Supporting Information for MGraphDTA: Deep Multiscale Graph Neural Network for Interpretable Drug-target binding affinity Prediction

Ziduo Yang[a#], Weihe Zhong[a#], Lu Zhao[a,b], Calvin Yu-Chian Chen[a,c,d]*

[a]*Artificial Intelligence Medical Center, School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, 510275, China*

[b]*Department of Clinical Laboratory, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510655, China*

[c]*Department of Medical Research, China Medical University Hospital, Taichung 40447, Taiwan*

[d]*Department of Bioinformatics and Medical Engineering, Asia University, Taichung, 41354, Taiwan*

[#] *Equal contribution*

***** *Corresponding Authors*

Calvin Yu-Chian Chen, Ph.D.

Dean of Artificial Intelligence Medical Center, Professor of School of Intelligent Systems Engineering, Sun Yat-sen University.

TEL: 02039332153

E-mail: chenyuchian@mail.sysu.edu.cn

S1. Machine learning methods

In this study, three traditional PCM methods, namely random forest (RF), support vector machine (SVM), and feed-forward neural network (FNN) were developed to predict binding affinities. All traditional PCM methods and graph-based methods shared the same training, validation, and test sets. For traditional PCM methods, we set a list of hyper-parameters to optimize and reported the results under the best hyper-parameters. Note that hyper-parameters were only optimized in the datasets using the random split setting, and keep fixed for all other split settings to save computation cost.

S1.1 Input representation

In the present study, we chose the most common features for PCM models: extended connectivity fingerprints (ECFP)[1,2] and protein sequence composition descriptors (PSC)[3] for compound and protein representation. The molecular fingerprints and protein descriptors were calculated using RDkit[4] and propy[3]. For each drug-target pair, we obtained a 8676-dimensional feature vector by combining ECFP and PSC. The input feature vectors were standardized by removing the mean and scaling to unit variance. Since the combined feature vectors were sparse, that contained only a small portion of no-zero values, we used a decision tree regressor to select important features. The decision tree regressor was implemented using a python library sklearn[5] with default parameters.

S1.2 Random forest

RF is a popular algorithm based on ensemble learning ideas for both classification and regression problems that operate by constructing a strong classifier/regressor by an ensemble of many decision trees[6]. For RF, the following hyper-parameters were optimized using Optuna[7]: n_estimators {50, 100, 150, 200}, max_depth {2 to 8}, min_samples_leaf {1, 3, 5, 10}, min_impurity_decrease {0 to 0.01}.

S1.3 Support vector machine

SVM[8] is one of the most popular machine learning methods, and it can be used for classification and regression. In this work, the SVM was implemented using a python

library sklearn[5]. The radial basis function was used as the kernel and the following hyper-parameters were optimized using Optuna[7]: C {0.1 to 10} and gamma values {0 to 0.5}.

S1.4 Feed-forward neural network

A feedforward neural network is an artificial neural network wherein the information moves from the input nodes, through the different nodes, and to the output nodes in only one direction. In this work, the FNN was implemented using a python library PyTorch[9]. For FNN, the following hyper-parameters were optimized using Optuna[7]: number_of_hidden_units {512, 1024, 2048}, number_of_hidden_layers {2, 3, 4}, dropout {0, 0.1, 0.2, 0.5}

Reference

1 E. B. Lenselink, N. Ten Dijke, B. Bongers, G. Papadatos, H. W. T. Van Vlijmen, W. Kowalczyk, A. P. IJzerman and G. J. P. Van Westen, *J. Cheminform.*, 2017, **9**, 1–14.

2 A. Mayr, G. Klambauer, T. Unterthiner, M. Steijaert, J. K. Wegner, H. Ceulemans, D.-A. Clevert and S. Hochreiter, *Chem. Sci.*, 2018, **9**, 5441–5451.

3 D.-S. Cao, Q.-S. Xu and Y.-Z. Liang, *Bioinformatics*, 2013, **29**, 960–962.

4 G. Landrum, *Components*, 2011.

5 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg and others, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

6 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32.

7 T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

8 V. V Zernov, K. V Balakin, A. A. Ivaschenko, N. P. Savchuk and I. V Pletnev, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 2048–2056.

9 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, *Adv. Neural Inf. Process. Syst.*, 2019, **32**, 8026--8037.

S2. Supplemental Tables

Table S1. The features used in regression tasks (Davis, KIBA, Metz, and ToxCast)

Name	Description	Dim
Atom type	[H, C, N, O, F, Cl, S, Br, I] (one-hot)	9
Atomic Num.	The atomic number (integer)	1
Acceptor	Accepts electrons [0/1] (binary)	1
Donor	Donates electrons [0/1](binary)	1
Aromatic	In an aromatic system [0/1](binary)	1
Hybridization	[sp, sp2, sp3] (one hot)	3
Hydrogens	Number of connected hydrogens (integer)	1
Formal charge	Formal charge of the atom (integer)	1
Explicit valence	Explicit valence of the atom (integer)	1
Implicit valence	Implicit valence of the atom (integer)	1
Explicit Hs.	Number of implicit Hs the atom is bound to (integer)	1
Radical electrons	Number of radical electrons for the atom (integer)	1

Table S2. The features used in classification tasks (Human and *C.elegans*)

Name	Description	Dim
Atom type	[C, N, O, S, F, Si, P, Cl, Br, Mg, Na, Ca, Fe, As, Al, I, B, V, K, Tl, Yb,Sb, Sn, Ag, Pd, Co, Se, Ti, Zn, H,Li, Ge, Cu, Au, Ni, Cd, In, Mn, Zr, Cr, Pt, Hg, Pb, other][a] (one-hot)	44
Degree	Number of covalent bonds [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] (one-hot)	11
Hydrogens	Number of connected hydrogens [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] (one-hot)	11
Implicit valence	Implicit valence of the atom [[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] (one-hot)]	11
Hybridization	[sp, sp2, sp3, sp3d, sp3d2, other] (one-hot)	6
Aromatic	Whether the atom is part of an aromatic system [0/1] (binary)	1
Chirality	Whether the atom is a chiral center [0/1] (binary)	1
Chirality type	[R, S] (one-hot)	2

[a] The Human and *C.elegans* datasets contained compounds with various types of atoms.

Table S3. Search range and selected values of hyperparameters for MGraphDTA

Hyperparameter	Search range	Selected value
Number of multiscale blocks	[1, 2, 3, 4, 5]	3
Number of graph convolutional layers in each multiscale block	[2, 3, 4, 5, 6, 7, 8, 9, 10]	8
The hidden channel number of each graph convolutional layer	[32, 64, 96, 128, 160]	64
The brach number of MCNN	[2, 3, 4, 5]	3

S3. Supplemental Figures

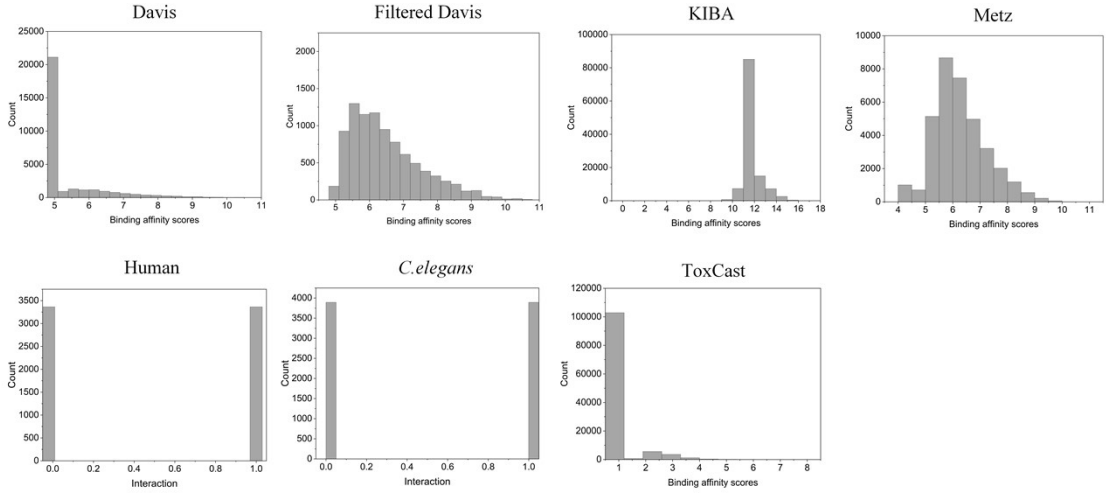

Figure S1. Distribution of the binding affinities (labels) in the Davis, filtered Davis, KIBA, Metz, Human, *C.elegans*, and ToxCast datasets used in our experiments.

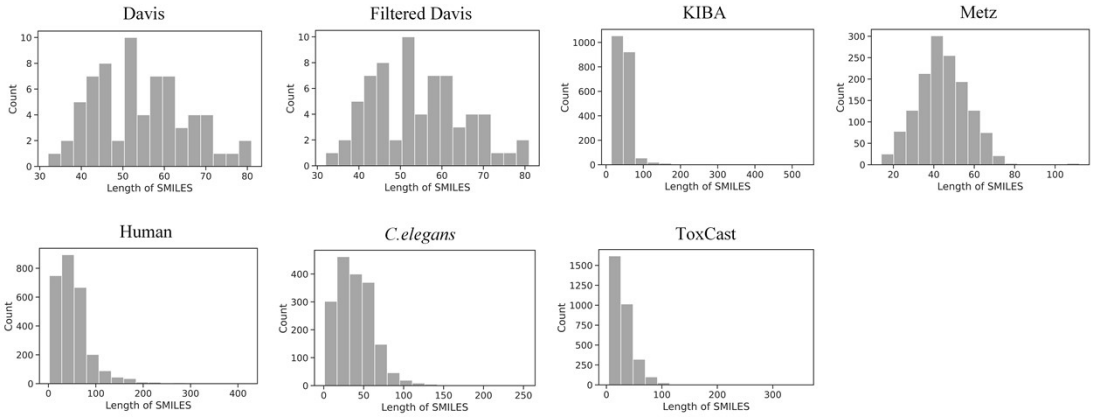

Figure S2. Distribution of the lengths of the SMILES strings in the Davis, filtered Davis, KIBA, Metz, Human, *C.elegans*, and ToxCast datasets

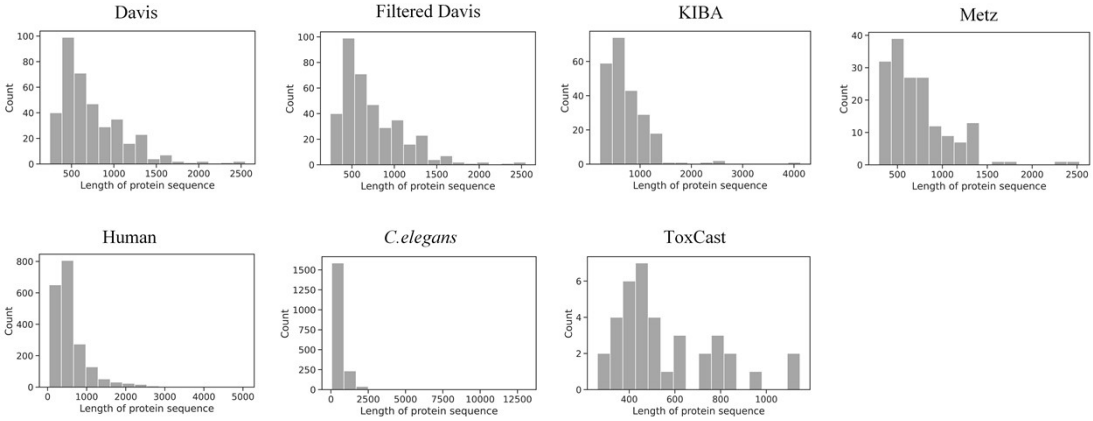

Figure S3. Distribution of the lengths of the protein sequences in the Davis, filtered Davis, KIBA,

Metz, Human, *C.elegans*, and ToxCast datasets

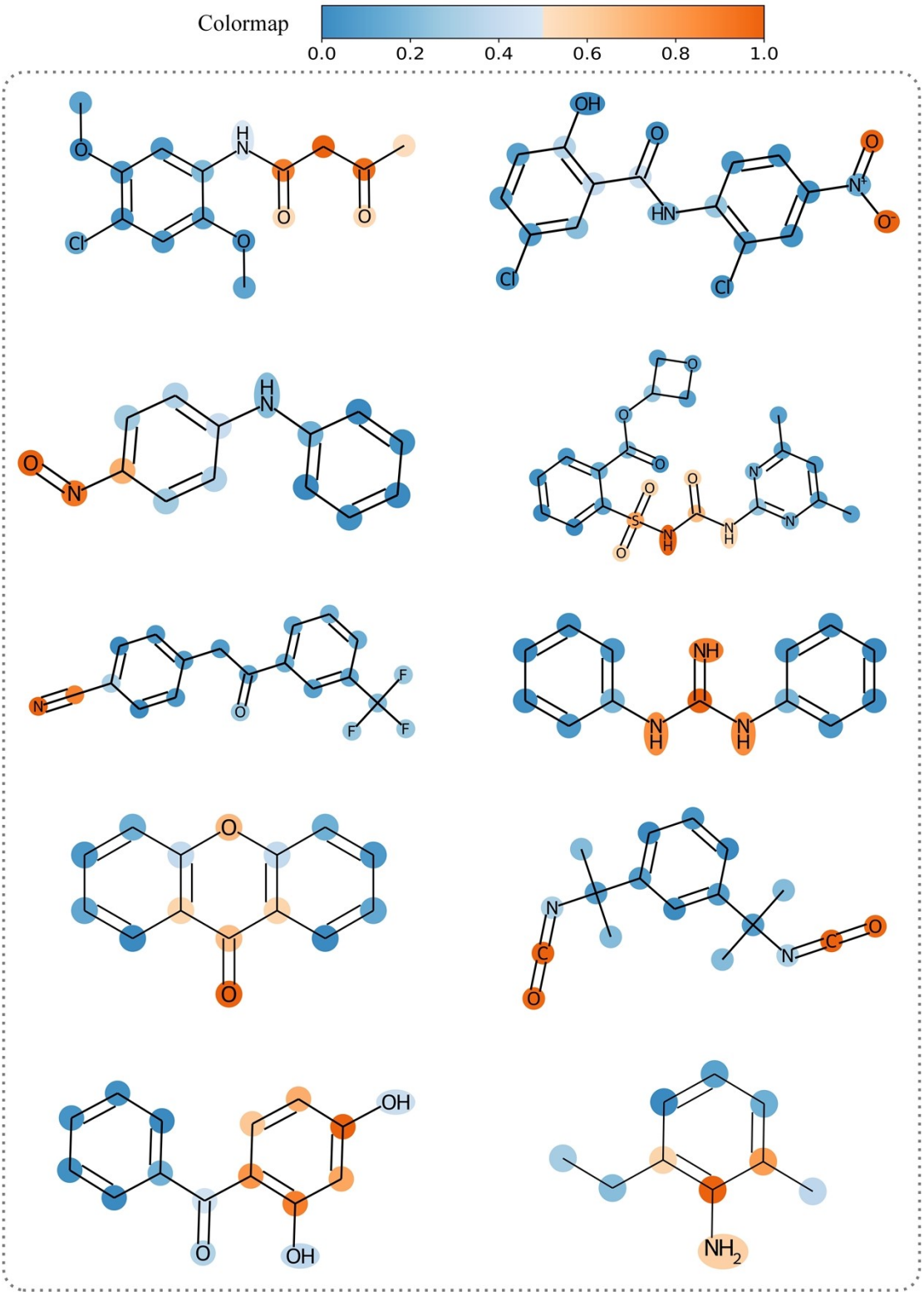

Figure S4. Additional examples of Grad-AAM (MGNN).

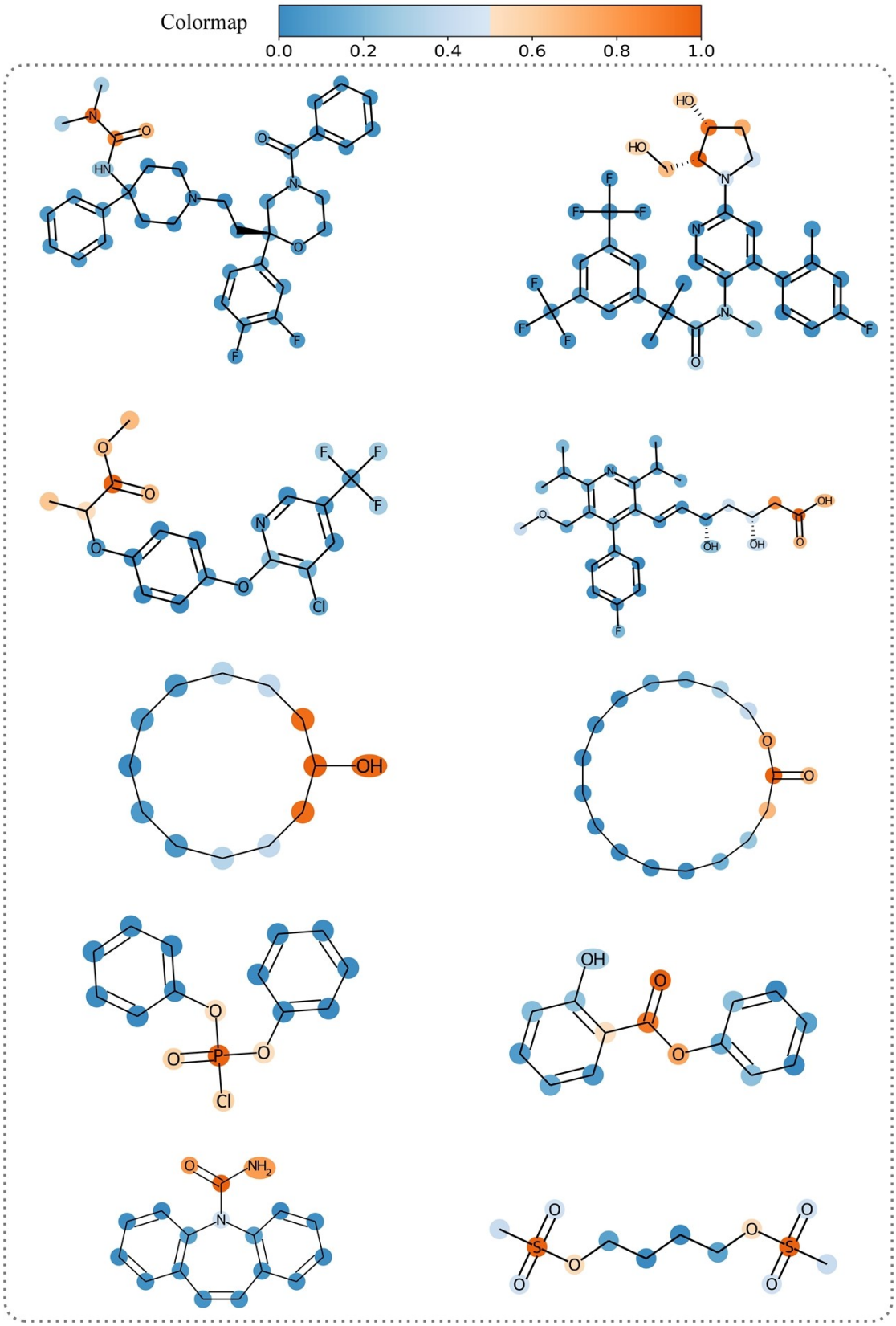

Figure S5. Additional examples of Grad-AAM (MGNN).