

Supplementary Materials - TUNA: A Target-aware Unified Network for Protein-Ligand Binding Affinity Prediction via Multi-Modal Feature Integration

Jaesuk Yoon¹, Yeojin Kim¹, Hyunju Lee^{*1}

¹Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology,
Gwangju, Republic of Korea

^{*}Corresponding author, E-mail: hyunjlee@gm.gist.ac.kr

S.I. CONFIGURATION OF THE GCN

We utilize PyTorch’s built-in *GCNConv* module to construct a two-layer graph convolutional network (GCN). As illustrated in the first equation, the input node features $\mathbf{X}^{(0)}$ are first processed by two consecutive GCN layers:

$$\mathbf{X}^{(1)} = \hat{A}\mathbf{X}^{(0)}\mathbf{W}^{(0)}, \quad (1)$$

$$\mathbf{X}^{(2)} = \hat{A}\mathbf{X}^{(1)}\mathbf{W}^{(1)}, \quad (2)$$

where, \hat{A} denotes the normalized adjacency matrix with self-loops, which is handled within the PyTorch *GCNConv* implementation as $\hat{A} = A + I$. The matrices $\mathbf{W}^{(0)}$ and $\mathbf{W}^{(1)}$ represent learnable weight parameters in the first and second convolution layers, respectively.

To stabilize training and normalize the output features, we apply batch normalization using PyTorch’s built-in *BatchNorm1d* function. The final output is obtained by applying ReLU activation to the batch-normalized features.

$$\mathbf{X}^{out} = \text{ReLU}(\text{BN}(\mathbf{X}^{(2)})) \quad (3)$$

S.II. WORKFLOW OF A 2D LIGAND ENCODER BASED ON GRAPH DIFFUSION BLOCKS

The initial input is first processed through a single graph diffusion block. Subsequently, the dimension of the features is progressively expanded through iterative feature expansion. At each step, the graph diffusion block increases the dimensionality following the form $\text{input_dimension} + \text{growth_rate} \times i$, by aggregating the cumulative features from previous layers. As illustrated in the Figure S1, this expansion process is repeated 8 times.

Once the feature expansion is completed, the final high-dimensional representation is passed through a feature reduction process, which reduces the dimensionality to $\text{input_dimension}/2$. This sequence of feature expansion followed by reduction is repeated 3 times. Finally, a linear layer projects the output to generate the 2D ligand feature. The overview of the 2D ligand encoder is in Figure S1.

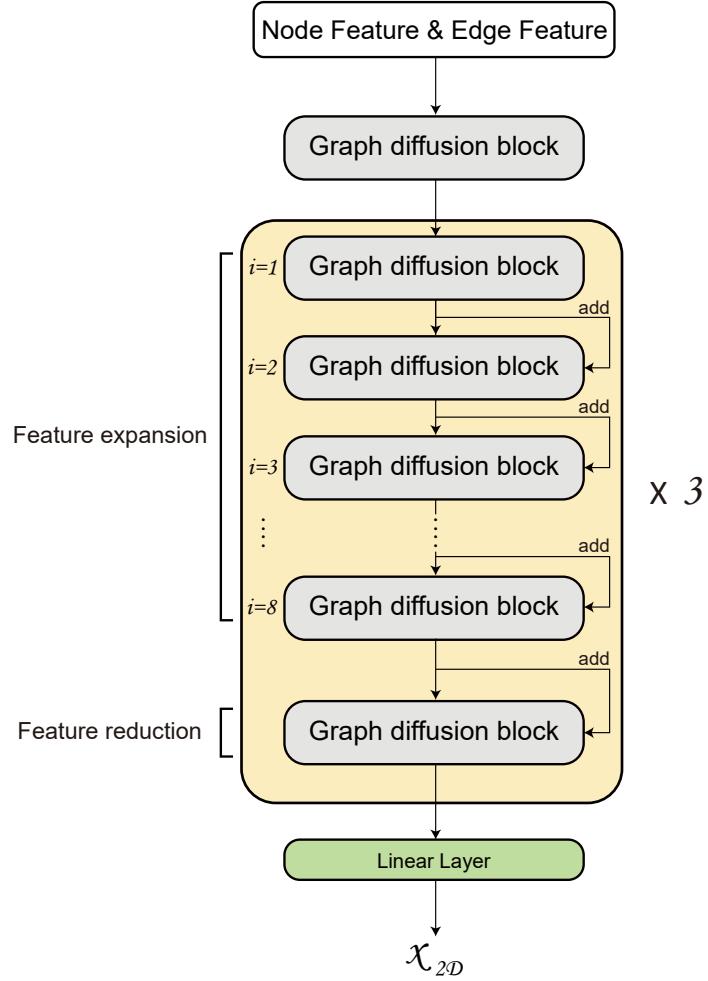


Fig. S1. Structure of the 2D ligand encoder composed of graph diffusion blocks with iterative feature expansion and reduction.

S.III. FULL RESULTS ON THE BINDINGDB DATASET

TABLE S1
OVERALL RMSE RESULTS ON THE BINDINGDB DATASET

Dataset	Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Random Split
Test	TUNA-Pred (fpocket)	1.375	1.359	1.322	1.410	1.412	0.842
	TUNA-Pred (DeepSurf)	1.391	1.363	1.335	1.343	1.297	0.829
	Deepaffinity	1.478	1.434	1.453	1.494	1.481	0.868
	DeepDTA	1.414	1.585	1.465	1.376	1.530	0.867
	GraphDTA	0.339	0.260	0.249	0.338	0.324	0.820
Kinase	TUNA-Pred (fpocket)	1.326	1.570	1.418	1.400	1.390	1.314
	TUNA-Pred (DeepSurf)	1.398	1.490	1.183	1.177	1.191	1.400
	Deepaffinity	1.613	1.963	1.778	1.608	1.693	1.652
	DeepDTA	1.705	1.747	1.686	1.562	1.821	1.624
	GraphDTA	0.171	0.187	0.064	0.150	0.184	0.163
Channel	TUNA-Pred (fpocket)	1.370	1.564	1.413	1.336	1.622	1.447
	TUNA-Pred (DeepSurf)	1.420	1.625	1.662	1.551	1.498	1.736
	Deepaffinity	1.413	1.467	1.471	1.444	1.495	1.449
	DeepDTA	1.645	1.628	1.657	1.589	1.636	1.655
	GraphDTA	0.168	0.136	0.388	0.259	0.229	0.169
ER	TUNA-Pred (fpocket)	1.472	1.339	1.442	1.984	1.432	1.678
	TUNA-Pred (DeepSurf)	1.361	1.451	1.362	1.351	1.309	1.378
	Deepaffinity	1.806	1.846	1.621	1.426	1.717	1.687
	DeepDTA	1.752	1.961	1.917	1.715	2.020	1.789
	GraphDTA	0.202	0.182	0.018	0.144	-0.166	0.167
GPCR	TUNA-Pred (fpocket)	1.392	1.437	1.396	1.365	1.408	1.459
	TUNA-Pred (DeepSurf)	1.369	1.384	1.310	1.274	1.325	1.375
	Deepaffinity	1.526	1.503	1.485	1.712	1.559	1.623
	DeepDTA	1.775	1.861	1.807	1.848	1.815	1.662
	GraphDTA	0.151	0.120	0.153	0.114	0.128	0.111

TABLE S2
OVERALL PCC RESULTS ON THE BINDINGDB DATASET

Dataset	Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Random Split
Test	TUNA-Pred (fpocket)	0.451	0.455	0.498	0.434	0.427	0.822
	TUNA-Pred (DeepSurf)	0.412	0.461	0.529	0.470	0.513	0.838
	Deepaffinity	0.380	0.423	0.398	0.399	0.411	0.820
	DeepDTA	0.481	0.442	0.468	0.367	0.397	0.813
	GraphDTA	0.339	0.260	0.249	0.338	0.324	0.820
Kinase	TUNA-Pred (fpocket)	0.271	0.247	0.331	0.281	0.342	0.354
	TUNA-Pred (DeepSurf)	0.253	0.219	0.394	0.373	0.381	0.279
	Deepaffinity	0.228	0.312	0.190	0.227	0.143	0.273
	DeepDTA	0.227	0.111	0.308	0.179	0.270	0.247
	GraphDTA	0.171	0.187	0.064	0.150	0.184	0.163
Channel	TUNA-Pred (fpocket)	0.159	-0.110	0.147	0.292	-0.066	0.082
	TUNA-Pred (DeepSurf)	0.217	-0.201	0.084	-0.032	0.072	-0.148
	Deepaffinity	0.110	0.162	0.154	0.213	0.211	0.160
	DeepDTA	-0.012	-0.024	-0.005	0.012	0.035	0.031
	GraphDTA	0.168	0.136	0.388	0.259	0.229	0.169
ER	TUNA-Pred (fpocket)	0.323	0.381	0.178	0.191	0.241	0.173
	TUNA-Pred (DeepSurf)	0.349	0.223	0.203	0.239	0.341	0.315
	Deepaffinity	0.128	0.303	0.298	0.039	0.053	0.011
	DeepDTA	0.217	0.184	0.121	0.080	0.178	0.096
	GraphDTA	0.202	0.182	0.018	0.144	-0.166	0.167
GPCR	TUNA-Pred (fpocket)	0.206	0.197	0.223	0.211	0.193	0.178
	TUNA-Pred (DeepSurf)	0.222	0.236	0.232	0.233	0.262	0.231
	Deepaffinity	0.203	0.181	0.237	0.170	0.209	0.101
	DeepDTA	0.137	0.111	0.176	0.116	0.126	0.181
	GraphDTA	0.151	0.120	0.153	0.114	0.128	0.111

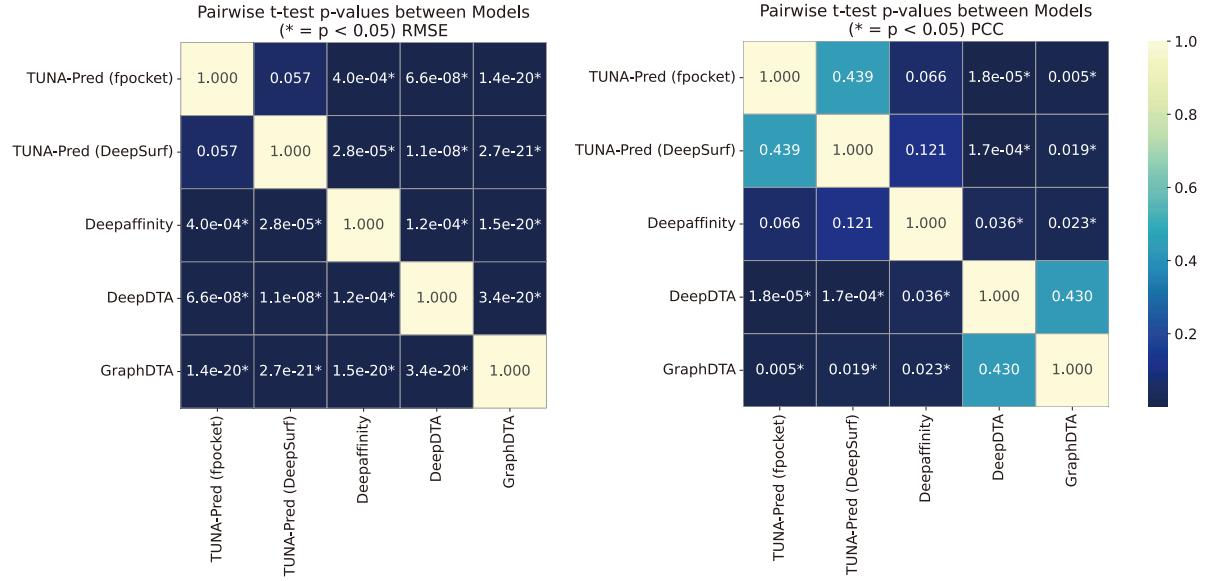


Fig. S2. Pairwise t-test p-values between baseline models on the entire BindingDB dataset (30 results: Test, Kinase, Channel, ER, and GPCR with five folds and random split). Left: RMSE comparison. Right: PCC comparison.

S.IV. EVALUATION OF MODEL BIAS ACROSS PROTEIN FAMILIES

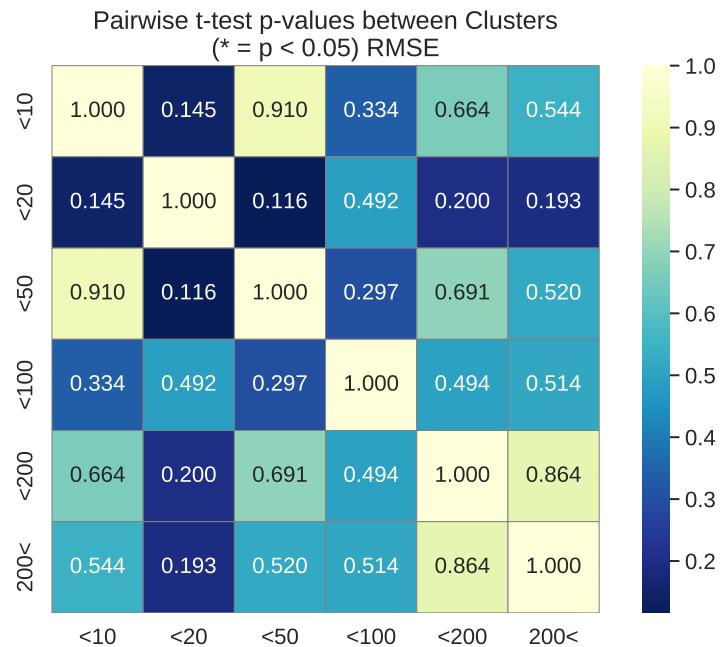


Fig. S3. Pairwise t-test p-values between protein clusters of different sizes in the BindingDB test set (RMSE). No statistically significant difference was observed between small and large protein families, indicating that the model performance was unbiased across protein clusters.

TABLE S3

PERFORMANCE COMPARISON ON TEST_2013 DATASET ACROSS PROTEIN FAMILIES USING REAL POCKET AND POCKET PREDICTION TOOLS (DEEPSURF AND FPOCKET)

Clan ID	n	TUNA	TUNA-Pred (DeepSurf)	TUNA-Pred (fpocket)
AB_hydrolase	5	1.206	1.377	1.937
APC	3	1.445	1.032	2.285
Arginase	3	1.972	1.973	2.702
Beta-lactamase	3	0.445	2.304	1.685
Beta_propeller	3	2.007	1.477	2.367
Calycin	3	1.556	1.226	1.371
E-set	13	1.464	1.872	1.763
EGF	3	1.579	2.254	2.328
EPT_RTPC	3	1.187	1.563	1.266
GBD	3	1.817	2.444	1.694
GH_CE	3	1.861	2.745	2.076
GT-B	3	0.377	1.097	1.408
Hexosaminidase	3	1.857	1.699	2.362
Ion_channel	6	1.626	2.074	1.237
Kringle	6	1.603	1.623	2.094
NADP_Rossmann	6	1.061	1.340	1.619
PKinase	18	1.251	1.397	1.362
PSI	3	1.964	2.125	2.206
Peptidase_AA	7	1.225	2.316	2.150
Peptidase_MA	9	1.254	1.583	1.892
Peptidase_PA	3	0.731	0.962	1.429
Phosphatase	3	0.758	0.456	0.647
RNase_H	3	1.770	1.491	1.124
RdRP	6	1.854	1.168	1.433
SH2-like	5	0.721	1.104	1.308
SH3	2	0.875	0.674	0.579
TIM_barrel	18	1.276	1.615	1.565
Terp_synthase	3	1.211	1.550	1.690
Thioredoxin	3	1.330	1.080	0.415
Transthyretin	3	0.920	0.945	1.410
N/A	40	1.707	1.740	1.850

TABLE S4

PERFORMANCE COMPARISON ON TEST_2016 DATASET ACROSS PROTEIN FAMILIES USING REAL POCKET AND POCKET PREDICTION TOOLS (DEEPSURF AND FPOCKET)

Clan ID	n	TUNA	TUNA-Pred (DeepSurf)	TUNA-Pred (fpocket)
AB_hydrolase	5	0.959	1.656	1.977
ADP-ribosyl	5	0.605	0.694	0.513
APC	5	1.324	1.269	2.025
Beta-lactamase	10	1.056	1.391	1.512
Calycin	5	1.467	0.790	1.111
EGF	5	1.410	2.002	1.939
GH_CE	5	2.003	2.383	1.990
GT-B	5	0.491	1.181	1.734
HD_PDEase	10	1.479	1.690	1.668
Hexasaminidase	5	1.806	1.309	1.670
Ion_channel	10	1.155	1.794	1.166
Kringle	10	1.569	1.743	1.787
NADP_Rossmann	5	1.212	1.495	1.658
P53-like	5	1.796	2.298	2.607
PAN	5	1.035	0.996	0.853
PKinase	35	1.172	1.188	1.167
Peptidase_AA	11	1.349	2.106	1.689
Peptidase_MA	5	1.015	0.806	1.480
Peptidase_PA	5	0.492	0.972	1.236
Phosphatase	5	1.118	0.875	1.079
RNase_H	5	1.576	1.258	1.447
RdRP	5	1.033	0.689	1.306
SH2-like	15	1.289	1.266	1.371
SH3	4	0.986	1.464	1.005
TIM_barrel	20	1.385	1.381	1.529
Terp_synthase	5	1.454	1.453	1.710
Ubiquitin	5	1.530	1.821	1.073
N/A	75	1.402	1.580	1.704

S.V. ADDITIONAL ABLATION RESULTS FOR OTHER EVALUATION METRICS

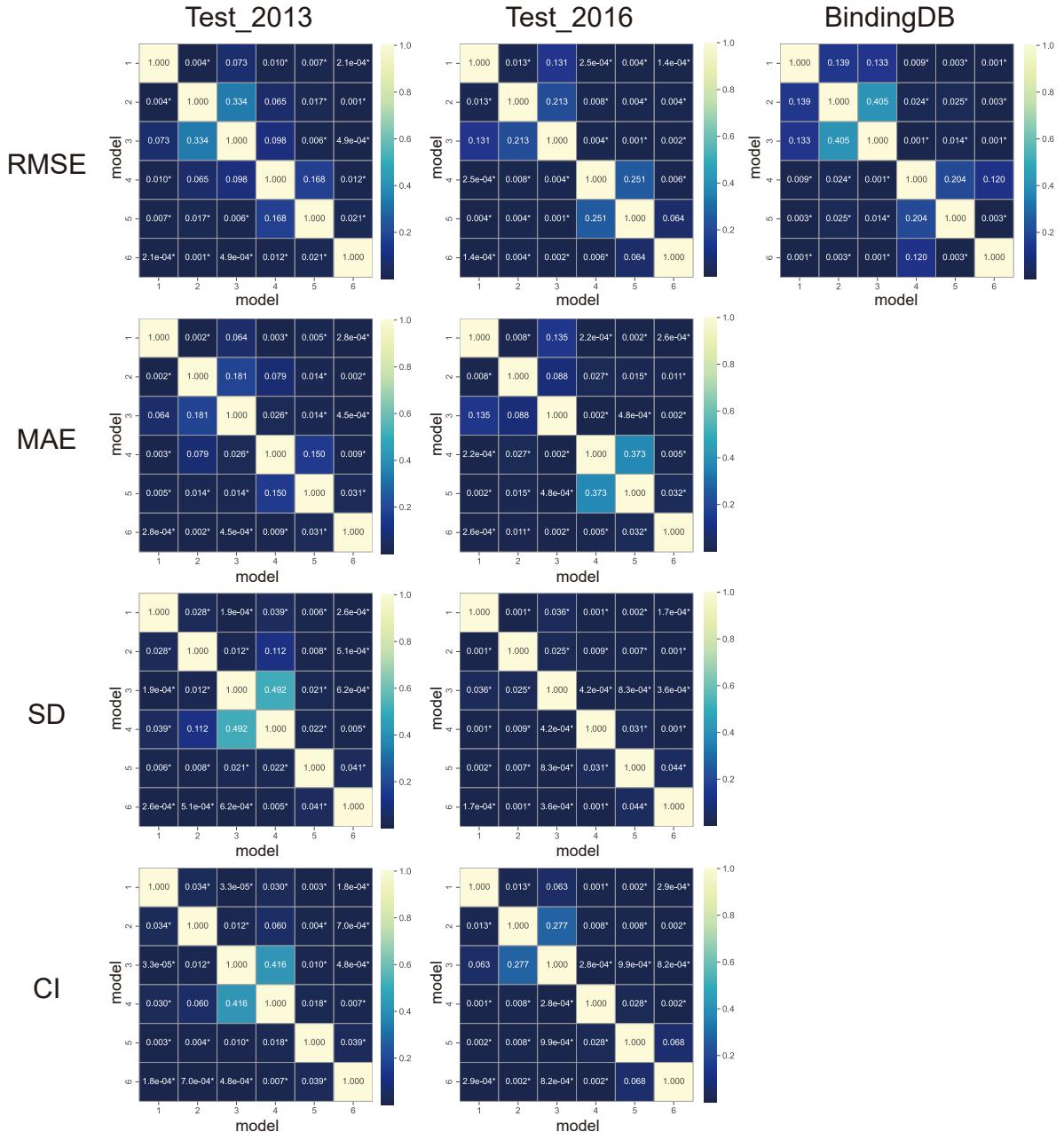


Fig. S4. Pairwise p-values for differences in performance metrics among six ablation models, evaluated on Test_2013, Test_2016, and the BindingDB test set. All models were trained using five-fold cross-validation on either the PDBbind or BindingDB training set. Each cell indicates the p-value comparing the performance of two models based on various metrics, including RMSE, MAE, SD and CI. Asterisks (*) denote statistically significant differences at the $p < 0.05$ level.