# Applying Novel Graph Neural Networks for Multi-Task Prediction of Carcinogenicity

Author: Hrach Yeghiazaryan
Supervisor: Zaven Navoyan

***Abstract*—As the necessary tests for the preliminary assessment of drug carcinogenicity are time-consuming and expensive, the methods to predict their results using machine learning models become more and more popular. However, it is often the case that the dataset with labeled molecules is small. We try to address the problem by training the model on several tasks simultaneously using the one-primary multiple-auxiliaries approach which enables us to leverage other datasets besides the one for the primary task thus solving the problem of small datasets. We show that the results for in-vivo rat carcinogenicity improve when using carefully chosen auxiliary tasks.**

## I. Introduction

Toxicity assessment is an essential component of the drug development process, integral to ensuring that chemical compounds are safe for human use before they reach clinical trials. This process involves a comprehensive series of evaluations, both in vitro and in vivo. In vitro tests are conducted on a variety of biological substrates, including cell cultures, microbial cultures, and tissue samples. These tests are crucial for initial toxicity screenings and help in identifying any cytotoxic or microbiologically adverse effects of the compounds under study [1]. In vivo testing, conducted on animal models, further assesses the pharmacokinetics and pharmacodynamics of the drugs, providing a detailed understanding of their metabolism, distribution, and potential adverse effects on living organisms.

One of the primary goals of these assessments is to ensure patient safety. By identifying potentially harmful substances early in the drug discovery process, researchers can prevent these compounds from advancing to human trials [2]. This not only protects potential trial participants but also helps pharmaceutical companies avoid the financial and reputational costs associated with late-stage drug failures. Late-stage failures are particularly expensive due to the vast amounts of money already invested in the drug's development. Carcinogenicity tests are a major aspect of in vivo evaluations. These tests are designed to determine whether a substance could cause cancer in a living organism. Typically lasting two years, these tests are both time-intensive and costly, with expenses ranging from two to four million dollars per compound. They also require the use of several hundred animals, raising ethical and logistical concerns.

In recent years, there has been significant progress in the field of machine learning, which has had a transformative impact on the drug discovery process. Advancements in deep learning, particularly the development of algorithms based on graph theory, have revolutionized how drug screenings are conducted. These technological improvements allow for the prescreening of large libraries of compounds, efficiently identifying those that are likely to be toxic or carcinogenic. This method significantly reduces the number of compounds that need to proceed to more costly and ethically challenging in vivo testing [3]–[6].

Graph Neural Networks are at the forefront of this research. These networks are capable of predicting molecular behavior based on the structural information encoded within the graph representation of molecules [6]–[12].

Another promising area of research is the application of multitask training approaches. In multitask learning, a single model is trained on multiple tasks simultaneously [13], [14]. This approach can significantly enhance the model's predictive accuracy and efficiency, potentially leading to better outcomes in drug safety evaluations and a reduction in the reliance on animal testing.

As the field of toxicology continues to evolve, the integration of machine learning into traditional processes is proving to be invaluable. Not only does this integration enhance the efficiency and accuracy of toxicity assessments, but it also offers the potential to drastically reduce the costs and ethical concerns associated with traditional methods. The ongoing development and refinement of these technologies are crucial for the future of drug discovery and the broader field of biomedical research.

One of the in vivo carcinogenicity tests is checking whether a substance will cause cancer in rats. The rat carcinogenicity is the animal model that is the most frequently used one for preliminary assessment of human carcinogenicity. The purpose of the present work is to show that multi-task learning affects the results of predicting the rat carcinogenicity positively.

## II. Literature Review

Traditionally, the physicochemical or biological properties of molecules have been tested using various laboratory tests. With the rapid development of machine learning, specialized methods have been developed to assess whether a molecule possesses certain properties before performing laboratory tests, thus greatly reducing the costs of assessment [3]–[6]. Only molecules with favorable toxicity profiles may proceed to actual wet lab tests. One of the pioneering approaches for predicting molecule properties is to train a model on so-called molecule fingerprints. Molecule fingerprints are essentially vectors that describe the structure of the molecule with a multi-hot vector. They encode the presence of specific chemical features or patterns within a molecule. There is a variety of

fingerprints, such as ECFP, MACCS, PubChem fingerprints [15]–[17], etc. They differ by the structural properties they include in the feature vector and the way they calculated. After the molecular fingerprints are extracted, they can be used to train various traditional machine learning models and neural network architectures [18]–[20].

Among the various machine learning approaches for solving problems in toxicology and drug design, the methods working with graph data have become increasingly popular in recent years [6]–[12]. The reason for that is that molecules are naturally representable as graphs, where atoms are the vertices and bonds between them are edges. This so-called graph representation of molecules led to a demand for developing machine learning approaches tailored for working with graph data. Those approaches coalesce under the umbrella term Machine Learning on Graphs.

Given a set of molecules, it is a natural question to ask whether each of them possesses a certain property (e.g. is toxic). Formally put, we want to classify the molecules into the ones that have the property and the ones that do not. Now, given that we are able to represent each molecule as a graph, the problem of molecule classification reduces to the problem of graph classification which is covered by Machine Learning on Graphs [21].

Perhaps the most widely used algorithms in the context of graph classification are Graph Neural Networks (GNN) [22]–[27]. The idea behind GNNs is message passing between the neighboring nodes so that during forward propagation each node aggregates the messages with information coming from its neighborhood.

Additionally, in a variety of problems, it was shown that in order to improve the results of predictions for a task, it is beneficial to train the model to predict several tasks at once. Those other tasks included in the training process are assumed to be related with the task for which it is desired to make the results better [28]. There is a present work on using GNN architectures where the so-called multi-task approach is utilized to make the results of predicting molecule properties better [14].

## III. Datasets

In this section, we describe various tests that biologists believe to be related with rat carcinogenicity.

### A. Rat Carcinogenicity (Carcino_Rat)

During this test, a drug is given to several hundreds of rats, and then, those rats are observed over a long period of time to check whether the drug causes cancer in rats. As there is a correlation between cancer causation in humans and rats, this is a useful preliminary test to assess whether the drug will cause cancer in humans [29].

### B. H2AX Agonist (H2AX)

This is an assay to identify small molecule agonists of H2AX. This high-throughput assay allows to identification of compounds that cause DNA damage, particularly DNA double-strand breaks [30].

### C. Ames test (Ames)

The Ames test is a widely used bacterial assay designed to assess the mutagenic potential of chemical substances by observing their ability to induce mutations in specific genes of Salmonella bacteria. It serves as a rapid and cost-effective tool in evaluating the potential carcinogenicity of various compounds and is commonly employed in regulatory toxicology and drug development [31].

### D. Androgen Agonist (AR_Agonist)

An androgen agonist is a substance that mimics the action of male hormones like testosterone, activating androgen receptors in the body. It's used in medicine to treat conditions like hypogonadism and prostate cancer, and it's sometimes misused in sports doping for performance enhancement [32].

### E. Androgen Antagonist (AR_Antagonist)

Androgen antagonists are drugs that block the action of male hormones like testosterone, commonly used in prostate cancer therapy to inhibit tumor growth fueled by androgens [32].

### F. Androgen Binding (AR_Binding)

Androgen binding refers to the interaction with male hormones (androgens) like testosterone and receptors on cancer cells. This binding can stimulate cancer growth, particularly in androgen-sensitive cancers like prostate cancer [32].

### G. Chromosomal Abberation In Vivo (CA_In_Vivo)

Chromosomal aberrations in vivo refer to structural or numerical abnormalities in animal chromosomes within cancer cells, caused by mutations or other factors. These abberations play a significant role in cancer development and progression, contributing to tumor growth, metastasis, and drug resistance. Detecting and identifying these abberations are crucial for cancer diagnosis, prognosis, and treatment planning [33]–[35].

### H. Chromosomal Abberation In Vitro (CA_In_Vitro)

In cancer research conducted in vitro, chromosomal aberrations are studied outside of living organisms, typically in cell cultures. These aberrations, such as deletions, duplications, or translocations, mimic those found in vivo and are crucial for understanding cancer biology, drug development, and therapeutic interventions [33]–[35].

### I. Estrogen Agonist (ER_Agonist)

An estrogen agonist is a substance that activates estrogen receptors in the body, mimicking the effects of estrogen hormones. These agonists are used therapeutically in hormone replacement therapy for menopausal symptoms, osteoporosis, and certain hormone-sensitive cancers like breast cancer [36].

### J. Estrogen Antagonist (ER_Antagonist)

An estrogen antagonist is a substance that blocks or inhibits the action of estrogen hormones by binding to estrogen receptors without activating them. These antagonists are used in medicine to treat estrogen-sensitive conditions such as breast cancer and endometriosis, where reducing estrogen activity is beneficial [36].

### K. Estrogen Binding (ER_Binding)

Estrogen binding refers to the interaction between estrogen hormones and estrogen receptors in cells. When estrogen binds to its receptors, it triggers a cascade of cellular responses that regulate various physiological processes, including growth, development, and reproduction. Understanding estrogen binding is crucial in hormone-related conditions and diseases such as breast cancer and osteoporosis, where estrogen activity plays a significant role [36].

### L. Micronucleus In Vivo (MN_In_Vivo)

In vivo micronucleus assays are tests conducted in living organisms to assess genotoxicity by observing the presence of micronuclei in cells, typically in blood or bone marrow samples. Micronuclei are small, additional nuclei that can form when chromosomes or chromosome fragments are not properly segregated during cell division. These assays are important in toxicology and environmental health research for evaluating the potential mutagenic and carcinogenic effects of chemicals, pollutants, and other substances on whole organisms [37].

### M. Micronucleus In Vitro (MN_In_Vitro)

In vitro micronucleus assays are laboratory tests used to assess genotoxicity by observing the formation of micronuclei in cultured cells. These assays are valuable tools in toxicology and drug development for evaluating the potential mutagenic and carcinogenic effects of chemicals, pharmaceuticals, and environmental agents [38].

## IV. Methodology

A typical GNN architecture has two stages in its forward propagation procedure, first, neighborhood aggregation, and second, fully connected tower propagation.

### A. Neighborhood aggregation

The most classic approach for neighborhood aggregation is Graph Convolutional Network (GCN). The GCN algorithm is the following:

$$h_v^{(0)} = x_v \tag{1}$$

$$h_v^{l+1} = \sigma(W_l \sum_{u \in N(v)} \frac{h_u}{|N(v)|} + B_l h_v), \qquad l \in \{0, \ldots, L-1\} \tag{2}$$

$$z_v = h_v^{(L)} \tag{3}$$

where $x_v$ is the initial feature vector of node $v$, $N(v)$ is the one-hop neighborhood of node $v$, $W_l$ and $B_l$ for $l \in \{0, \ldots, L-1\}$ are learnable matrices, and $z_v$ is the final embedding of node $v$ after $L$ rounds of aggregations.

### B. Readout and MLP

After performing the neighborhood aggregation procedure, each node has an embedding encompassing the information coming from its neighborhood and its initial feature vector. In order to make graph-level predictions, that is, to predict a label for an entire graph, we have to somehow summarize embeddings of the nodes of that graph to obtain an embedding describing the entire graph. A simple approach to doing that is to sum or average embeddings of the nodes, or even perform max pooling on them. The function that summarizes the node embeddings into a graph embedding is called a readout function.

After the embedding for the entire graph is obtained it is propagated through an MLP that outputs the label. Therefore, the loss is defined with respect to the output of the MLP, and since the output depends on the weights of both, MLP and graph convolutions, during the training all those weights are optimized together.

### C. Proposed model

Du et. al. [14] enhanced the prediction quality of ADMET properties of molecules by training the model not only to predict labels for the desired task but also on several other tasks by leveraging the paradigm of "one primary, multiple auxiliaries". The approach suggests to train the model to predict multiple tasks by treating one of them as a primary. For this research, we adopted their approach to improve the predictions on rat carcinogenicity task, by taking rat carcinogenicity as a primary task and several other tasks that we know from biological knowledge are related to it as auxiliaries. We make use of the ResGCN model proposed in [14] to conduct the experiments (Figure 1).

In the neighborhood aggregation phase of this architecture, besides the classic GCN convolution presented above, there is also a residual connection added, in a way that for each node of a graph, the embedding of that node is passed through one linear layer and then added to the embedding obtained through GCN convolution. This is similar to the approach leveraged in ResNet [39]. In the phase of readout, the weighted mean of the node embeddings is computed. These weights are learnable, and are task-dependent, thus for each task, the node embeddings are aggregated with different weights. After the readout step, the graph embeddings for auxiliary tasks are ready, but for the primary task, we get its embedding that is obtained after the readout, and gate it with each of the auxiliary task embeddings separately, then we sum those gated feature vectors to obtain the final embedding for the primary task. Finally, each embedding propagates through the MLP corresponding to its task. The loss function for each of the tasks is BCE loss, which also balances the weights for classes for each task separately in order to address the issues associated with class imbalances.

## V. Experiments

We take as a baseline the model trained on Carcino_Rat single-task, meaning that the ResGCN model is fitted by taking

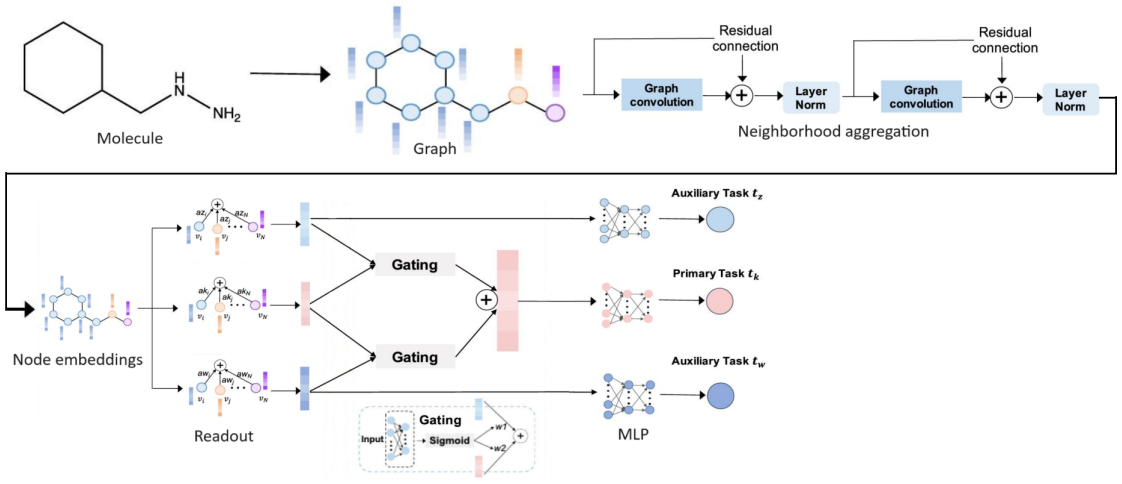


Fig. 1. ResGCN architecture

Carcino_Rat as a primary, and no tasks as auxiliaries. We want to understand, whether taking some of the tasks described in Section III as auxiliaries will improve the results of the model on Carcino_Rat. We compare models in the following way: we take their scores on multiple splits and then compare the mean AUC-s on Carcino_Rat of each model on those splits. Additionally, we check on how many splits one model superior to the other in predicting Carcino_Rat labels. We perform the Mann-Whitney U-test on the results of the obtained AUC scores for Carcino_Rat in order to get a statistical estimate. Wherever mentioned above, the AUC scores for Carcino_Rat are the validation scores. We decided to evaluate the models on validation sets because during the training we noticed that the AUC score fluctuates a lot depending on the split, and since the test set is just one set, we thought it would not accurately show which model was better.

### A. Auxiliary Selection

The first decision to make is which tasks should be chosen to participate as auxiliaries in order to improve the results on Carcino_Rat. Indeed, from the point of view of biology, all of them should be useful, but from the point of view of machine learning, it is not obvious whether including each of those would help or harm the prediction accuracy for Carcino_Rat. To perform the auxiliary selection we make 20 stratified random splits, and then, we train and evaluate the baseline model on those 20 splits. Then, for each of the auxiliary tasks, we take them exclusively as an auxiliary to Carcino_Rat as the primary task, then train, evaluate, and compare the results with the baseline. The tasks with which the accuracy of Carcino_Rat would be significantly greater than the baseline will be considered useful auxiliaries and will be included in the final training.

### B. Final Experiment

After selecting which tasks should participate in the training as auxiliaries to Carcino_Rat, we actually train the model with Carcino_Rat being the primary task and the selected tasks as auxiliary tasks. Additionally, in order to understand whether it makes sense to choose between the auxiliaries, we decided to train a model that involved all auxiliaries.

## VI. Results

### A. Results For Auxiliary Selection

We see that the p-values of the U-test are not significant anywhere (Table I), and we thought that the reason for that may be that we made too few splits for the test to be able to adequately compare the performances. Thus we decided to choose the auxiliaries based on how many validation splits the model with the auxiliary outperformed the baseline. This way, we chose Ames, AR_Agonist, CA_In_Vitro, and CA_In_Vivo as auxiliaries. Additionally, in order to have more precise comparisons for the final experiment, we made 50 splits and conducted the final experiments on them.

### B. Results For Training With Multiple Auxiliaries

We now make 50 splits and train the model with multiple auxiliaries(Table II).

We see, that when involving those selected auxiliaries in the training process, the mean AUC goes up by 2%, the model then beats the baseline on 38 splits out of 50, and finally, as we have enough splits, we perform a t-test and obtain a significant p-value. This indeed shows that multitask learning has the potential to improve the results for the problem of carcinogenicity prediction. As of the model trained on all tasks, we see, the results are worse compared to the model with selected auxiliaries.

| Auxiliary | Mean AUC | Num Better Than Single | U-stat | p-value |
|---|---|---|---|---|
| Single | 0.754 | | | |
| H2AX | 0.762 | 11 | 215.5 | 0.685 |
| Ames | 0.769 | **14** | 227.5 | 0.465 |
| AR_Agonist | 0.762 | **14** | 220 | 0.598 |
| AR_Antagonist | 0.760 | 12 | 211.5 | 0.766 |
| AR_Binding | 0.758 | 12 | 209 | 0.818 |
| CA_In_Vitro | 0.763 | **15** | 223.5 | 0.534 |
| CA_In_Vivo | 0.761 | **14** | 222 | 0.561 |
| ER_Agonist | 0.754 | 9 | 199 | 0.989 |
| ER_Antagonist | 0.747 | 8 | 164.5 | 0.344 |
| ER_Binding | 0.750 | 10 | 187 | 0.735 |
| MN_In_Vitro | 0.759 | 11 | 207.5 | 0.850 |
| MN_In_Vivo | 0.757 | 10 | 206 | 0.882 |

TABLE I
RESULTS FOR AUXILIARY SELECTION

| Auxiliary | Mean AUC | Num Better Than Single | t-stat | p-value |
|---|---|---|---|---|
| Single | 0.737 | | | |
| Selected Auxiliaries | 0.757 | 38 | 2.226 | 0.01 |
| All Auxiliaries | 0.752 | 33 | 1.59 | 0.06 |

TABLE II
RESULTS FOR FINAL EXPERIMENTS

## VII. CONCLUSION AND FUTURE WORK

To conclude, the experiments have shown that the multi-task learning approach is well-suitable with the graph neural networks and enables them to learn to predict rat carcinogenicity better than in the single-task setting. And so, we've shown that by using cheaper tests we can enhance the quality of predictions for a more expensive rat carcinogenicity test. For future work, it is planned to use more expressive graph convolutions, such as GIN and GINE, as well as to leverage a stricter approach for the auxiliary selection.

## REFERENCES


[1] D. Krewski, D. Acosta Jr, M. Andersen, H. Anderson, J. C. Bailar III, K. Boekelheide, R. Brent, G. Charnley, V. G. Cheung, S. Green Jr *et al.*, "Toxicity testing in the 21st century: a vision and a strategy," *Journal of Toxicology and Environmental Health, Part B*, vol. 13, no. 2-4, pp. 51–138, 2010.

[2] F. Pognan, M. Beilmann, H. Boonen, A. Czich, G. Dear, P. Hewitt, T. Mow, T. Oinonen, A. Roth, T. Steger-Hartmann *et al.*, "The evolving role of investigative toxicology in the pharmaceutical industry," *Nature reviews drug discovery*, vol. 22, no. 4, pp. 317–335, 2023.

[3] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer *et al.*, "Applications of machine learning in drug discovery and development," *Nature reviews Drug discovery*, vol. 18, no. 6, pp. 463–477, 2019.

[4] S. Dara, S. Dhamercherla, S. S. Jadav, C. M. Babu, and M. J. Ahsan, "Machine learning in drug discovery: a review," *Artificial Intelligence Review*, vol. 55, no. 3, pp. 1947–1999, 2022.

[5] L. Patel, T. Shukla, X. Huang, D. W. Ussery, and S. Wang, "Machine learning methods in drug discovery," *Molecules*, vol. 25, no. 22, p. 5277, 2020.

[6] 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.

[7] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," *Advances in neural information processing systems*, vol. 28, 2015.

[8] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International conference on machine learning*. PMLR, 2017, pp. 1263–1272.

[9] 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, pp. 595–608, 2016.

[10] A. Fout, J. Byrd, B. Shariat, and A. Ben-Hur, "Protein interface prediction using graph convolutional networks," *Advances in neural information processing systems*, vol. 30, 2017.

[11] K. Do, T. Tran, and S. Venkatesh, "Graph transformation policy network for chemical reaction prediction," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 750–760.

[12] N. Xu, P. Wang, L. Chen, J. Tao, and J. Zhao, "Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions," *arXiv preprint arXiv:1905.09558*, 2019.

[13] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[14] B.-X. Du, Y. Xu, S.-M. Yiu, H. Yu, and J.-Y. Shi, "Admet property prediction via multi-task graph learning under adaptive auxiliary task selection," *Iscience*, vol. 26, no. 11, 2023.

[15] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *Journal of chemical information and modeling*, vol. 50, no. 5, pp. 742–754, 2010.

[16] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse, "Reoptimization of mdl keys for use in drug discovery," *Journal of chemical information and computer sciences*, vol. 42, no. 6, pp. 1273–1280, 2002.

[17] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant, "Pubchem: a public information system for analyzing bioactivities of small molecules," *Nucleic acids research*, vol. 37, no. suppl_2, pp. W623–W633, 2009.

[18] C. Xu, F. Cheng, L. Chen, Z. Du, W. Li, G. Liu, P. W. Lee, and Y. Tang, "In silico prediction of chemical ames mutagenicity," *Journal of chemical information and modeling*, vol. 52, no. 11, pp. 2840–2847, 2012.

[19] M. Yang, B. Tao, C. Chen, W. Jia, S. Sun, T. Zhang, and X. Wang, "Machine learning models based on molecular fingerprints and an extreme gradient boosting method lead to the discovery of jak2 inhibitors," *Journal of Chemical Information and Modeling*, vol. 59, no. 12, pp. 5002–5012, 2019.

[20] L. Xie, L. Xu, R. Kong, S. Chang, and X. Xu, "Improvement of prediction performance with conjoint molecular fingerprint in deep learning," *Frontiers in pharmacology*, vol. 11, p. 606668, 2020.

[21] F. Errica, M. Podda, D. Bacciu, and A. Micheli, "A fair comparison of graph neural networks for graph classification," *arXiv preprint arXiv:1912.09893*, 2019.

[22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[23] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[24] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio *et al.*, "Graph attention networks," *stat*, vol. 1050, no. 20, pp. 10–48 550, 2017.
[25] P. Veličković, W. Fedus, W. L. Hamilton, P. Liò, Y. Bengio, and R. D. Hjelm, "Deep graph infomax," *arXiv preprint arXiv:1809.10341*, 2018.
[26] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.
[27] W. Hu, B. Liu, J. Gomes, M. Zitnik, P. Liang, V. Pande, and J. Leskovec, "Strategies for pre-training graph neural networks," *arXiv preprint arXiv:1905.12265*, 2019.
[28] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 12, pp. 5586–5609, 2021.
[29] N. C. for Biotechnology Information, "Pubchem bioassay record for aid 1208, dsstox (cpdbas) carcinogenic potency database summary rat bioassay results," *PubChem*, 2024.
[30] ——, "Pubchem bioassay record for aid 1224896, qhts assay to identify small molecule agonists of h2ax: Summary," *PubChem*, 2024.
[31] K. Hansen, S. Mika, T. Schroeter, A. Sutter, A. Ter Laak, T. Steger-Hartmann, N. Heinrich, and K.-R. Muller, "Benchmark data set for in silico prediction of ames mutagenicity," *Journal of chemical information and modeling*, vol. 49, no. 9, pp. 2077–2081, 2009.
[32] K. Mansouri, N. Kleinstreuer, A. M. Abdelaziz, D. Alberga, V. M. Alves, P. L. Andersson, C. H. Andrade, F. Bai, I. Balabin, D. Ballabio *et al.*, "Compara: collaborative modeling project for androgen receptor activity," *Environmental Health Perspectives*, vol. 128, no. 2, p. 027002, 2020.
[33] R. Corvi and F. Madia, "Eurl ecvam genotoxicity and carcinogenicity consolidated database of ames positive chemicals," *Eur. Comm. Jt. Res. Cent.[Dataset]*, 2018.
[34] F. Madia and R. Corvi, "Eurl ecvam genotoxicity and carcinogenicity consolidated database of ames negative chemicals," *Eur. Comm. Jt. Res. Cent.[Dataset]*, 2020.
[35] T. Morita, Y. Shigeta, T. Kawamura, Y. Fujita, H. Honda, and M. Honma, "In silico prediction of chromosome damage: comparison of three (Q)SAR models," *Mutagenesis*, vol. 34, no. 1, pp. 91–100, 07 2018. [Online]. Available: https://doi.org/10.1093/mutage/gey017
[36] K. Mansouri, A. Abdelaziz, A. Rybacka, A. Roncaglioni, A. Tropsha, A. Varnek, A. Zakharov, A. Worth, A. M. Richard, C. M. Grulke *et al.*, "Cerapp: collaborative estrogen receptor activity prediction project," *Environmental health perspectives*, vol. 124, no. 7, pp. 1023–1033, 2016.
[37] C. B. R. Benigni and C. L. Battistelli, "Chemical toxicity: Structures and experimental data," *ISSTOX Chemical Toxicity Databases*, 2021.
[38] D. Baderna, D. Gadaleta, E. Lostaglio, G. Selvestrel, G. Raitano, A. Golbamaki, A. Lombardo, and E. Benfenati, "New in silico models to predict in vitro micronucleus induction as marker of genotoxicity," *Journal of hazardous materials*, vol. 385, p. 121638, 2020.
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.