Electronic Supplementary Material (ESI) for RSC Advances.

iEdgeDTA: integrated edge information and 1D graph convolutional neural networks for binding affinity prediction

Supplementary information

Table S1

Detailed drug feature representation

Name	Definition	Size	Range of value
Node feature (66)			
Atom symbol (char)	Atomic symbol	18	["H", "B", "C", "N", "O", "F", "Mg", "Si", "P", "S", "Cl", "Cu", "Zn", "Se", "Br", "Sn", "I", and Unknown]
Atomic chiral tag (type)	Type of chirality	4	[CHI_UNSPECIFIED (chirality that hasn't been specified), CHI_TETRAHEDRAL_CW (tetrahedral: clockwise rotation (SMILES @@)), CHI_TETRAHEDRAL_CCW (tetrahedral: counter-clockwise rotation), CHI_OTHER (unrecognized type of chirality)]
Degree of atom (integer)	Number of directly-bonded neighbors in the molecule including Hs	8	from 1 to 7, and unknown if degree > 7
Number of formal charges (integer)	Number of formal charges in the molecule	11	from -5 to 5
Number of explicit and implicit Hs (integer)	Total number of Hs (implicit and explicit) that this atom is bound to	7	from 0 to 6
Number of radical electron (integer)	Number of unpaired electrons for an atom (radical)	8	from 0 to 7
Atom hybridization (type)	Orbital hybridisation, introduced to explain molecular structure	8	[UNSPECIFIED, S, SP, SP2, SP3, SP3D, SP3D2, OTHER]

Is aromatic (bool)	Whether the atom is aromatic (molecule has cyclic; ring-shaped structures)	1	1 if true, 0 if false
Is in ring (bool)	Whether the atom is in an aromatic ring	1	1 if true, 0 if false
Edge Feature (18)			
Bond type (type)	A type of bond between atoms	4	[SINGLE, DOUBLE, TRIPLE, AROMATIC]
Bond direction (type)	The direction of the bond	7	[NONE, BEGIN_WEDGE, BEGIN_DASH, END_DOWN_RIGHT, END_UP_RIGHT, EITHER_DOUBLE, UNKNOWN]
Bond stereo configuration (type)	The stereo configuration of the bond	6	[STEREONONE,STEREOANY, STEREOZ, STEREOE, STEREOCIS, STEREOTRANS]
Is bond conjugated (bool)	Whether the bond is considered to be conjugated	1	1 if true, 0 if false

Table S2

Performance of iEdgeDTA evaluated in the same setting as GraphDTA, DGraphDTA, WGNN and MGraphDTA; all results of GraphDTA, DGraphDTA, WGNN and MGraphDTA were obtained from their original publications.

Dataset	Model	MSE	CI
Davis	GraphDTA	0.229	0.893
	DGraphDTA	0.202	0.904
	WGNN	0.208	0.898
	MGraphDTA	0.207	0.900
	Our (iEdgeDTA)	**0.195**	**0.906**
KIBA	GraphDTA	0.139	0.891
	DGraphDTA	0.126	**0.904**
	WGNN	0.130	0.900
	MGraphDTA	0.128	0.902
	Our (iEdgeDTA)	**0.124**	0.897

Figure S1

Dataset distribution

Davis: Around 25% (109/442) of the proteins contain more than 1024 amino acids.

KIBA: Around 20% (42/229) of the proteins contain more than 1024 amino acids.

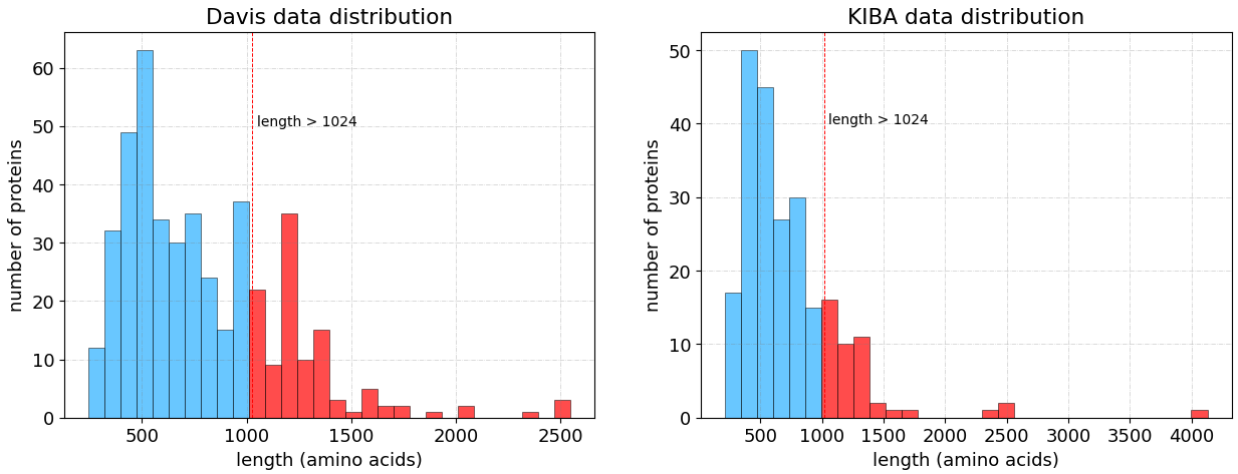

Figure S2

Cross-validation and testing; the model that performed best in the cross-validation was selected, and the weight of each cross-validation fold of the selected model was used to evaluate the performance with the independence test set.

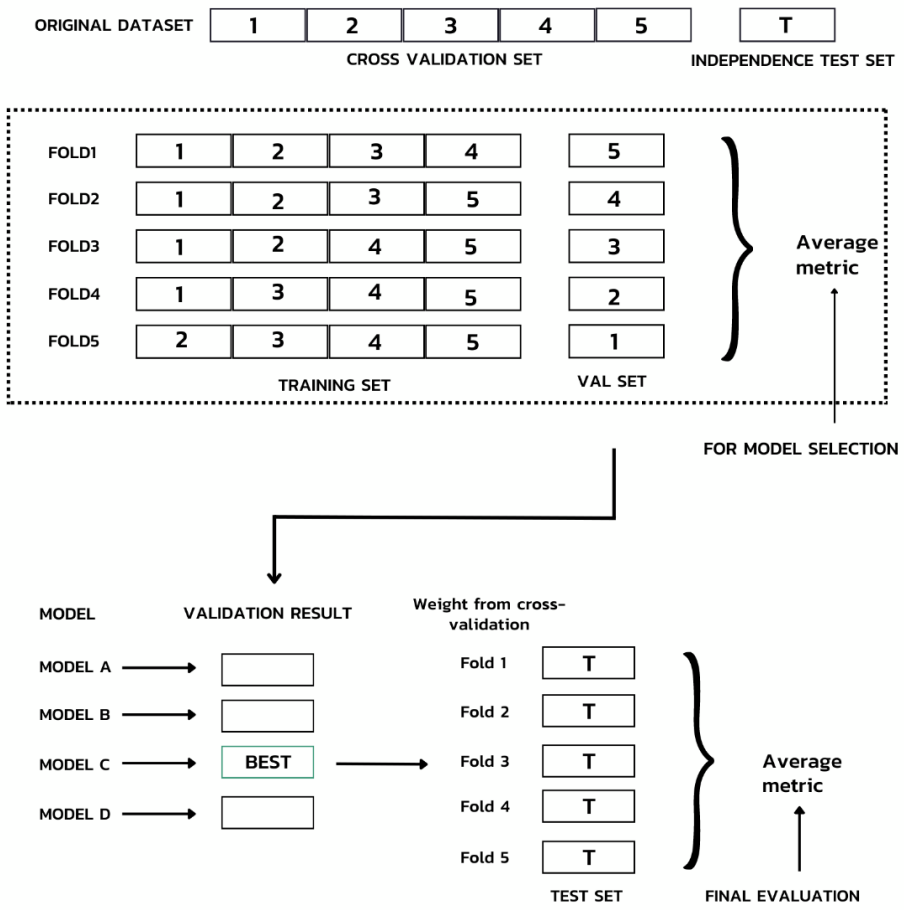

Figure S3

Performance of various dropout probabilities on Davis's five-fold cross validation set, where the error bar indicates the standard deviation.
MSE : Mean squared error (lower better)
CI: Concordance index (higher better)