A Multitask Graph Neural Network Framework for Ames Mutagenicity Prediction

Abigail E. Teitgen[1,2], Eugenia Ulzurrun[2], Nuria E. Campillo[2], Eduardo R. Hernández[1]
[1]Instituto de Ciencia de Materiales de Madrid (CSIC), Calle de Sor Juana Inés de la Cruz, 3, 28049 Madrid, Spain
[2]Centro de Investigaciones Biológicas Margarita Salas (CSIC), Calle de Ramiro de Maeztu, 9, 28040, Madrid, Spain

abigail.teitgen@csic.es

The process of drug screening and development is long, complex, and costly, making early and reliable assessment of toxicity a central challenge in pharmaceutical research. Computational models, including Quantitative Structure–Activity Relationship (QSAR) approaches, have shown promise in reducing the need for costly experimental testing and have evolved substantially with advances in artificial intelligence [1,2]. Deep learning methods can be used to predict the outcome of the Ames mutagenicity test, which relies on multiple strains of *Salmonella typhimurium* to detect different mutagenic mechanisms [3]. Recent work has demonstrated that a deep multitask learning (MTL) approach that accounts for individual strain responses improves mutagenicity prediction [4]. However, most existing models rely on predefined molecular descriptors, which often lack both expressive power and interpretability [5].

Graph neural networks (GNNs) provide a chemically intuitive alternative by operating directly on molecular graphs and supporting the integration of structural information [5]. Here, we present a MTL-GNN framework for predicting Ames mutagenicity that explicitly models strain-specific responses while incorporating two- and three-dimensional molecular information (Figure 1). The proposed model outperforms existing QSAR and deep learning approaches, including models from the Ames/QSAR International Challenge [6], and demonstrates high sensitivity for identifying mutagenic compounds. Importantly, the framework is highly interpretable: GNNExplainer analysis reveals that the model learns known structural alerts associated with mutagenicity, as well as additional molecular patterns and three-dimensional contextual features in a strain-specific manner. The model further identifies previously unreported molecular substructures potentially associated with mutagenicity. Beyond mutagenicity prediction, this work highlights a generalizable graph-based MTL framework applicable to molecular and materials property prediction, supporting rational design in both drug discovery and materials science.

References

[1] Gini, G., Zanoli, F., Gamba, A, Raitano, G., and Benfenati, E., SAR and QSAR in Environmental Research, 30 (2019) 617–642.
[2] Tropsha, A., Isayev, O., Varnek, A., Schneider, G., and Cherkasov, A., Nature Reviews Drug Discovery, 23 (2023) 141–155.
[3] Mortelmans, K. and Zeiger, E., Mutation Research, 455 (2000) 29–60.
[4] Martínez, M. J., Sabando, M. V., Soto, A. J., Roca, C., Requena-Triguero, C., Campillo, N. E., Páez, J. A., and Ponzoni, I. Journal of Chemical Information and Modeling, 62 (2022) 6342–6351.
[5] Wojtuch, A., Danel, T., Podlewska, S., and Maziarka, Ł., Journal of Cheminformatics, 15 (2023).
[6] Honma, M. et al. Mutagenesis 34 (2019) 3–16.

Figures

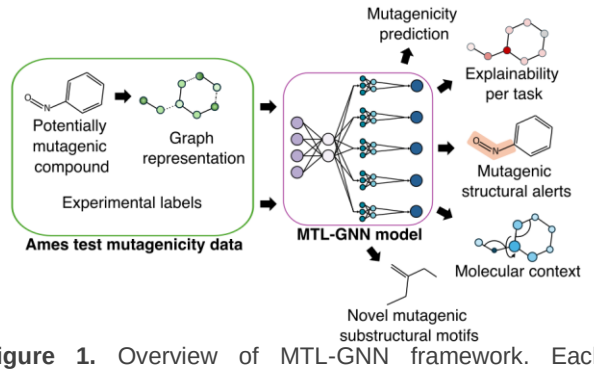

Figure 1. Overview of MTL-GNN framework. Each compound is represented as a graph, which includes various node and edge features to encode chemical information. In addition to predicting mutagenicity, our MTL-GNN model also allows for explainability analysis, screening for mutagenic structural alerts in addition to assessment of overall molecular context, and prediction of novel mutagenic substructural alerts.