Unraveling Molecular Carcinogenicity Through Computational Insights

Zhu, Bryan (School: Bellevue Senior High School)

Carcinogenicity assessment is vital in public health and regulatory compliance in multiple industries. In vivo bioassays are costly and time-consuming, while in vitro assays can be simplistic, highlighting the need for alternatives. This study presents CarcNet, a deep learning framework for predicting molecular carcinogenicity. CarcNet integrates a Graph Attention Network for molecular graphs, focusing on critical functional groups and reactive sites; a self-attention mechanism for binary fingerprints, capturing dependencies between substructures; and a Kolmogorov-Arnold layer for continuous features, refining their representations with more expressiveness. Evaluated through five-fold cross-validation, CarcNet achieves superior performance in key metrics, outperforming state-of-the-art methods. It demonstrates robust generalization on external datasets and generated carcinogen prediction for 5 million+ unlabeled molecules. Notably, CarcNet identified carcinogenicity in FDA-approved drugs recently classified as Group 1 carcinogens (Tacrolimus, Voriconazole, Hydrochlorothiazide) and certain cancer treatments (Tamoxifen, Busulfan, Teniposide). In particular, Hydrochlorothiazide has been a widely prescribed antihypertensive drug for 60+ years. Counterfactual analysis reveals CarcNet's ability to recognize established carcinogenic functional groups, such as nitroso and aromatic amines. These results suggest CarcNet's potential for early identification of carcinogens in drug development and chemical safety evaluation. By providing rapid, accurate predictions while maintaining interpretability, CarcNet can significantly streamline the drug discovery pipeline, reduce dependency on animal testing, and enhance the efficiency of risk assessment processes.

Awards Won: