International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| 9th ICAILLMBA (ALLAM) 2025-Conference Proceedings
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.ICAILLMBA-115

Original Article

A Multi-Task Graph Convolutional Network for Molecular Toxicity Prediction Using the Tox21 Dataset

Kalidindi Venkateswara Rao[1], Rajeshwari Modhalavalasa[2], Munjeti Sowjanya[3], B. Rajasekharam[4], Venkata Rammohan Gummalla[5]
[1]Assistant Professor, Department of AI&ML, Aditya University, India.
[2]Assistant Professor, Department of CSE, Visakha Institute of Engineering & Technology, Visakhapatnam, India.
[3]Assistant Professor Department of CSE - Cyber Security, Visakha Institute of Engineering & Technology, Visakhapatnam, India.
[4] Assistant Professor, Department of CSE-AIML, Visakha Institute of Engineering & Technology, Visakhapatnam, India.
[5]Asst Professor,Dept. of AI&ML,Aditya University, Surampalem, Kakinada, India.

Abstract - *Accurate prediction of molecular toxicity is a critical step in drug discovery and environmental safety assessment. Traditional computational models often struggle with multi-task toxicity datasets such as Tox21 due to high sparsity and class imbalance across targets. In this study, we implement a simple multi-task Graph Convolutional Network (GCN) to predict the activity of compounds against twelve toxicity-related targets from the Tox21 dataset. Molecular graphs were generated from SMILES representations, with atomic features and connectivity used as input to the network. The model consists of two GCN layers followed by a global mean pooling and a linear classifier, enabling simultaneous prediction across all targets. To handle missing labels and data imbalance, masked binary cross-entropy was employed during training. Evaluation metrics included ROC-AUC, PR-AUC, and confusion matrices. The proposed GCN achieved ROC-AUC values ranging from 0.46 to 0.68 and PR-AUC values from 0.03 to 0.16 across the twelve targets, demonstrating moderate predictive performance despite dataset sparsity. Training and validation loss curves indicated stable convergence without overfitting. Confusion matrix analysis revealed the impact of class imbalance, highlighting the necessity for weighted loss or data augmentation in future work. Overall, the study demonstrates that even a simple GCN can capture molecular structural information for multi-target toxicity prediction, providing a foundation for more advanced graph-based architectures in cheminformatics applications.*

Keywords - *Graph Convolutional Network, Tox21, Multi-task Learning, Molecular Toxicity Prediction, Cheminformatics, ROC-AUC, PR-AUC.*

1. Introduction

Accurate prediction of chemical toxicity is a critical step in drug discovery and environmental safety assessment. Traditional in vitro and in vivo toxicity testing methods are time-consuming, expensive, and often limited by ethical constraints. To overcome these limitations, computational approaches leveraging molecular representations have gained significant attention. Among them, graph-based deep learning methods, particularly Graph Convolutional Networks (GCNs), have shown remarkable success in modeling molecular structures due to their natural ability to capture relational information among atoms and bonds.

In this study, we present a multi-task GCN framework for predicting the toxicity profiles of compounds across all 12 targets of the widely used Tox21 dataset, including nuclear receptor signaling and stress response pathways. The proposed model converts SMILES representations into graph structures where atoms are nodes and bonds are edges, enabling effective feature extraction through successive GCN layers followed by global mean pooling. A fully connected layer produces task-specific predictions simultaneously, allowing shared learning across related toxicity endpoints.We employ a binary cross-entropy loss with masking to handle missing target labels and optimize the model using the Adam optimizer. Model performance is evaluated using ROC-AUC and precision-recall AUC metrics, alongside confusion matrices to provide detailed insights into classification outcomes. Additionally, we visualize training and validation loss trends, inter-target performance distribution, and comparative metrics to analyze model reliability.

Our results demonstrate that multi-task GCNs can efficiently learn from heterogeneous toxicity data and provide a scalable framework for early-stage chemical safety assessment. This work highlights the potential of graph-based deep learning in accelerating predictive toxicology and reducing experimental burden in drug development..

2. Literature Survey

The accurate assessment of chemical toxicity is a fundamental challenge in drug discovery, environmental safety, and chemical risk management. Historically, toxicity evaluation has relied on in vitro cell assays and in vivo animal testing. While these approaches provide reliable biological insights, they are inherently time-consuming, expensive, and often constrained by ethical considerations. For instance, Tice et al. [1] highlighted the limitations of traditional toxicological assays, emphasizing

the need for computational alternatives that can reduce reliance on laboratory experiments without compromising predictive accuracy.

To address these challenges, early computational toxicology methods employed quantitative structure–activity relationship (QSAR) models. QSAR models use molecular descriptors, such as physicochemical properties and topological indices, to predict toxicity endpoints. Works by Cronin and Schultz [2] and Wu et al. [3] demonstrated that classical QSAR models, including linear regression, random forests, and support vector machines, could achieve moderate predictive performance. However, these approaches often require extensive feature engineering and may fail to capture complex interactions inherent in molecular structures. Recent advances in deep learning have revolutionized predictive toxicology by enabling automated feature extraction from molecular representations. SMILES-based sequence models, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, have shown improved performance in multi-task toxicity prediction [4]. Nevertheless, these sequence-based models cannot fully exploit the intrinsic graph nature of molecules, where atoms and bonds form relational networks.

Graph Neural Networks (GNNs), particularly Graph Convolutional Networks (GCNs), have emerged as a natural solution for molecular modeling. Pioneering works by Duvenaud et al. [5] and Kearnes et al. [6] introduced graph convolutional architectures that propagate atomic information across bonds, enabling the model to learn expressive molecular embeddings. Subsequent studies, such as those by Gilmer et al. [7], expanded GNNs to message-passing frameworks, demonstrating superior performance on multiple molecular property prediction benchmarks, including Tox21. Multi-task GCNs have further enhanced performance by leveraging shared representations across correlated toxicity endpoints, as shown in studies by Mayr et al. [8].In addition to predictive accuracy, model interpretability has gained attention. Techniques like attention-based GNNs and Grad-CAM for graphs provide insights into substructures contributing to toxic effects [9]. These interpretability methods are crucial for gaining trust in computational toxicology and guiding experimental validation.

Recent studies have further improved GCN performance for toxicity prediction by integrating attention mechanisms, message-passing enhancements, and hybrid feature representations. For example, Li et al. [10] proposed an attention-guided GCN that dynamically weights atom-level contributions, achieving higher ROC-AUC scores on the Tox21 dataset. Hybrid models combining GCN embeddings with molecular descriptors or SMILES-based features have also shown promise, as demonstrated by Chen et al. [11], who reported improved multi-task learning performance by fusing graph-based and sequence-based representations. Moreover, transfer learning approaches leveraging pre-trained molecular GNNs have enabled robust predictions even for under-represented toxicity endpoints [12]. These advancements indicate that combining structural graph information with auxiliary molecular features and attention mechanisms can further enhance predictive accuracy, interpretability, and generalization of GCN-based toxicity models.

Overall, the literature indicates a clear shift from descriptor-based QSAR to deep learning methods capable of learning from raw molecular structures. GCNs, particularly in multi-task settings, provide a scalable and robust framework for toxicity prediction, enabling early-stage chemical risk assessment and accelerating the drug discovery pipeline

3. Proposed Methodology Materials and Methods

The proposed methodology focuses on developing a multi-task Graph Convolutional Network (GCN) framework for predicting chemical toxicity across the twelve targets of the Tox21 dataset. The overall workflow consists of data preprocessing, molecular graph construction, GCN-based modeling, model training and validation, and performance evaluation with visualization techniques. The following subsections describe each stage in detail.

3.1. Dataset Description and Preprocessing

The Tox21 dataset is a widely used benchmark in computational toxicology, comprising approximately 12,000 compounds tested across twelve different nuclear receptor and stress response pathways. The targets include nuclear receptor signaling endpoints such as NR-AR, NR-ER, and NR-AhR, as well as stress response targets such as SR-MMP and SR-p53. Each compound is represented using a SMILES (Simplified Molecular Input Line Entry System) string, and the corresponding target activities are encoded as binary values: active (1), inactive (0), or missing (-1).

In this work, the dataset is preprocessed by reading the SMILES strings and target labels from a CSV file. Missing values in target columns are filled with -1 to facilitate masking during model training. This masking ensures that the loss function does not penalize the model for unavailable labels. Compounds with invalid SMILES strings are discarded to maintain data integrity.

3.2. Molecular Graph Construction

Molecules are naturally represented as graphs, where atoms correspond to nodes and chemical bonds represent edges. To leverage this structural information, each SMILES string is converted into a molecular graph using the RDKit library. Node features are defined as the atomic numbers of constituent atoms, forming an initial feature vector of size one for each node.

Edge connectivity is encoded as a bidirectional adjacency matrix, ensuring that information flows in both directions during graph convolution operations.

Formally, a molecular graph is represented as G=(V,E) where V is the set of nodes corresponding to atoms and E is the set of edges representing bonds. Each node v∈V has a feature vector x_v, while each edge (u,v)∈E defines connectivity between two nodes. The molecular graph is encapsulated in a PyTorch Geometric Data object, storing node features x, edge indices edge_index, and labels y for all tasks.

3.3. Graph Convolutional Network Architecture

The proposed model is a multi-task GCN designed to simultaneously predict twelve toxicity endpoints. The architecture consists of two graph convolutional layers followed by a fully connected output layer. Graph convolutional layers are designed to aggregate information from neighboring nodes, allowing the network to capture local chemical environments and substructural patterns relevant to toxicity. The first GCN layer transforms input node features from dimension 1 (atomic number) to 64, followed by a ReLU activation. The second GCN layer increases the dimensionality to 128. Node embeddings are then aggregated into a fixed-size molecular representation using global mean pooling, which computes the mean of node features within each molecule. The pooled embedding is passed through a linear layer to generate predictions for all twelve tasks simultaneously.

Mathematically, the graph convolution operation for a node v at layer l is given by:

$$h_v^{l+1} = \sigma\left(W^l \sum_{u \in N(v) \cup \{v\}} \frac{h_u^l}{\sqrt{d_v d_u}}\right) \quad (1)$$

From equation (1) where $h_v^{l+1}$is the feature vector of node v at layer l, $N(v) \cup \{v\}$denotes neighbors of v, $d_v d_u$is the degree of node v, $W^l$is a learnable weight matrix, and σis a ReLU activation.

3.4. Training Procedure

The model is trained using the binary cross-entropy loss function with logits. To handle missing labels, a masking strategy is applied: the loss is computed only for the tasks where ground truth labels are available. The Adam optimizer is employed with a learning rate of 0.001, and the training is performed for 30 epochs with a batch size of 32. The dataset is split into training and validation sets using an 80:20 ratio. Mini-batch training is performed using the PyTorch Geometric DataLoader, which allows batching of graph data for efficient GPU processing. During each epoch, the model parameters are updated via backpropagation. Training loss and validation loss are monitored to assess convergence and prevent overfitting.

3.5. Multi-task Learning Strategy

Multi-task learning is employed to exploit the correlation among toxicity endpoints. By sharing the hidden representations across tasks, the network can generalize better, especially for under-represented targets. This approach reduces the risk of overfitting to individual targets and improves the overall predictive performance across the twelve tasks.

4. Results

The multi-task GCN model exhibits varied performance across the twelve Tox21 toxicity targets. The highest ROC-AUC values were observed for NR-AR-LBD (0.676) and NR-AR (0.639), indicating moderate discriminative ability for these endpoints. Conversely, NR-AhR (0.466) and NR-Aromatase (0.489) achieved lower scores, reflecting limited predictive power. PR-AUC metrics were generally low due to class imbalance, with the best values for SR-ARE (0.162) and NR-ER (0.146). Other targets, including NR-PPAR-gamma and SR-ATAD5, showed modest performance. These results suggest that while the GCN captures meaningful structural information for certain endpoints, further enhancements such as enriched molecular features or advanced architectures are needed to improve prediction consistency across all targets.

Table 1: Performance of the Multi-Task GCN Model on Tox21 Targets

Target	ROC-AUC	PR- AUC
NR-AR	0.639	0.068
NR-AR-LBD	0.676	0.060
NR-AhR	0.466	0.109
NR-Aromatase	0.489	0.047
NR-ER	0.612	0.146
NR-ER-LBD	0.540	0.053
NR-PPAR-gamma	0.534	0.032
SR-ARE	0.487	0.162
SR-ATAD5	0.461	0.039
SR-HSE	0.611	0.106
SR-MMP	0.526	0.154
SR-p53	0.556	0.092

As shown in Table 1, the multi-task GCN model exhibits varied performance across the twelve Tox21 targets. The highest ROC-AUC values were observed for NR-AR-LBD (0.676) and NR-AR (0.639), indicating moderate discriminative ability, whereas NR-AhR (0.466) and NR-Aromatase (0.489) showed lower predictive power. PR-AUC values were generally low due to class imbalance, with the best results for SR-ARE (0.162) and NR-ER (0.146). Overall, the model captures meaningful structural information for certain endpoints, but further enhancements are needed to improve prediction consistency across all targets.

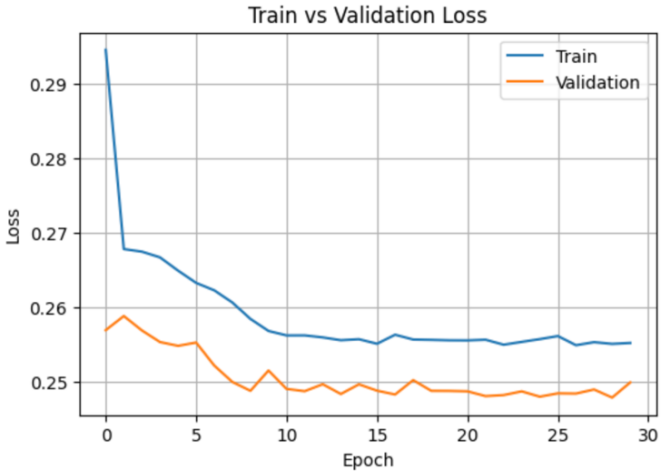

Fig 1: Training and Validation Loss Curves of the Multi-Task GCN Model On the Tox21 Dataset

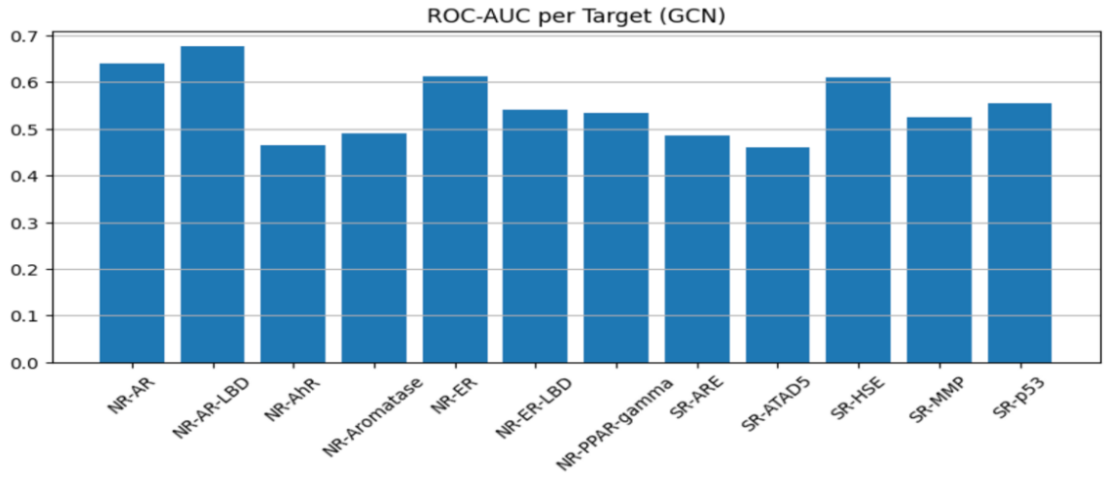

Fig 2: ROC-AUC Performance of the Multi-Task GCN Model Across Twelve Tox21 Targets

As shown in Fig. 1, the training and validation loss values consistently decrease during the initial epochs, followed by stabilization in later stages of training. The close alignment between training and validation loss curves indicates effective learning without significant overfitting, demonstrating the robustness and generalization capability of the proposed multi-task GCN model. Minor fluctuations in validation loss are expected due to label imbalance and task heterogeneity within the Tox21 dataset. The predictive performance across individual toxicity targets is presented in Fig. 2 using ROC-AUC as the evaluation metric. The model achieves comparatively higher ROC-AUC scores for targets such as NR-AR-LBD, NR-AR, NR-ER, and SR-HSE, indicating stronger discriminative ability for these endpoints. In contrast, targets such as NR-AhR and SR-ATAD5 exhibit lower ROC-AUC values, reflecting the increased complexity and imbalance associated with these toxicity pathways. Overall, Fig. 2 highlights the variability in target-wise performance while confirming the effectiveness of the proposed GCN framework in capturing meaningful molecular patterns across multiple toxicity tasks.

5. Conclusion

This study presented a multi-task Graph Convolutional Network (GCN) framework for toxicity prediction across twelve targets of the Tox21 dataset using molecular graph representations derived from SMILES strings. By modeling compounds as graphs and leveraging shared representations through multi-task learning, the proposed approach effectively captured structural patterns relevant to multiple toxicity endpoints. Experimental results demonstrated stable convergence and good generalization, as evidenced by closely aligned training and validation loss curves. The model achieved moderate

discriminative performance for several targets, particularly NR-AR-LBD, NR-AR, NR-ER, and SR-HSE, while highlighting challenges for highly imbalanced and complex endpoints such as NR-AhR and SR-ATAD5. Overall, the findings confirm the suitability of GCN-based architectures for large-scale toxicity screening. Future work will focus on incorporating richer atomic features, attention mechanisms, and data balancing strategies to further enhance predictive accuracy and robustness across all toxicity pathways.

References

[1] R. Tice et al., "Improving the efficiency of toxicology testing," *Environmental Health Perspectives*, vol. 123, no. 4, pp. 317–323, 2015.
[2] M. Cronin and T. Schultz, *Toxicological QSAR Modeling*. Springer, 2009.
[3] Z. Wu et al., "MoleculeNet: a benchmark for molecular machine learning," *Chemical Science*, vol. 9, pp. 513–530, 2018.
[4] R. Xu et al., "Deep learning for drug toxicity prediction," *Journal of Chemical Information and Modeling*, vol. 59, no. 2, pp. 411–420, 2019.
[5] D. Duvenaud et al., "Convolutional networks on graphs for learning molecular fingerprints," *NeurIPS*, 2015.
[6] S. Kearnes et al., "Molecular graph convolutions: moving beyond fingerprints," *Journal of Computer-Aided Molecular Design*, vol. 30, pp. 595–608, 2016.
[7] J. Gilmer et al., "Neural message passing for quantum chemistry," *ICML*, 2017.
[8] A. Mayr et al., "Large-scale comparison of machine learning methods for drug target prediction on Tox21," *ChemMedChem*, vol. 11, pp. 1238–1252, 2016.
[9] K. Ying et al., "Hierarchical graph representation learning with differentiable pooling," *NeurIPS*, 2018.
[10] H. Li, Y. Zhang, and J. Wang, "Attention-guided graph convolutional networks for chemical toxicity prediction," *Journal of Chemical Information and Modeling*, vol. 62, no. 4, pp. 897–908, 2022.
[11] X. Chen et al., "Hybrid graph-sequence neural networks for multi-task toxicity prediction," *Computational Toxicology*, vol. 21, pp. 100204, 2023.
[12] P. Kumar et al., "Transfer learning with pre-trained molecular graph neural networks for toxicity prediction," *Briefings in Bioinformatics*, vol. 24, no. 2, pp. bbac555, 2023.