

SUPPORTING INFORMATION

GRAPH-BASED TRANSFORMER TO PREDICT THE OCTANOL-WATER PARTITION COEFFICIENT

Vyacheslav Grigorev^a, Nikita Serov^b, Timur Gimadiev^{a,b}, Assima Poyezzhayeva^b and Pavel Sidorov^c

^aA.M. Butlerov Institute of Chemistry, Kazan Federal University, 18 Kremlyovskaya Str., Kazan, 420008, Russia

^bFederal Research Center "Kazan Scientific Center of the Russian Academy of Sciences", Lobachevskogo Str. 2/31, 420111 Kazan, Russia

^cInstitute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Kita 21 Nishi 10, Kita-ku, Sapporo, 001-0021, Japan.

E-mail: pavel.sidorov@icredd.hokudai.ac.jp

Table of contents:

Table S1: Statistical significance of differences in R^2 between models on the SGNN dataset (Tukey HSD test);

Table S2: Statistical significance of differences in RMSE between models on the SGNN dataset (Tukey HSD test);

Table S3: Statistical significance of differences in MAE between models on the SGNN dataset (Tukey HSD test);

Table S4: Statistical significance of differences in R^2 between models on the GLP dataset (Tukey HSD test);

Table S5: Statistical significance of differences in RMSE between models on the GLP dataset (Tukey HSD test);

Table S6: Statistical significance of differences in MAE between models on the GLP dataset (Tukey HSD test);

Table S1: Statistical significance of differences in R^2 between models on the SGNN dataset (Tukey HSD test);

Model 1	Model 2	Mean difference	p-adj	Lower CI	Upper CI	Reject H_0
Graphormer	StructGNN	-0.0028	0.4638	-0.0075	0.0019	False
Graphormer	CheMeleon	-0.0088	0.0001	-0.0135	-0.0041	True
Graphormer	Attentive FP	-0.0182	0.0000	-0.0229	-0.0135	True
Graphormer	Chemprop	-0.0130	0.0000	-0.0177	-0.0083	True
Graphormer	RF	-0.1297	0.0000	-0.1344	-0.1251	True
StructGNN	CheMeleon	-0.0060	0.0066	-0.0107	-0.0013	True
StructGNN	Attentive FP	-0.0154	0.0000	-0.0201	-0.0107	True
StructGNN	Chemprop	-0.0102	0.0000	-0.0149	-0.0056	True
StructGNN	RF	-0.1270	0.0000	-0.1317	-0.1223	True
CheMeleon	Attentive FP	-0.0094	0.0000	-0.0140	-0.0047	True
CheMeleon	Chemprop	-0.0042	0.0959	-0.0089	0.0005	False
CheMeleon	RF	-0.1209	0.0000	-0.1256	-0.1162	True
Attentive FP	Chemprop	0.0051	0.0262	0.0004	0.0098	True
Attentive FP	RF	-0.1116	0.0000	-0.1163	-0.1069	True
Chemprop	RF	-0.1167	0.0000	-0.1214	-0.1120	True

Table S2: Statistical significance of differences in RMSE between models on the SGNN dataset (Tukey HSD test);

Model 1	Model 2	Mean difference	p-adj	Lower CI	Upper CI	Reject H_0
Graphormer	StructGNN	0.0054	0.9494	-0.0138	0.0247	False
Graphormer	CheMeleon	0.0253	0.0052	0.0061	0.0446	True
Graphormer	Attentive FP	0.0650	0.0000	0.0457	0.0842	True
Graphormer	Chemprop	0.0436	0.0000	0.0243	0.0628	True
Graphormer	RF	0.3602	0.0000	0.3410	0.3795	True
StructGNN	CheMeleon	0.0199	0.0398	0.0007	0.0391	True
StructGNN	Attentive FP	0.0595	0.0000	0.0403	0.0788	True
StructGNN	Chemprop	0.0381	0.0000	0.0189	0.0574	True
StructGNN	RF	0.3548	0.0000	0.3356	0.3740	True
CheMeleon	Attentive FP	0.0396	0.0000	0.0204	0.0589	True
CheMeleon	Chemprop	0.0182	0.0706	-0.0010	0.0375	False
CheMeleon	RF	0.3349	0.0000	0.3157	0.3541	True
Attentive FP	Chemprop	-0.0214	0.0231	-0.0407	-0.0022	True
Attentive FP	RF	0.2953	0.0000	0.2760	0.3145	True
Chemprop	RF	0.3167	0.0000	0.2974	0.3359	True

Table S3: Statistical significance of differences in MAE between models on the SGNN dataset (Tukey HSD test);

Model 1	Model 2	Mean difference	p-adj	Lower CI	Upper CI	Reject H ₀
Graphormer	StructGNN	0.0096	0.0862	-0.0009	0.0200	False
Graphormer	CheMeleon	0.0225	0.0000	0.0120	0.0329	True
Graphormer	Attentive FP	0.0420	0.0000	0.0316	0.0525	True
Graphormer	Chemprop	0.0385	0.0000	0.0281	0.0490	True
Graphormer	RF	0.2944	0.0000	0.2839	0.3048	True
StructGNN	CheMeleon	0.0129	0.0095	0.0025	0.0234	True
StructGNN	Attentive FP	0.0325	0.0000	0.0220	0.0429	True
StructGNN	Chemprop	0.0290	0.0000	0.0185	0.0394	True
StructGNN	RF	0.2848	0.0000	0.2744	0.2953	True
CheMeleon	Attentive FP	0.0195	0.0001	0.0091	0.0300	True
CheMeleon	Chemprop	0.0160	0.0010	0.0056	0.0265	True
CheMeleon	RF	0.2719	0.0000	0.2615	0.2824	True
Attentive FP	Chemprop	-0.0035	0.9009	-0.0139	0.0069	False
Attentive FP	RF	0.2524	0.0000	0.2419	0.2628	True
Chemprop	RF	0.2559	0.0000	0.2454	0.2663	True

Table S4: Statistical significance of differences in R² between models on the GLP dataset (Tukey HSD test);

Model 1	Model 2	Mean difference	p-adj	Lower CI	Upper CI	Reject H ₀
Graphormer	StructGNN	-0.0115	0.0000	-0.0131	-0.0099	True
Graphormer	CheMeleon	-0.0149	0.0000	-0.0165	-0.0133	True
Graphormer	Attentive FP	-0.0154	0.0000	-0.0170	-0.0138	True
Graphormer	Chemprop	-0.0187	0.0000	-0.0203	-0.0171	True
Graphormer	RF	-0.1222	0.0000	-0.1238	-0.1206	True
StructGNN	CheMeleon	-0.0034	0.0000	-0.0050	-0.0018	True
StructGNN	Attentive FP	-0.0039	0.0000	-0.0055	-0.0023	True
StructGNN	Chemprop	-0.0072	0.0000	-0.0088	-0.0056	True
StructGNN	RF	-0.1107	0.0000	-0.1123	-0.1091	True
CheMeleon	Attentive FP	-0.0005	0.9268	-0.0021	0.0011	False
CheMeleon	Chemprop	-0.0038	0.0000	-0.0054	-0.0022	True
CheMeleon	RF	-0.1073	0.0000	-0.1089	-0.1057	True
Attentive FP	Chemprop	-0.0033	0.0000	-0.0049	-0.0017	True
Attentive FP	RF	-0.1068	0.0000	-0.1084	-0.1052	True
Chemprop	RF	-0.1035	0.0000	-0.1051	-0.1019	True

Table S5: Statistical significance of differences in RMSE between models on the GLP dataset (Tukey HSD test);

Model 1	Model 2	Mean difference	p-adj	Lower CI	Upper CI	Reject H ₀
Graphormer	StructGNN	0.0534	0.0000	0.0452	0.0616	True
Graphormer	CheMeleon	0.0669	0.0000	0.0587	0.0751	True
Graphormer	Attentive FP	0.0708	0.0000	0.0626	0.0790	True
Graphormer	Chemprop	0.0836	0.0000	0.0754	0.0918	True
Graphormer	RF	0.4232	0.0000	0.4150	0.4314	True
StructGNN	CheMeleon	0.0136	0.0004	0.0054	0.0218	True
StructGNN	Attentive FP	0.0174	0.0000	0.0092	0.0256	True
StructGNN	Chemprop	0.0303	0.0000	0.0221	0.0385	True
StructGNN	RF	0.3699	0.0000	0.3617	0.3781	True
CheMeleon	Attentive FP	0.0038	0.6973	-0.0044	0.0120	False
CheMeleon	Chemprop	0.0167	0.0000	0.0085	0.0249	True
CheMeleon	RF	0.3563	0.0000	0.3481	0.3645	True
Attentive FP	Chemprop	0.0129	0.0008	0.0047	0.0210	True
Attentive FP	RF	0.3525	0.0000	0.3443	0.3607	True
Chemprop	RF	0.3396	0.0000	0.3314	0.3478	True

Table S6: Statistical significance of differences in MAE between models on the GLP dataset (Tukey HSD test);

Model 1	Model 2	Mean difference	p-adj	Lower CI	Upper CI	Reject H ₀
Graphormer	StructGNN	0.0441	0.0000	0.0373	0.0509	True
Graphormer	CheMeleon	0.0479	0.0000	0.0411	0.0547	True
Graphormer	Attentive FP	0.0684	0.0000	0.0616	0.0752	True
Graphormer	Chemprop	0.0650	0.0000	0.0582	0.0718	True
Graphormer	RF	0.3104	0.0000	0.3036	0.3172	True
StructGNN	CheMeleon	0.0038	0.5185	-0.0030	0.0106	False
StructGNN	Attentive FP	0.0243	0.0000	0.0176	0.0311	True
StructGNN	Chemprop	0.0209	0.0000	0.0141	0.0277	True
StructGNN	RF	0.2663	0.0000	0.2595	0.2731	True
CheMeleon	Attentive FP	0.0205	0.0000	0.0137	0.0273	True
CheMeleon	Chemprop	0.0171	0.0000	0.0103	0.0239	True
CheMeleon	RF	0.2625	0.0000	0.2557	0.2693	True
Attentive FP	Chemprop	-0.0035	0.6234	-0.0102	0.0033	False
Attentive FP	RF	0.2420	0.0000	0.2352	0.2488	True
Chemprop	RF	0.2454	0.0000	0.2386	0.2522	True