation of the GIN model, a multi-layer perceptron (MLP) is employed, comprising four hidden layers, each containing ten perceptrons. The output layer, consistent with the GCN model, consists of 128 perceptrons.

2) Experimental Results: In the GIN model, we observed significant improvements in validation accuracy and ROC AUC score compared to the baseline GCN model, even at an early stage. The final validation accuracy of **73.23%** as shown in Fig.6 and ROC score of **0.816** as shown in Fig.7 can be attributed to the enhanced expressiveness of the Graph Isomorphism Network Operator utilized in GIN. Notably, this improved accuracy is consistently maintained over the course of 150 epochs, suggesting that while GIN may not necessarily provide an advantage in achieving higher-quality training results, it excels at capturing relevant task information with fewer epochs.

C. Enhanced Aggregation and Extraction using Weighted Multi Headed Attention (MHA).

1) Architecture: The final setting as shown in Fig.8, involves adding a Multi Attention Head (MAH) mechanism as an operator in the last part of the AS-MAML model. The baseline GCN carries out a normal aggregation of the outputs of the Relu layers and passes it to a binary classification network as displayed in the baseline figure above which might be unable to extract all necessary information or give weightage to the important ones. By inculcating a MAH layer, we attempt to change this fact and try to make the best out of the convolutions. MAH takes these three values as input to the Key, Value and Query fields to identify patterns of significance. The weight factor *"w"* adds an extra bias to the inclusion of the attention layer while performing regularization.

2) Experimental Reults: Upon examination of the graphs presented in Fig.9 and Fig.10, it becomes evident that the adapted model consistently outperforms the baseline counterpart during the validation phase. The GCN+MAH model achieves a notable validation accuracy of approximately **69.62%**, showcasing a significant improvement over the baseline's attainment of 65%. Additionally, a discernible discrepancy of 0.055 units is observed in the AUC-ROC values, further substantiating the effectiveness of the modified architecture. This enhanced performance of the GCN+MAH

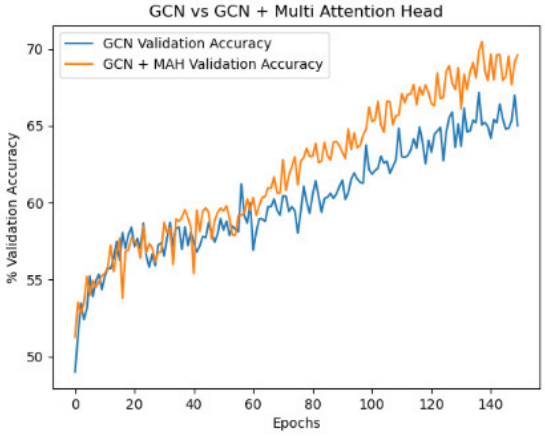

Fig. 9: A plot of validation accuracies of the baseline and the proposed GCN+MAH sub-architecture.

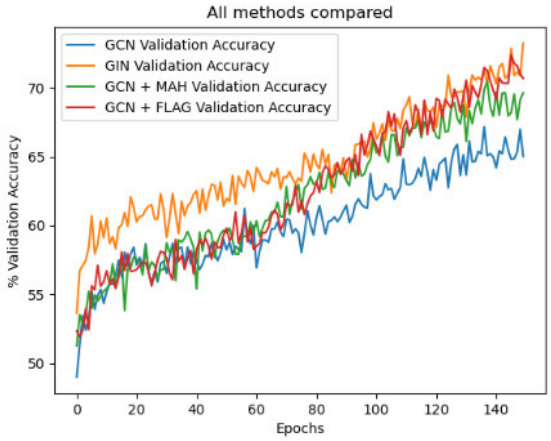

Fig. 11: Plot of validation accuracy for baseline and all three proposed sub-architectures.

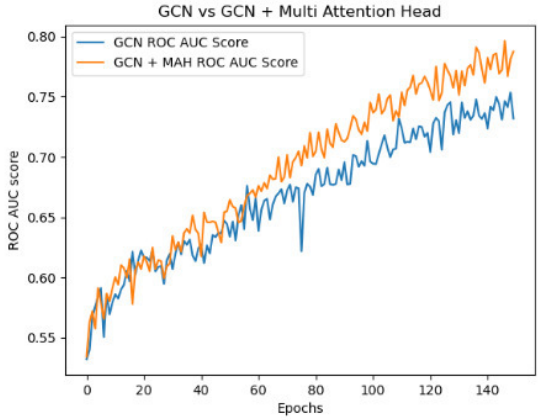

Fig. 10: A plot of ROC Score of the baseline and the proposed GCN+MAH sub-architecture.

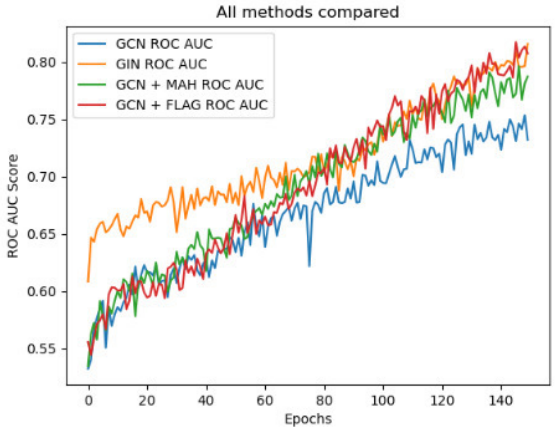

Fig. 12: Plot of ROC score for baseline and all three proposed sub-architectures.

model can be attributed to the model's abilty to simultaneously consider multiple attention patterns, enabling it to capture intricate data relationships and patterns more effectively. The experiments have been carried out with the weight factor "w" having a value of 0.4 which is another hyper parameter that we introduce.

VII. Discussion and Evaluations

Within this section, we elaborate on the training regimen and specifications pertinent to Few-Shot Learning (FSL), followed by a comprehensive evaluation of their collective outcomes along with an exhaustive assessment of the best method employed.

A. Few Shot Learning Specifications

In Table I, we list some of the common tunable parameters in a few shot learning scenarios and state the settings used for our testing.

B. Comprehensive Analysis

Table II and Fig.11, Fig.12, highlight the enhanced expressibility and feature representation ability of GIN operators, attention modules and data augmentation algorithms.

By adopting these modified architectures, which have the potential to advance the state-of-the-art results. We anticipate achieving improved performance and predictive capabilities in the context of the AS-MAML algorithm for few-shot learning tasks, particularly in toxicological classification of molecular compounds.

GIN's superiority is attributed to its order-agnostic aggregation operation, which ensures robustness and insensitivity to changes in node positions. In contrast, GCN's performance is influenced by the order of nodes in the neighborhood, making it more sensitive to node ordering. Another aspect contributing to GIN's efficacy is its higher expressiveness compared to GCN. While GCN focuses on local information within fixed neighborhoods, it faces limitations in capturing higher-order

TABLE I: Tunable Parameters and Values for FSL setting

Tunable Parameter	Significance	Value in Experiments
Train Shot	The number of labeled examples from the training set used for adapting the model during the few-shot learning process.	10
Validation Shot	The number of labeled examples from the validation set used for fine-tuning or evaluating the model during the few-shot learning process.	10
Train Query Set	The set of unlabeled examples from the training set that are used for prediction or evaluation after model adaptation.	15
Validation Query Set	The set of unlabeled examples from the validation set that are used for prediction or evaluation after fine-tuning or model evaluation.	15
Epochs	Count of how many times the full dataset was run through the model during training.	150
Learning Rate (Outer Loop)	A hyperparameter that determines the step size or rate at which the model's parameters are updated during the training process.	0.001
Learning Rate (Inner Loop)	A hyperparameter that determines the step size or rate at which the model's parameters are updated during the inner loop training process.	0.01

TABLE II: The table presents a detailed comparison between the accuracy and ROC values of the proposed methods.

Model and Algorithm	Validation Accuracy	Δ Accuracy Score	ROC-AUC Score	Δ ROC-AUC Score
GCN (Baseline)	65.02 %	-	0.732	-
GCN + FLAG	70.68 %	**+5.66 %**	0.806	**+0.074**
GIN	73.23 %	**+8.21 %**	0.816	**+0.084**
GCN + MHA	69.62 %	**+4.6 %**	0.787	**+0.055**

graph structures. Conversely, GIN's iterative message passing mechanism enables it to encompass more intricate and global structural patterns, making it more adept at handling complex molecular graphs.

Moreover, GIN's passing of the Weisfeiler-Lehman (WL) test [18], a theoretical measure of a GNN's expressive power, further validates its strength as a graph neural network. The WL test checks whether a GNN can differentiate non-isomorphic graphs with the same initial node labels. GIN's successful performance on this test showcases its stronger representational capacity compared to GCN and other methods.

Nevertheless, it's essential to acknowledge that the efficacy of GNN architectures may vary based on the specific dataset and task, with hyperparameter tuning and data preprocessing also impacting their overall performance.

VIII. Conclusion

In conclusion, the GIN method presented in this research paper establishes a new benchmark in drug discovery and toxicity prediction using the Tox21 data, exhibiting remarkable improvements of 8.21% in accuracy and 11.4% in ROC performance compared to existing GCN methods. Moreover, its exceptional performance in low labeled data scenarios, surpassing all other given methods, underscores its robustness and practicality. This novel approach holds immense promise for researchers and practitioners in pharmaceutical and chemical industries, providing valuable insights and advancements in these fields while inspiring further exploration and adoption of graph neural network-based methodologies for addressing real-world challenges

References

[1] Y. Jiang, D.-W. Sun, H. Pu, and Q. Wei, "Surface enhanced raman spectroscopy (sers): A novel reliable technique for rapid detection of common harmful chemical residues," *Trends in Food Science & Technology*, vol. 75, pp. 10–22, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924224417307045

[2] M. E. E. Alahi and S. C. Mukhopadhyay, "Detection methodologies for pathogen and toxins: A review," *Sensors*, vol. 17, no. 8, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/8/1885

[3] O. Vinyals, C. Blundell, T. P. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," *CoRR*, vol. abs/1606.04080, 2016. [Online]. Available: http://arxiv.org/abs/1606.04080

[4] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *CoRR*, vol. abs/1703.03400, 2017. [Online]. Available: http://arxiv.org/abs/1703.03400

[5] N. Ma, J. Bu, J. Yang, Z. Zhang, C. Yao, Z. Yu, S. Zhou, and X. Yan, "Adaptive-step graph meta-learner for few-shot graph classification," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, ser. CIKM '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1055–1064. [Online]. Available: https://doi.org/10.1145/3340531.3411951

[6] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM - a tutorial into long short-term memory recurrent neural networks," *CoRR*, vol. abs/1909.09586, 2019. [Online]. Available: http://arxiv.org/abs/1909.09586

[7] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2017. [Online]. Available: https://openreview.net/forum?id=SJU4ayYgl

[8] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *CoRR*, vol. abs/1810.00826, 2018. [Online]. Available: http://arxiv.org/abs/1810.00826

[9] K. Kong, G. Li, M. Ding, Z. Wu, C. Zhu, B. Ghanem, G. Taylor, and T. Goldstein, "Robust optimization as data augmentation for large-scale graphs," 2022.

[10] A. Mayr, G. Klambauer, T. Unterthiner, and S. Hochreiter, "DeepTox: Toxicity prediction using deep learning," *Frontiers in Environmental Science*, vol. 3, Feb. 2016. [Online]. Available: https://doi.org/10.3389/fenvs.2015.00080

[11] Z. Alperstein, A. Cherkasov, and J. T. Rolfe, "All SMILES VAE," *CoRR*, vol. abs/1905.13343, 2019. [Online]. Available: http://arxiv.org/abs/1905.13343

[12] X. Jiang, P. Ji, and S. Li, "Censnet: Convolution with edge-node switching in graph neural networks," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 7 2019, pp. 2656–2662. [Online]. Available: https://doi.org/10.24963/ijcai.2019/369

[13] G. Zhou, Z. Gao, Q. Ding, H. Zheng, H. Xu, Z. Wei, L. Zhang, and G. Ke, "Uni-mol: A universal 3d molecular representation learning framework," May 2022. [Online]. Available: https://doi.org/10.26434/chemrxiv-2022-jjm0j

[14] J. Baek, M. Kang, and S. J. Hwang, "Accurate learning of graph representations with graph multiset pooling," *CoRR*, vol. abs/2102.11533, 2021. [Online]. Available: https://arxiv.org/abs/2102.11533

[15] Z. Guo, C. Zhang, W. Yu, J. Herr, O. Wiest, M. Jiang, and N. V. Chawla, "Few-shot graph learning for molecular property prediction," *CoRR*, vol. abs/2102.07916, 2021. [Online]. Available: https://arxiv.org/abs/2102.07916

[16] J. Chen, Y.-W. Si, C.-W. Un, and S. W. I. Siu, "Chemical toxicity prediction based on semi-supervised learning and graph convolutional neural network," *Journal of Cheminformatics*, vol. 13, no. 1, Nov. 2021. [Online]. Available: https://doi.org/10.1186/s13321-021-00570-8

[17] C. Özdemir, "Avg-topk: A new pooling method for convolutional neural networks," *Expert Systems with Applications*, vol. 223, p. 119892, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423003937

[18] N. T. Huang and S. Villar, "A short tutorial on the weisfeiler-lehman test and its variants," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, jun 2021. [Online]. Available: https://doi.org/10.1109%2Ficassp39728.2021.9413523