Benchmarking GNN Performance in Virtual Screening

Anonymous

Abstract

This work benchmarks five state-of-the-art GNN variants for virtual screening on three public datasets, providing insight into performance trade-offs and model selection for compound prioritization.

1 Introduction

Virtual screening filters large chemical libraries to prioritize candidates. We evaluate GCN, GraphSAGE, GAT, SchNet, and D-MPNN with scaffold-based splits to simulate realistic testing.

2 Methods

Dataset Preparation:

- ChEMBL (classification), Tox21 (toxicity), ZINC (regression)

Model Training:

- Early stopping; Scaffold-based train/test splits

Metrics:

- ROC-AUC for classification; RMSE for regression

3 Results

4 Discussion

- D-MPNN achieved highest accuracy on ChEMBL.
- GraphSAGE performed best on Tox21 (toxicity).

5 Conclusion

Choice of architecture depends on target dataset and computational resources. Attention-based and message-passing variants tend to yield generalizable performance.

Model	ChEMBL (ROC-AUC)	Tox21 (ROC-AUC)	ZINC (RMSE)
GCN	0.87	0.76	0.34
GraphSAGE	0.85	0.81	0.33
GAT	0.89	0.78	0.32
SchNet	0.90	0.77	0.29
D-MPNN	0.92	0.80	0.28

Table 1: Benchmark results on three datasets

6 References

Benchmarks and code available upon request.