Supporting Information: Costless Performance Improvement in Machine Learning for Graph-based Molecular Analysis

Gyoung S. Na[a*], Hyun Woo Kim[a], and Hyunju Chang[*]

Chemical Data-Driven Research Center, Korea Research Institute of Chemical Technology (KRICT), Daejeon, 34114, Korea

E-mail: ngs0@krict.re.kr; hjchang@krict.re.kr

[a]These authors contributed equally to this work.

Table S1 shows the statistics of the target properties in the benchmark dataset.

Table S1: Statistics of the target properties in the benchmark datasets.

Dataset	Range	Average	Standard deviation
ESOL	[-11.60, 1.58]	-3.05	2.10
FreeSolv	[-25.47, 3.43]	-3.80	3.84
Lipophilicity	[-1.5, 4.5]	2.2	1.2
PDBbind	[0.4, 9.3]	6.2	1.7
QM7	[-2.9034, 5.1144]	-1.4059	0.9999
QM8-E1	[0.0653, 0.4964]	0.2168	0.0446
QM8-F1	[0.0, 0.5254]	0.0235	0.0579
QM9-HLG	[6.002, 46.969]	31.601	4.063
QM9-HC	[0.0246, 0.6221]	0.2511	0.0475
QM9-IP	[6.31, 196.62]	75.19	8.19

For the reproducibility of the proposed algorithm, we present the atomic features used

in our experiment. Table S2 shows the atomic features used in EGCN. Note that the atomic features are represented as the node features in EGCN and they are given by Python mendeleev package[1].

Table S2: Atomic features of EGCN.

Name	Unit	Comment
Atomic weight	-	-
Atomic radius	pm	-
Atomic volume	cm^3/mol	-
Dipole polarizability	a.u.	-
Fusion heat	kJ/mol	-
Thermal conductivity	W/(mK)	Thermal conductivity at 25 C
Van der Waals radius	pm	-
En. pauling	-	Pauling's scale of electronegativity

[1]https://mendeleev.readthedocs.io/en/stable/index.html