Benchmarking Pretrained Molecular Embedding Models for Molecular Representation Learning

Mateusz Praski[1], Jakub Adamczyk[1], Wojciech Czech[1]

[1]Faculty of Computer Science, AGH University of Krakow, al. Mickiewicza 30, 30-059, Krakow, Poland.

Pretrained neural networks have attracted significant interest in chemistry and small molecule drug design, with their embeddings widely used for molecular property prediction, virtual screening, and small data learning. However, there is no unified, large-scale comparison of these models in the literature, making their advantages over each other and over classical molecular fingerprints unclear.

This study presents the most extensive comparison of such models to date, evaluating 25 pretrained molecular embedding models across 25 datasets from MoleculeNet and Therapeutics Data Commons. Under a fair comparison framework using frozen embeddings and identical downstream classifiers, we assess models spanning graph neural networks, graph transformers, text-based transformers, and hybrid architectures. Model comparisons are conducted using a hierarchical Bayesian Bradley-Terry model, enabling robust statistical inference and principled handling of practical equivalence.

We arrive at a surprising result: nearly all neural models show negligible or no improvement over the baseline ECFP molecular fingerprint. Only the CLAMP model, which is also based on molecular fingerprints, performs statistically significantly better than the alternatives. These findings raise concerns about evaluation rigor in existing studies. We discuss potential causes, propose solutions, and offer practical recommendations for model selection in chemoinformatics applications.

[1] M. Praski ; J. Adamczyk ; W. Czech. Benchmarking Pretrained Molecular Embedding Models for Molecular Representation Learning. arXiv (2026).