[BIOINFO 2024]

Multimodal molecular representation learning enhanced by chemical knowledge-based fragmentation for toxicity prediction

Sumin Ha[1], Dongmin Bang[2,3], and Sun Kim[1,2,3,4,*]

[1] *Interdisciplinary Program in Artificial Intelligence, Seoul National University*
[2] *Interdisciplinary Program in Bioinformatics, Seoul National University* [3] *AIGENDRUG Co., Ltd.*
[4] *Department of Computer Science and Engineering, Seoul National University*
** Corresponding author: sunkim.bioinfo@snu.ac.kr*

Comprehensive toxicity in chemical and drug development evaluation across various organs is essential for enhancing safety assessments. This can lead to significant societal and economic benefits by reducing the time and cost of drug discovery. Despite the broad spectrum of toxicities, previous works have primarily focused on modeling individual toxicity predictions, lacking a comprehensive approach. To address this, we propose FATE-Tox (Fragment Attention Transformer for Equivariant Toxicity Prediction), deep graph learning model that dissects chemical compounds based on functional groups and bonds defined by three fragmentation techniques: BRICS, Murcko scaffolds, and RDKit-fragment. The model leverages unique fragments derived from each method to construct multi fragment-level graphs augmented with 3D conformer information. Traditional 3D coordinate-based equivariant graph neural networks (EGNNs) face limitations in accurately passing messages along defined chemical bonds (edges), which hinders their ability to fully capture chemical properties. Our model overcomes this by integrating both 2D and 3D structural information, enabling a multilayered analysis at both atomic and fragment levels. FATE-Tox demonstrates superior performance on seven toxicity-related benchmark datasets of MoleculeNet and TDC, outperforming existing models. Furthermore, the fragment-level graph instructs atom-wise grouping to enable interpretable attention visualizations as shown in extensive case studies, enhancing explainability and thus the model's interpretive power.