[BIOINFO 2025]

Reproducible Benchmarking of Molecular Representations and Fusion Strategy for Molecular Property Prediction

Minyoung Kim[1], and Mina Rho[1, 2, 3,*]

[1] *Department of Artificial Intelligence, Hanyang University*
[2] *Department of Computer Science, Hanyang University*
[3] *Department of Biomedical Informatics, Hanyang Univesity*
**Corresponding author: minarho@hanyang.ac.kr*

Molecular property prediction plays an essential role in drug discovery. Despite its importance, deep learning studies in this field often adopt non-uniform evaluation practices, where variations in data splitting and random seed selection obscure whether reported improvements truly reflect model capability or favorable experimental conditions. To address this issue, we systematically benchmarked six representative models, two from each of descriptor based, sequence based, and graph based approaches, across nine MoleculeNet datasets under controlled conditions with standardized splits and multiple random seeds. Results showed that model performance and relative rankings were highly sensitive to both data splitting strategies and seed selection, underscoring the importance of consistent and reproducible evaluation. Moreover, no single model achieved consistently superior performance across all datasets, which motivated the integration of complementary representations. Multi modal fusion of descriptors, sequences, and graphs achieved the most robust overall performance with an average ROC-AUC of 0.830 and an average RMSE of 1.088, yielding clear gains in datasets such as FreeSolv and ClinTox. However, analysis of ToxCast revealed that fusion can underperform when conflicting signals from different modalities override correct predictions or amplify decision boundary ambiguities. This duality suggests that while fusion enhances generalizability, its reliability depends on adaptive mechanisms to filter noise and balance modality contributions. Taken together, this study provides a controlled framework for reproducible benchmarking, identifies critical experimental factors that shape evaluation outcomes, and highlights both the challenges and opportunities of fusion learning in advancing molecular property prediction for drug discovery.