Supplemental: Taking a Respite from Representation Learning for Molecular Property Prediction

Jianyuan Deng[1], **Zhibo Yang**[2], **Hehe Wang**[3], **Iwao Ojima**[3], **Dimitris Samaras**[2], **and Fusheng Wang**[1,2,*]

[1]Stony Brook University, Department of Biomedical Informatics, Stony Brook, 11790, United States
[2]Stony Brook University, Department of Computer Science, Stony Brook, 11790, United States
[2]Stony Brook University, Department of Chemistry, Stony Brook, 11790, United States

Dataset	BACE			BBBP			HIV		
Model	RF	MOLBERT	GROVER[1]	RF	MOLBERT	GROVER[1]	RF	MOLBERT	GROVER[1]
AUROC	28	2	0	12	9	9	26	4	0
AUPRC	26	3	1	8	13	9	30	0	0
PPV	16	6	8	6	12	12	28	2	0
NPV	20	7	3	17	8	5	13	17	0
Dataset	ESOL			FreeSolv			Lipop		
Model	RF	MOLBERT	GROVER[1]	RF	MOLBERT	GROVER[1]	RF	MOLBERT	GROVER[1]
RMSE	0	0	30	2	0	28	1	27	2
MAE	0	0	30	3	0	27	2	25	3
R2	0	0	30	2	0	28	1	27	2
PEARSON_R	0	0	30	2	0	28	1	21	8

Table 1. Number of Single Fold where a Model Achieves the Best Performance using Different Metrics under Scaffold Split.

Dataset	BACE			BBBP			HIV		
Model	RF	MOLBERT	GROVER[1]	RF	MOLBERT	GROVER[1]	RF	MOLBERT	GROVER[1]
AUROC	3,924	136	0	2,105	1,013	942	3,711	349	0
AUPRC	3,797	206	57	1,370	1,934	756	4,060	0	0
PPV	2,455	599	1,046	632	1,725	1,703	4,054	6	0
NPV	3,499	369	192	2,980	653	427	1,857	2,203	0
Dataset	ESOL			FreeSolv			Lipop		
Model	RF	MOLBERT	GROVER[1]	RF	MOLBERT	GROVER[1]	RF	MOLBERT	GROVER[1]
RMSE	0	0	4,060	0	0	4,060	49	3,945	66
MAE	0	0	4,060	14	0	4,046	166	3,768	126
R2	0	0	4,060	42	0	4,018	13	4,039	8
PEARSON_R	0	0	4,060	0	0	4,060	11	3,463	586

Table 2. Number of 3-Fold Combinations where a Model Achieves the Best Performance using Different Metrics under Scaffold Split.

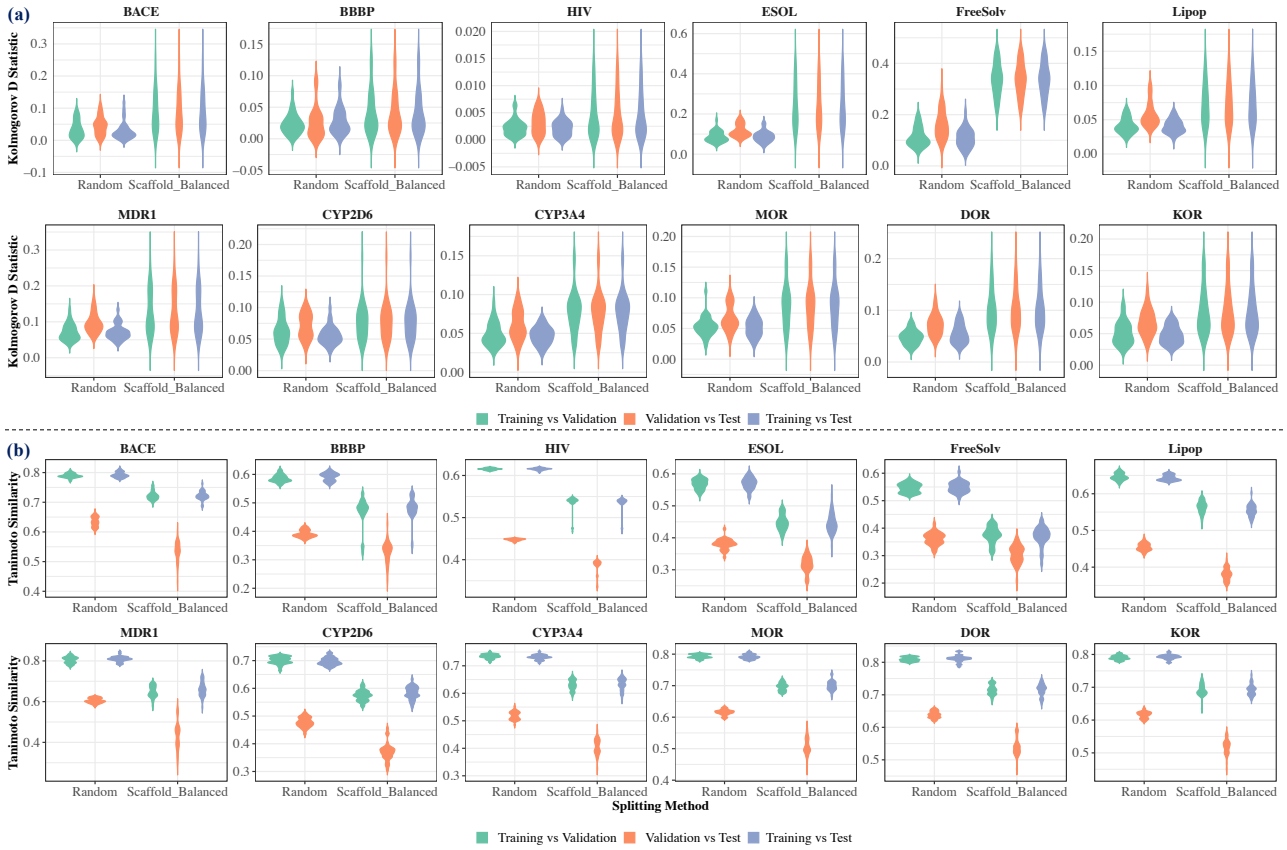

Figure 1. Distribution of the Kolmogrov *D* Statistic (a) and Tanimoto Similarity (b) among Training, Validation and Test Sets.

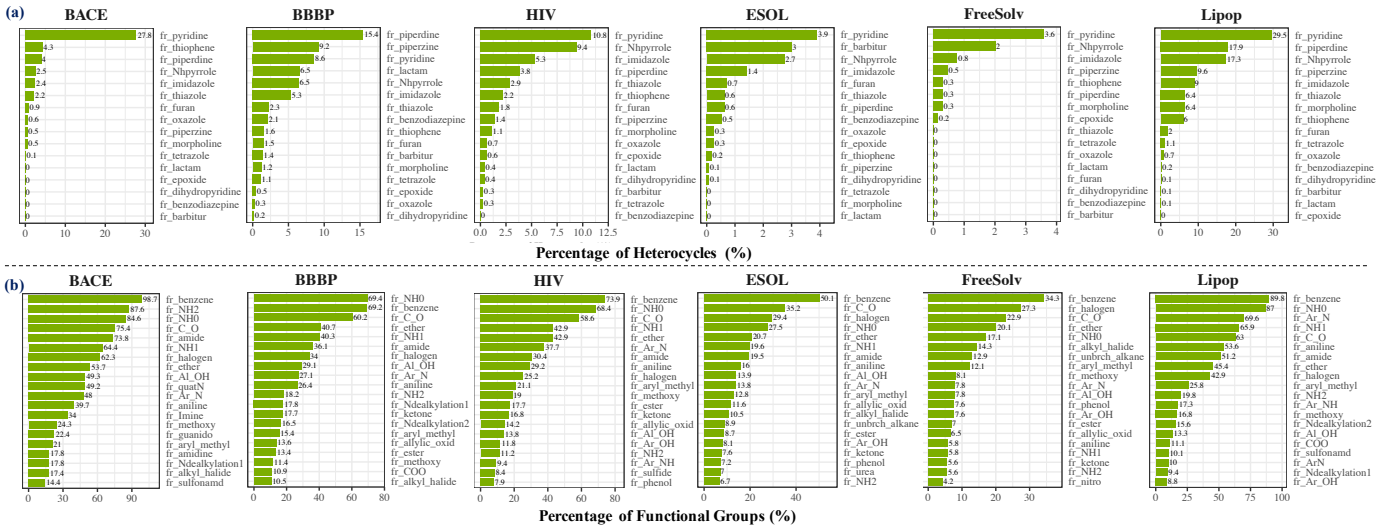

Figure 2. Top Fragments Prevalence for the Benchmark Datasets. (a). Top heterocycles (b). Top functional groups.

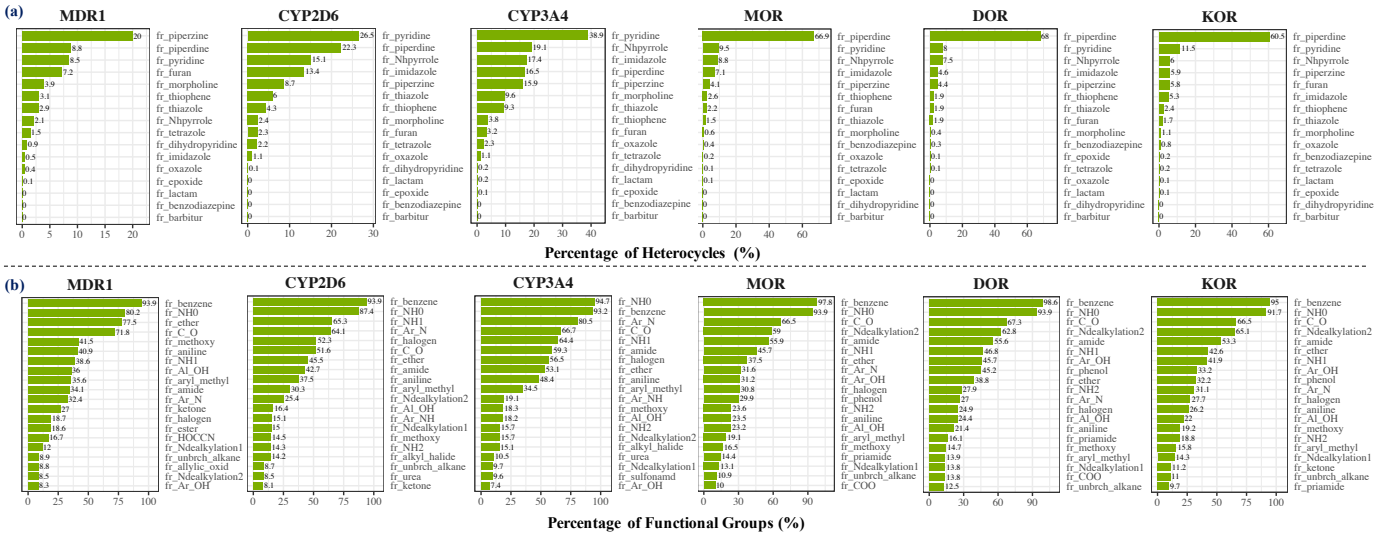

Figure 3. Top Fragments Prevalence for the Opioids-related Datasets. (a). Top heterocycles (b). Top functional groups.

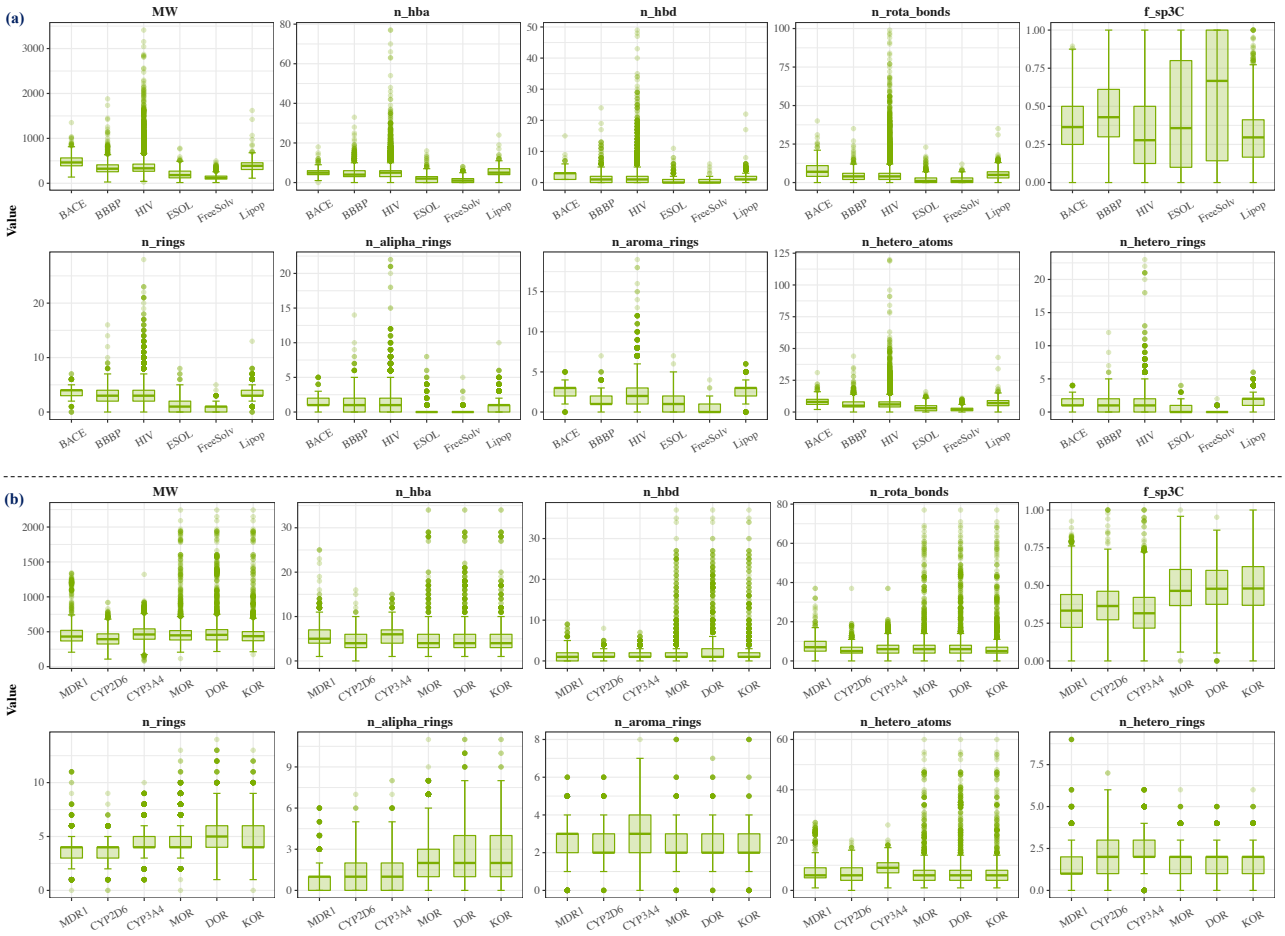

Figure 4. Distribution of Other Structural Traits for the Benchmark Datasets (a) and Opioids-related Datasets (b).

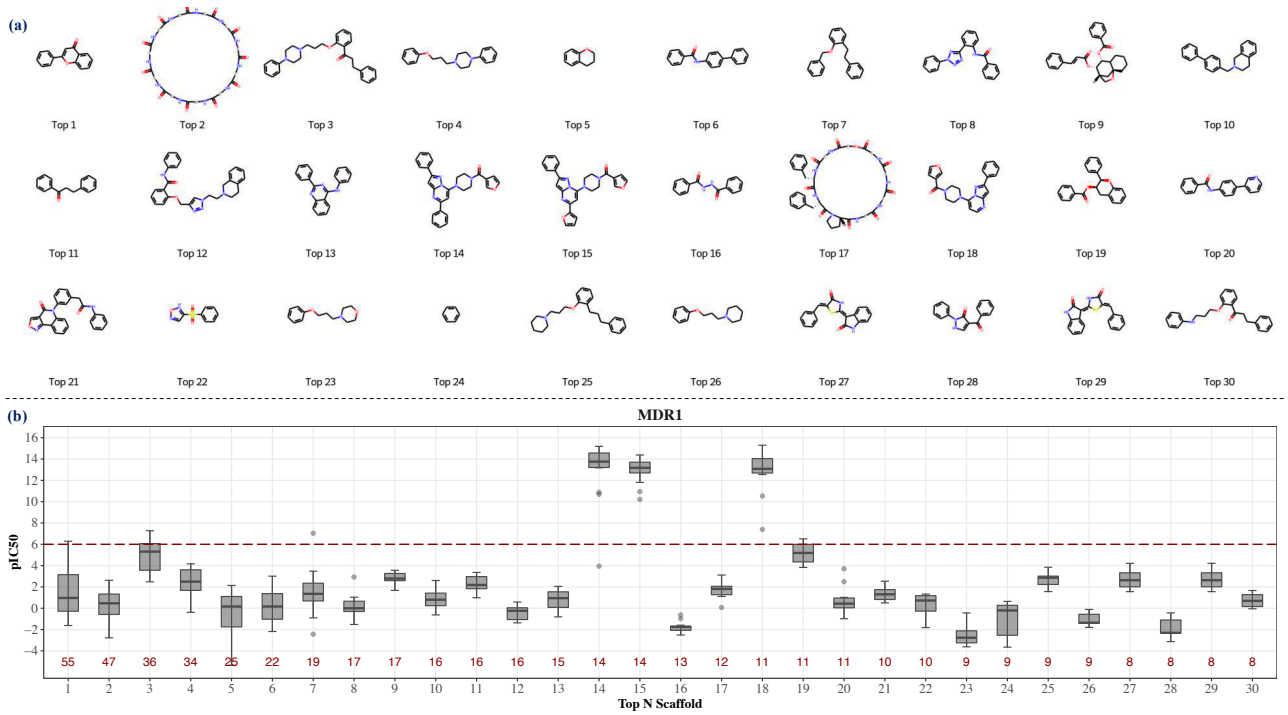

Figure 5. Scaffolds and Label Distribution in the MDR1 Dataset. (a). Top 30 scaffolds in the MDR1 dataset. (b). pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).

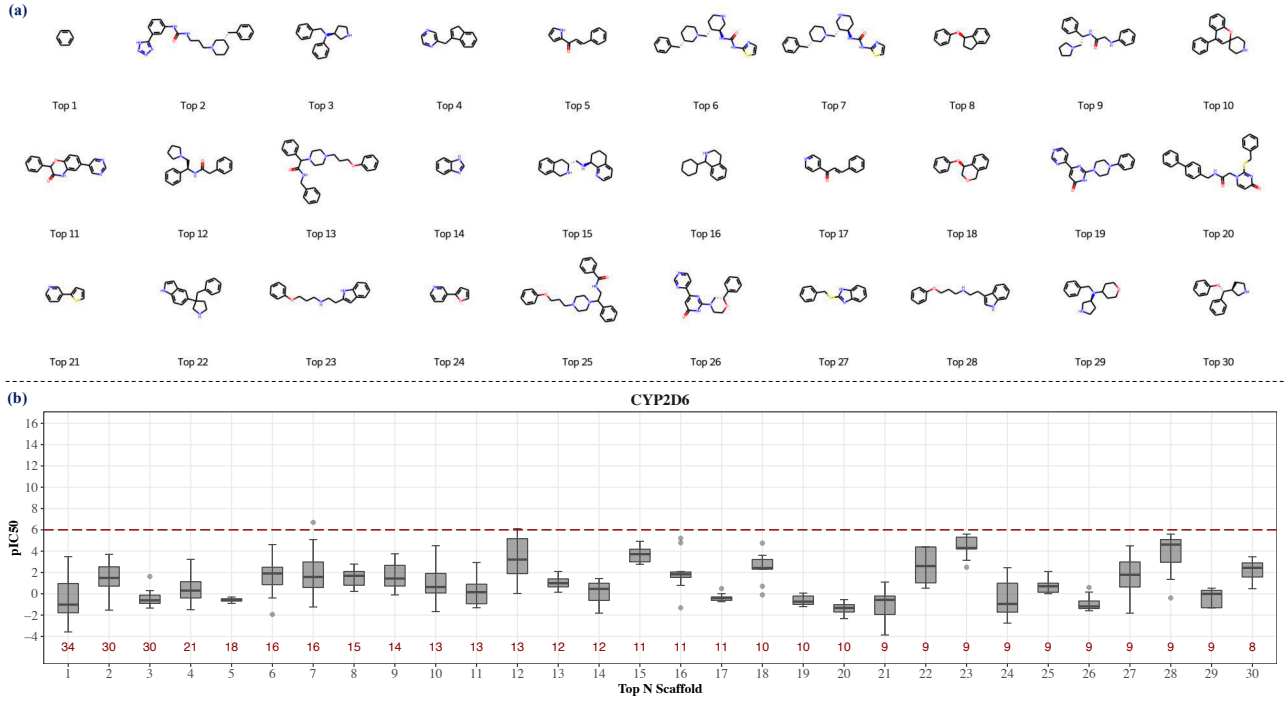

Figure 6. Scaffolds and Label Distribution in the CYP2D6 Dataset. (a). Top 30 scaffolds in the CYP2D6 dataset. (b). pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).

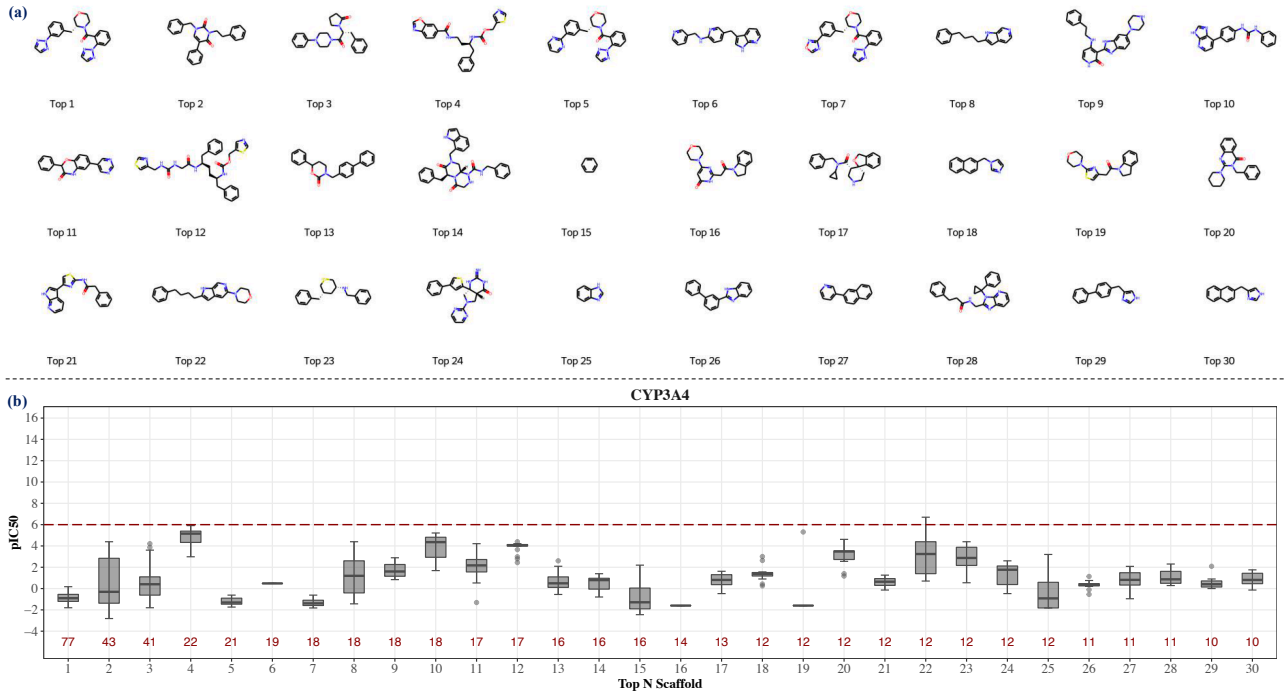

Figure 7. Scaffolds and Label Distribution in the CYP3A4 Dataset. (a). Top 30 scaffolds in the CYP3A4 dataset. (b). pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).

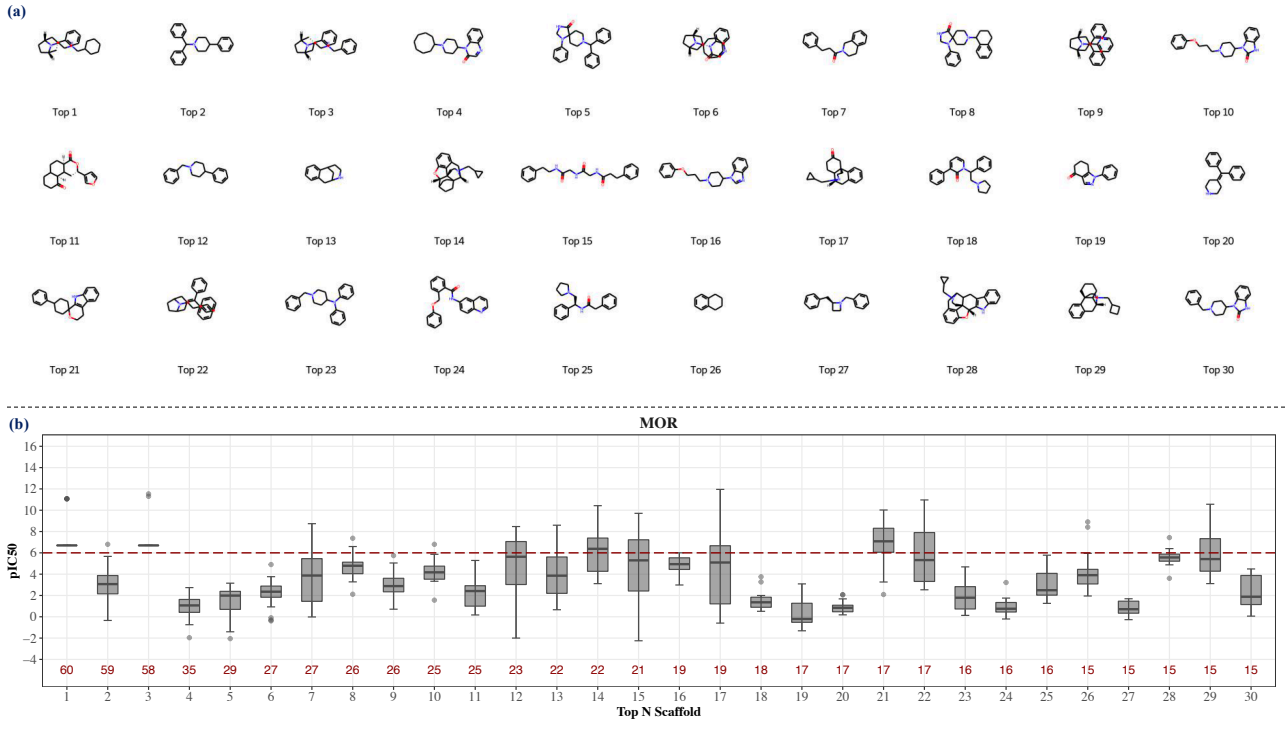

Figure 8. Scaffolds and Label Distribution in the MOR Dataset. (a). Top 30 scaffolds in the MOR dataset. (b). pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).

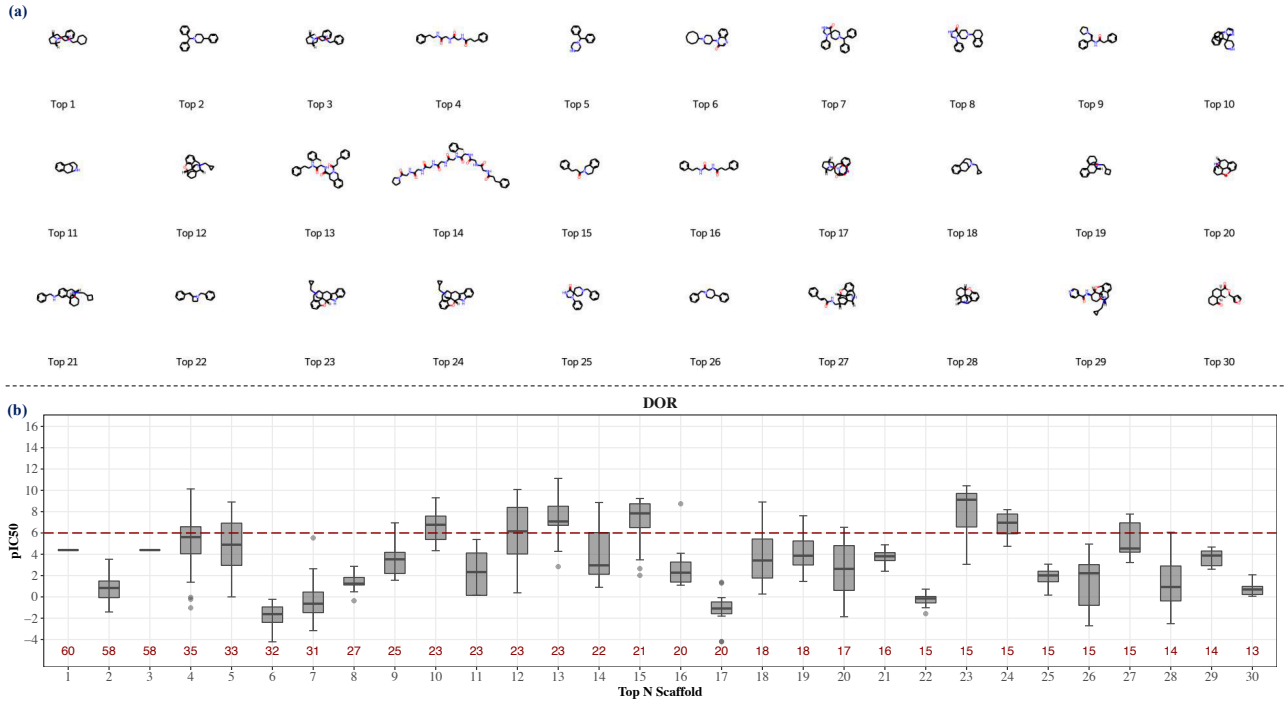

Figure 9. Scaffolds and Label Distribution in the DOR Dataset. (a). Top 30 scaffolds in the DOR dataset. (b). pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).

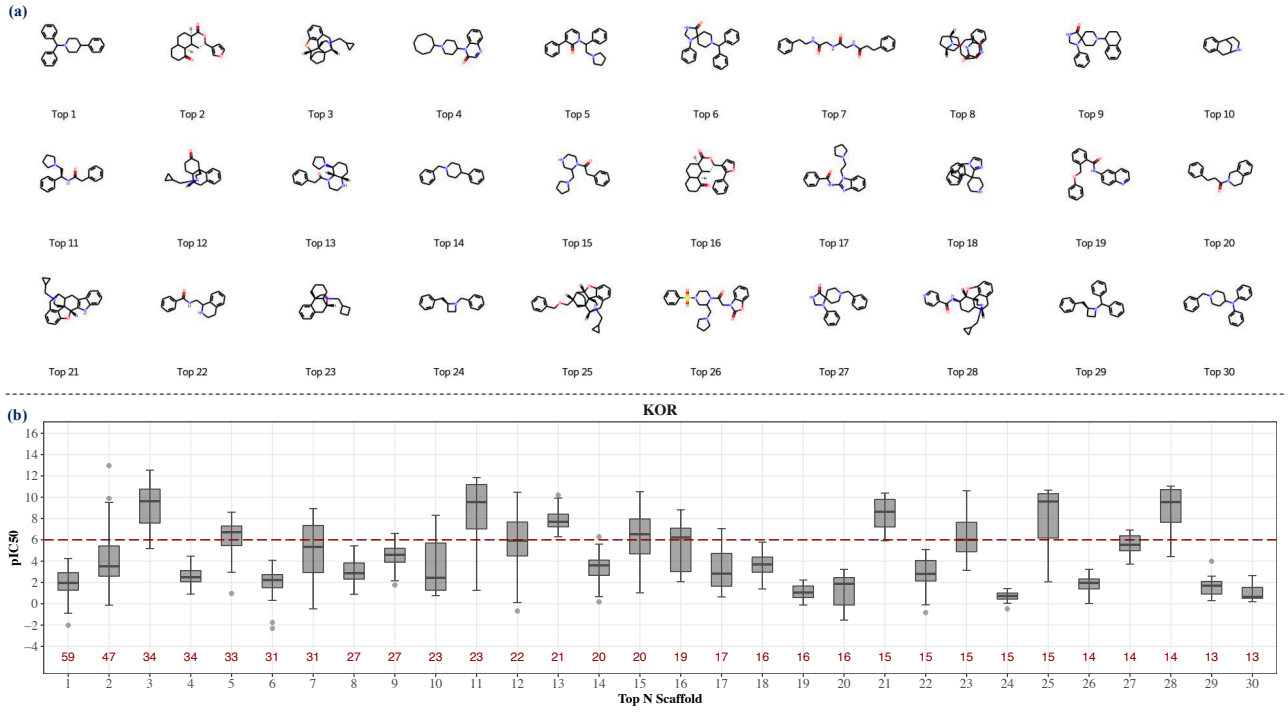

Figure 10. Scaffolds and Label Distribution in the KOR Dataset. (a). Top 30 scaffolds in the KOR dataset. (b). pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).

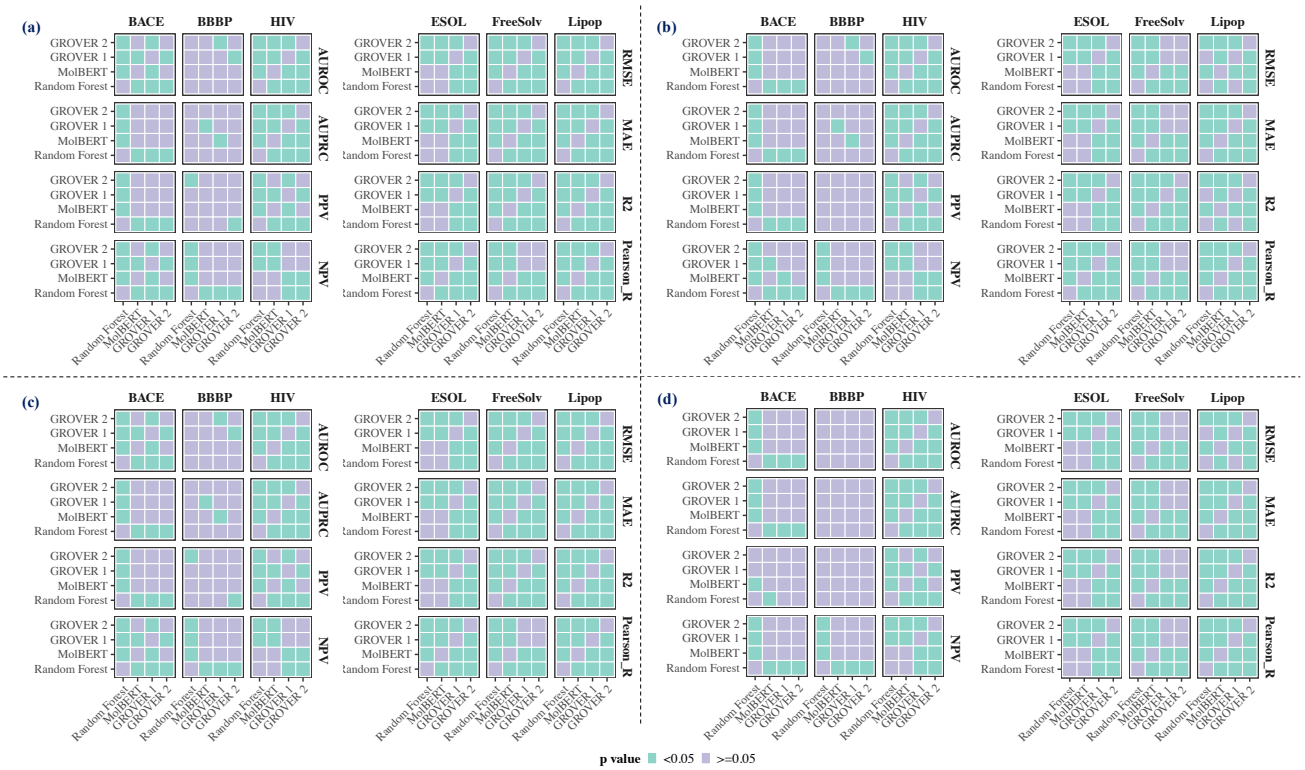

Figure 11. Pairwise Statistical Significance in Prediction Performance Comparison between Different models under Scaffold Split with Different Tests. (a). Paired *t* test (b). Unpaired *t* test (c). Wilcoxon signed-rank test (d). Wilcoxon rank-sum test.

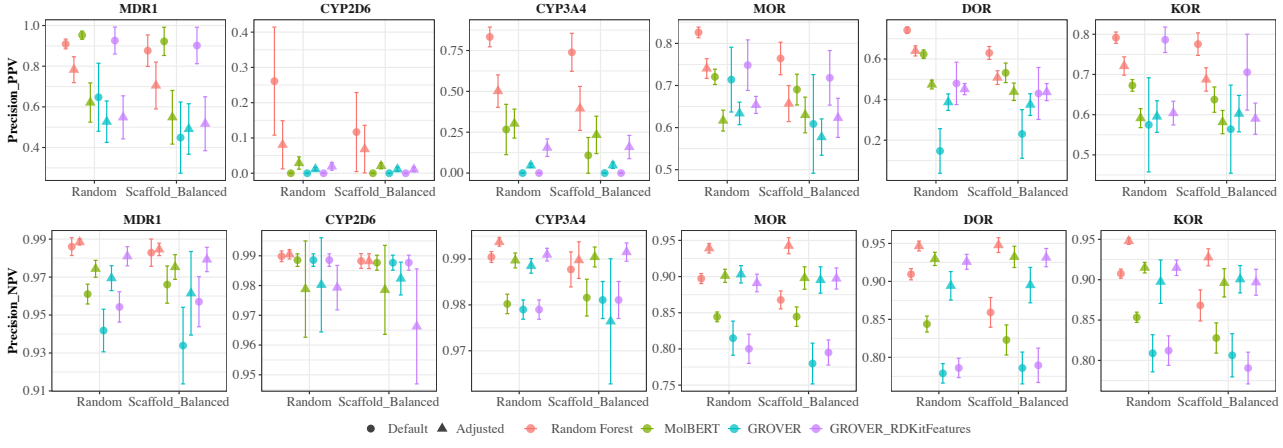

Figure 12. PPV and NPV in the Opioids-related Datasets with Default and Adjusted Probability Threshold.

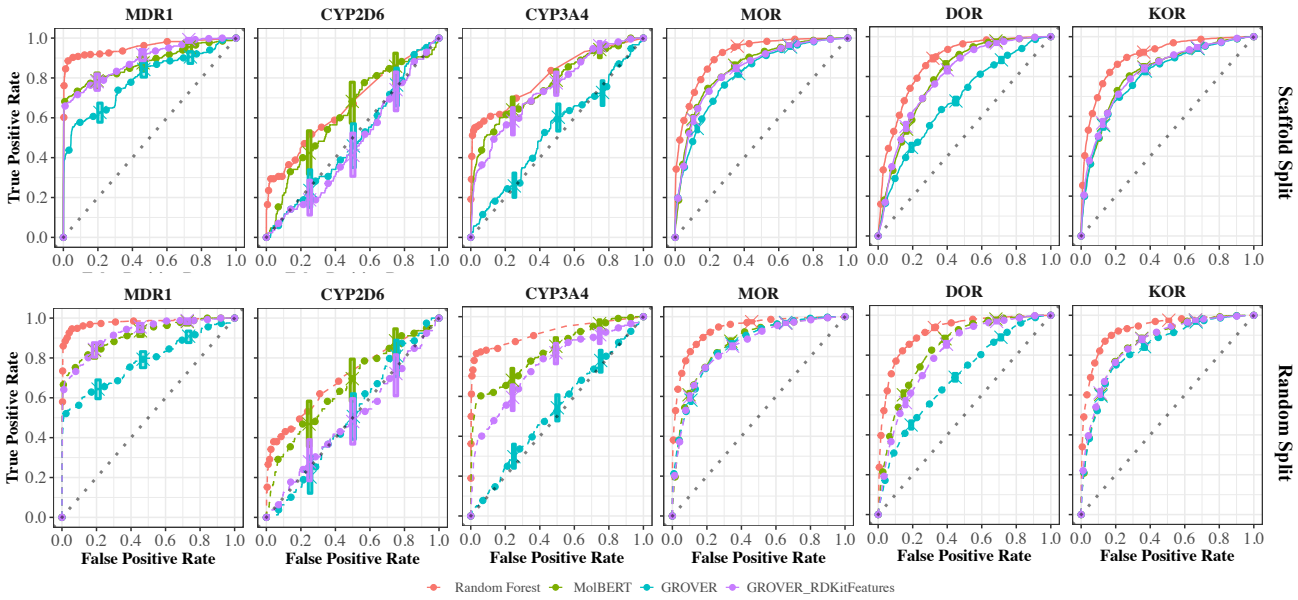

Figure 13. ROC Curves for the Opioids-related Datasets under Random and Scaffold Split.

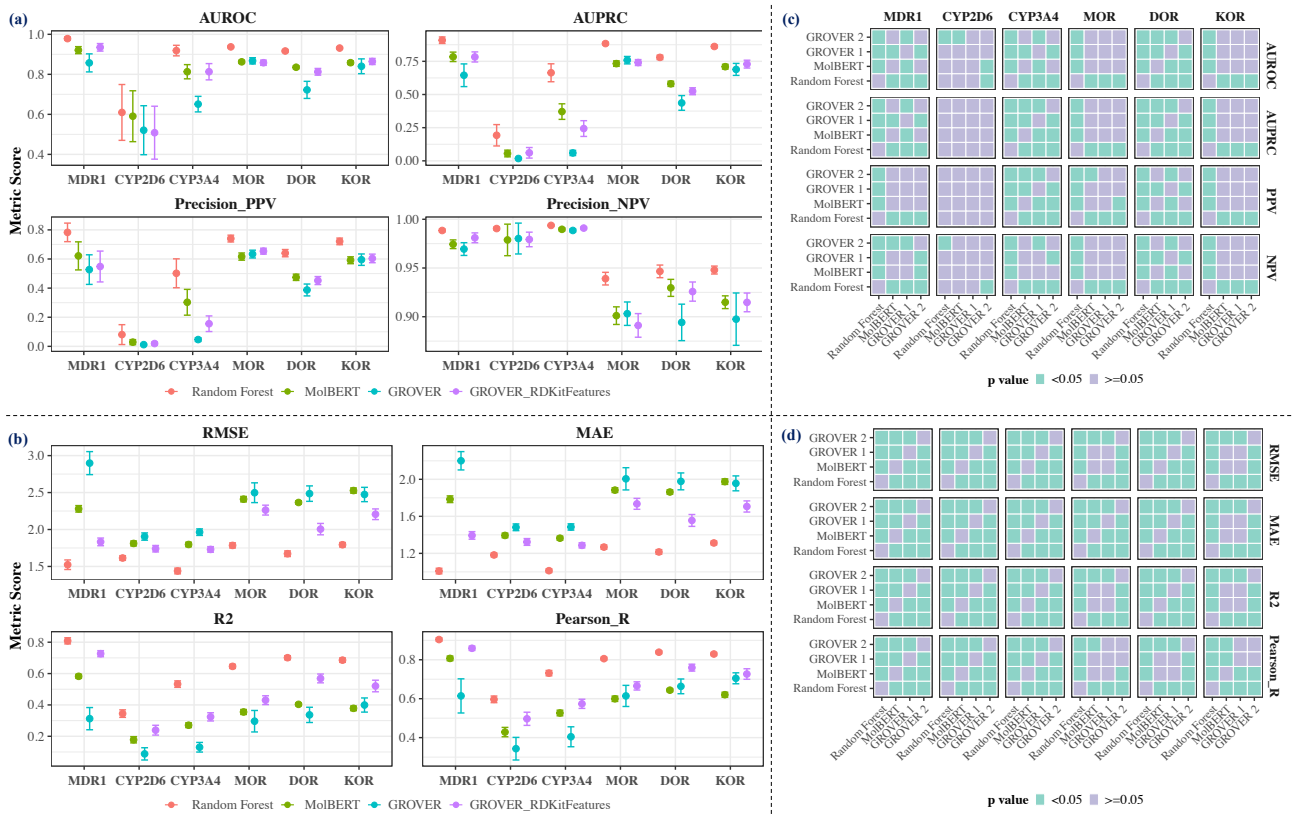

Figure 14. Prediction Performance in the Opioids-related Datasets under Random Split. (a). Point plot (mean±95% confidence interval over 30 folds) for the prediction performance in the classification setting. (b). Pairwise statistical significance in classification performance comparison between different models. (c). Point plot (mean±95% confidence interval over 30 folds) for the prediction performance in the regression setting. (d). Pairwise statistical significance in regression performance comparison between different models.