Deep Learning Applications in Drug Development: Graph Attention Mechanism

February 18, 2022

Presenter: Haudi Ghiassi-Nejad

CS 846 – Deep Learning for Biotechnology
Cheriton School of Computer Science

UNIVERSITY OF WATERLOO

Outline

- Introduction
- Short Survey
- Focused Paper: Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism
- Conclusion
- Future Work
- Questions

Introduction

Drug development is:

- Time consuming

  It takes more than 10 years for a drug to reach market

- Costly

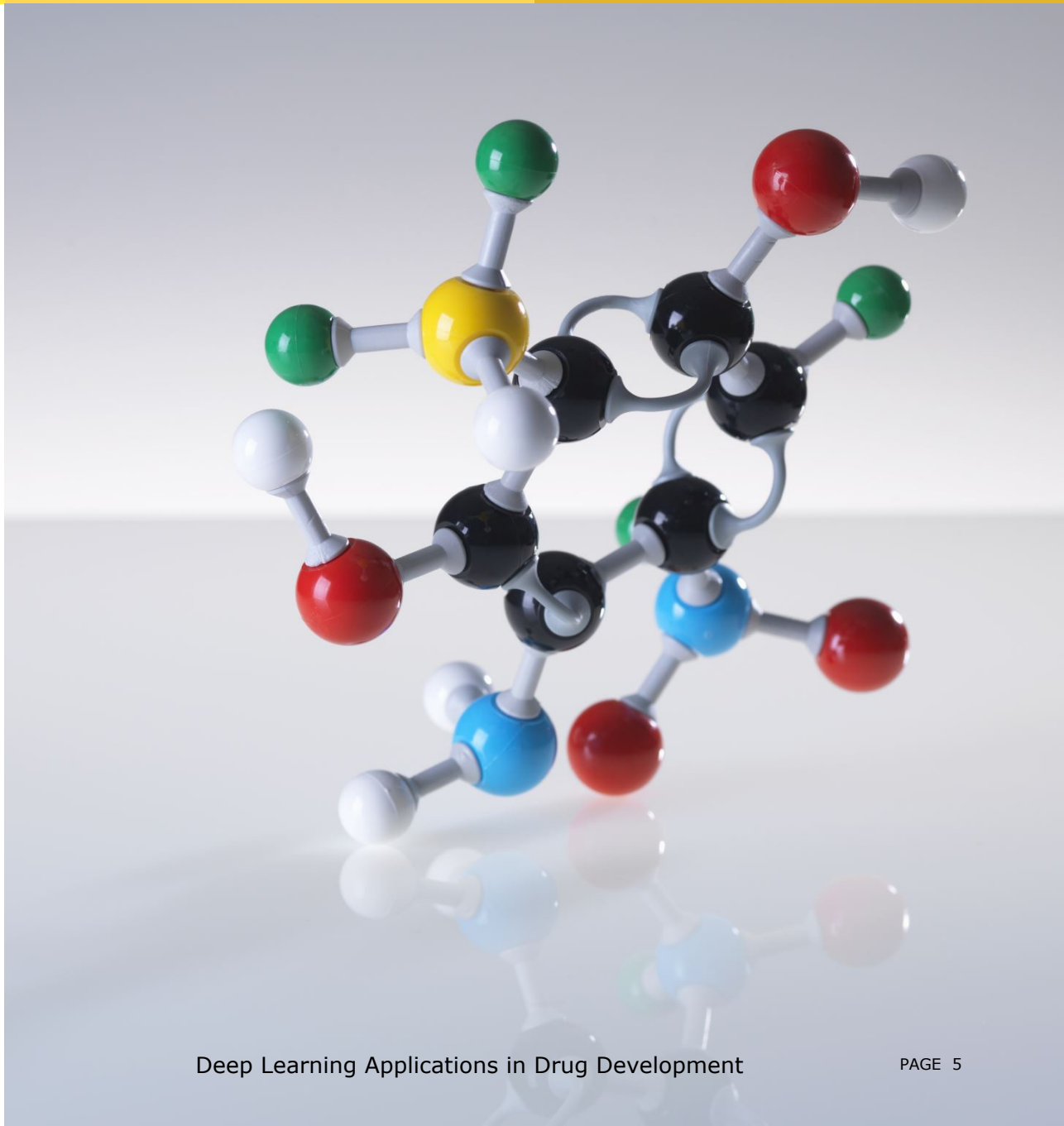
  The total costs can surpass 1 billion dollars per drug

Clinical Phases

- Preclinical Phase
  - Find the original drug candidate
  - Initial In-vivo and In-vitro tests
  - Toxicity and efficacy tests
- Phase 1: Tests on small group of humans mainly to detect adverse effects
- Phase 2: Efficacy in medium size groups
- Phase 3: Efficacy in larger groups of patients

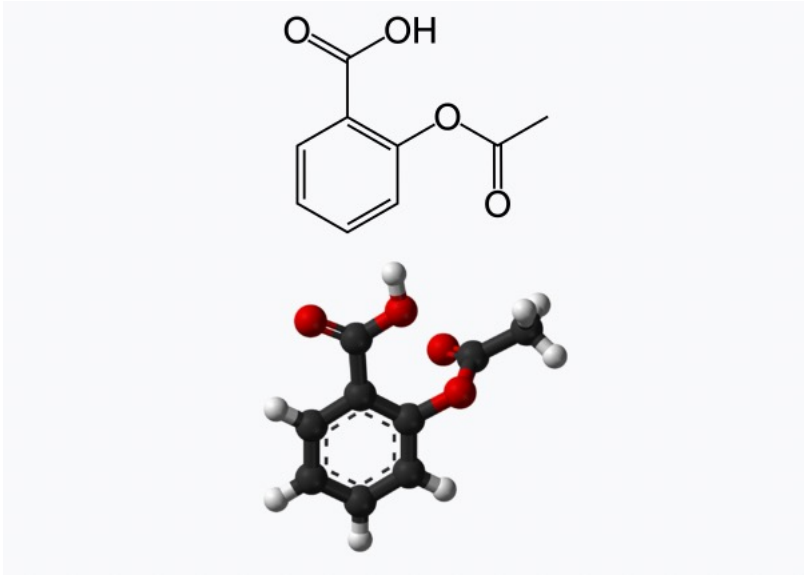
Small Molecules

Within the fields of molecular biology and pharmacology, a small molecule is a low molecular weight (< 900 Daltons) organic compound that may regulate a biological process, with a size on the order of 1 nm

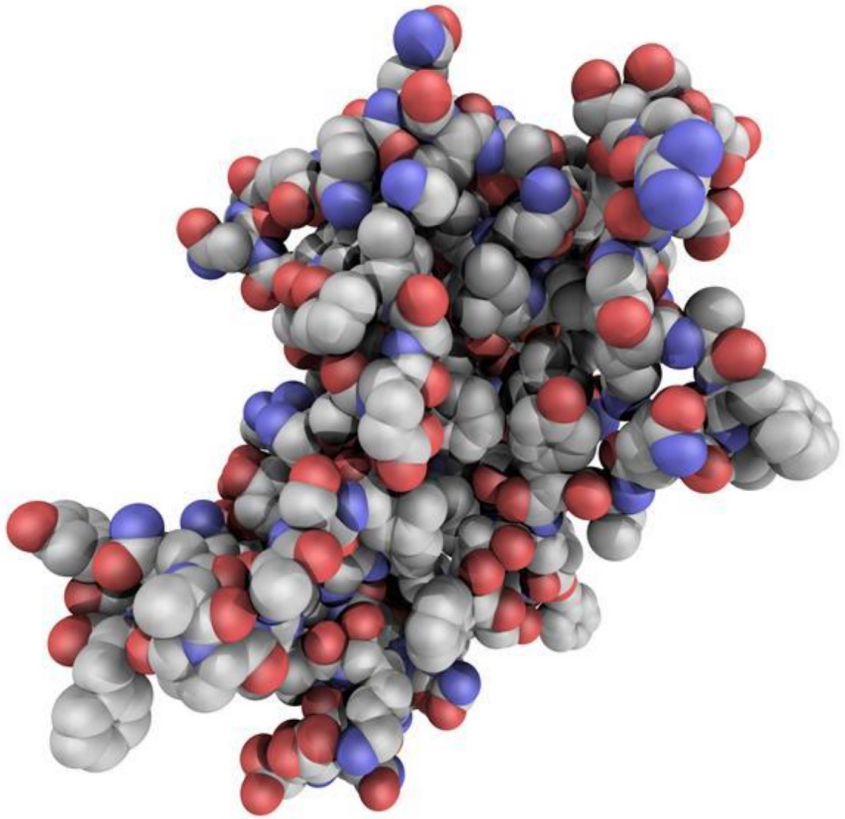


UNIVERSITY OF WATERLOO

Small Molecules vs Large Molecules



Aspirin



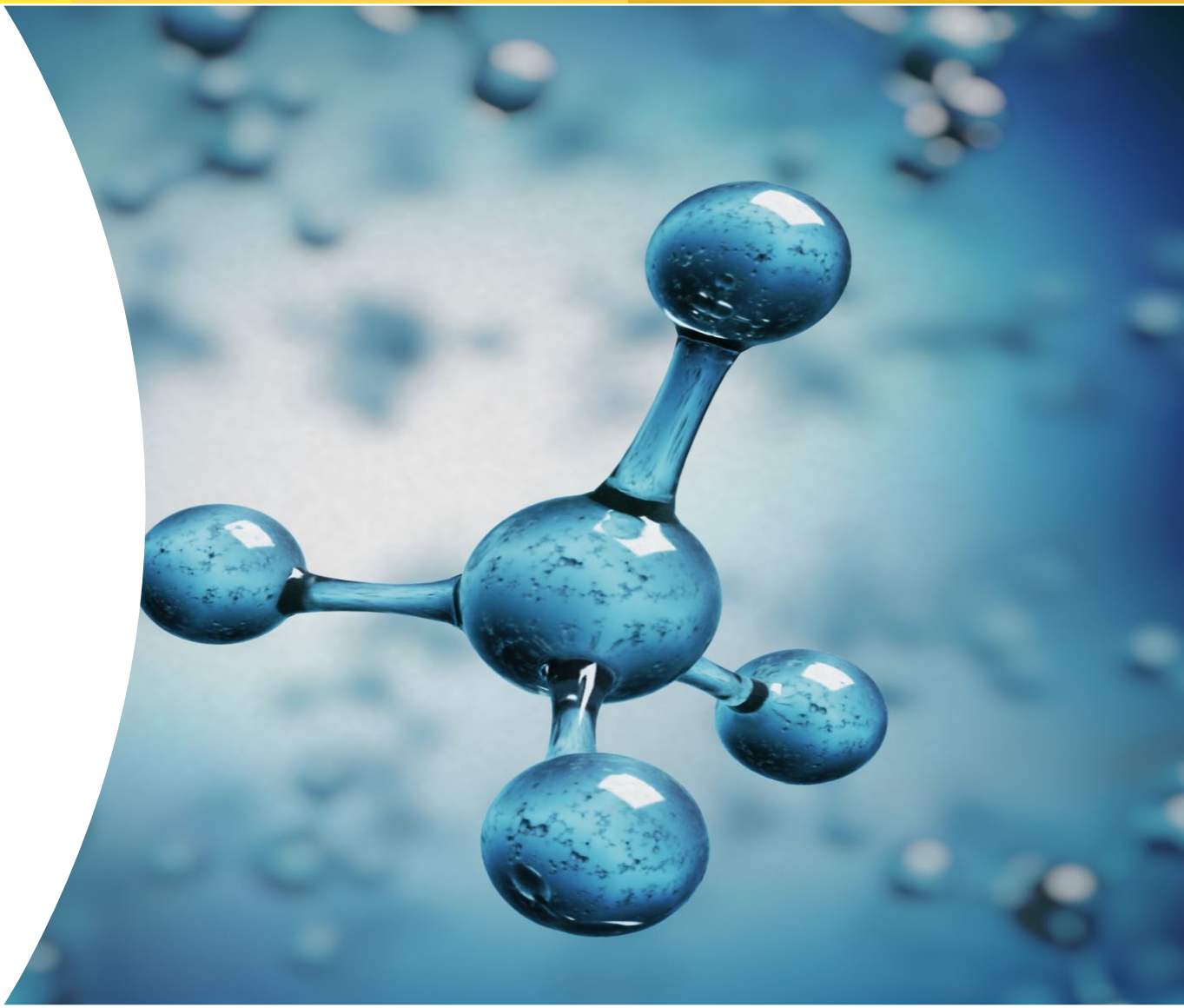
Human Growth Hormone

Problem Complexity: World of Chemical Compounds

- Theoretically, there are 10^{60} molecules with drug like capabilities

- It is impossible for human to go through all of them

- Computer programs are needed!



Short Survey I

Machine Learning Algorithms

Predict the physical and chemical properties of a molecule

Support vector machine (SVM): One of the oldest and most popular ones
Accuracy is reasonably good

Extreme gradient boosting (XGBoost): One of the newest ensemble learning algorithms, more optimization compared to traditional gradient boosting

Random forest (RF): Creates a strong classifier or regressor by an ensemble of individual decision trees

Deep neural networks (DNN)

Short Survey II

Graph Neural Networks - Mostly after 2016

- **Message Passing Neural Networks (MPNN):**
  - Message Passing phase: transmits information across the molecular graph to learn a molecular embedding
  - Readout phase: Computes a feature vector for the whole molecular graph
- **Graph Convolutional Neural Networks (GCN)**
  - Aggregate neighbors' information to produce an intermediate representation
  - transform the aggregated representations with linear projection and followed by a non-linear projection
- **Graph Attention Networks (GAT):** Similar to GCN but weighted aggragation
- **Attentive FP (We will cover this today)**

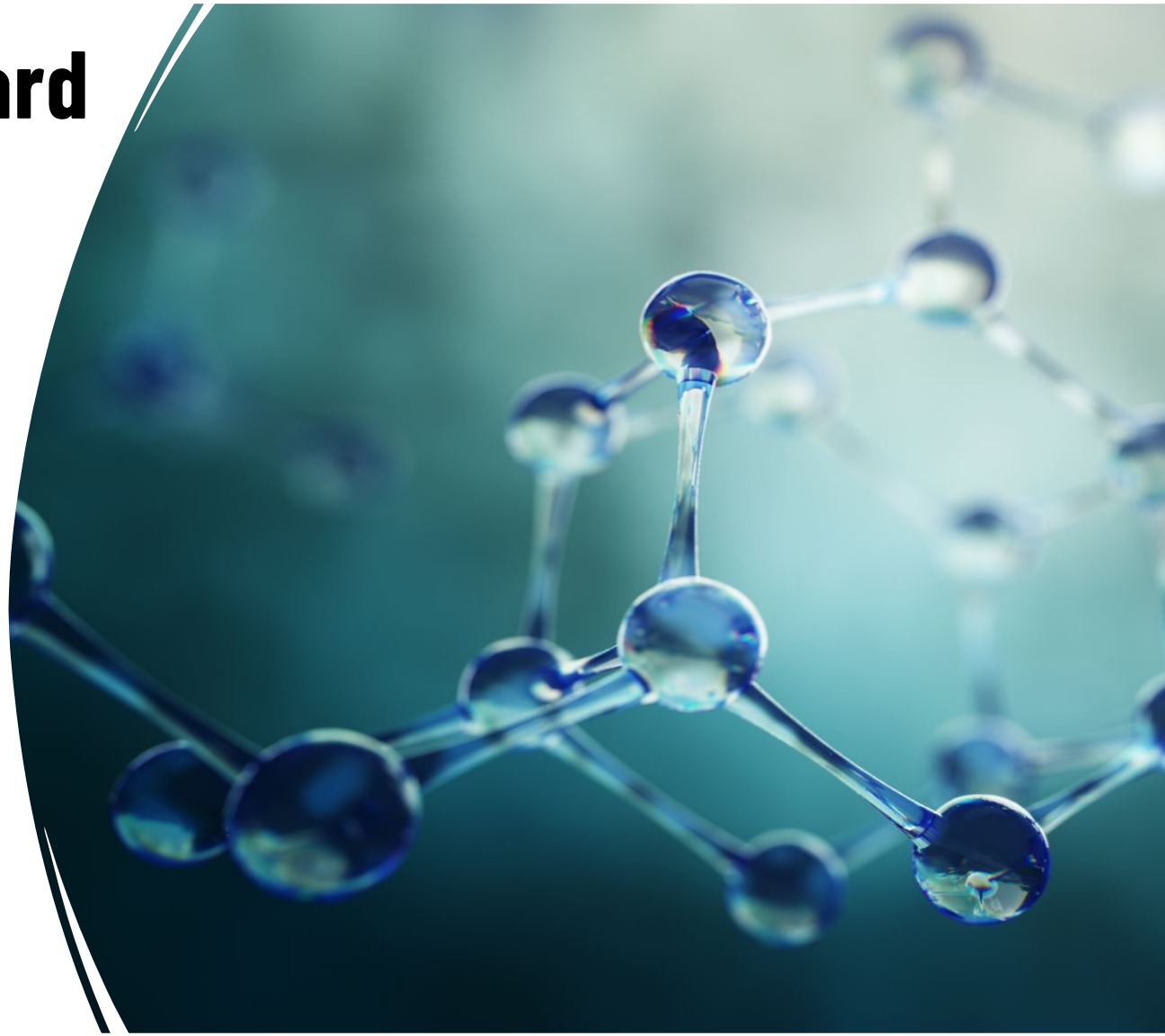
Paper: Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism

By Xiong et al.

UNIVERSITY OF WATERLOO

Molecular Understanding is Hard

- A molecular structure is usually composed of many-body interactions and complex electronic configurations
- Chemists have been working “at the edge of chaos” and in desperate need of augmented intelligence from AI.



Graph based vs Geometry based

Graph Based

- only the information concerning the topological arrangement of atoms

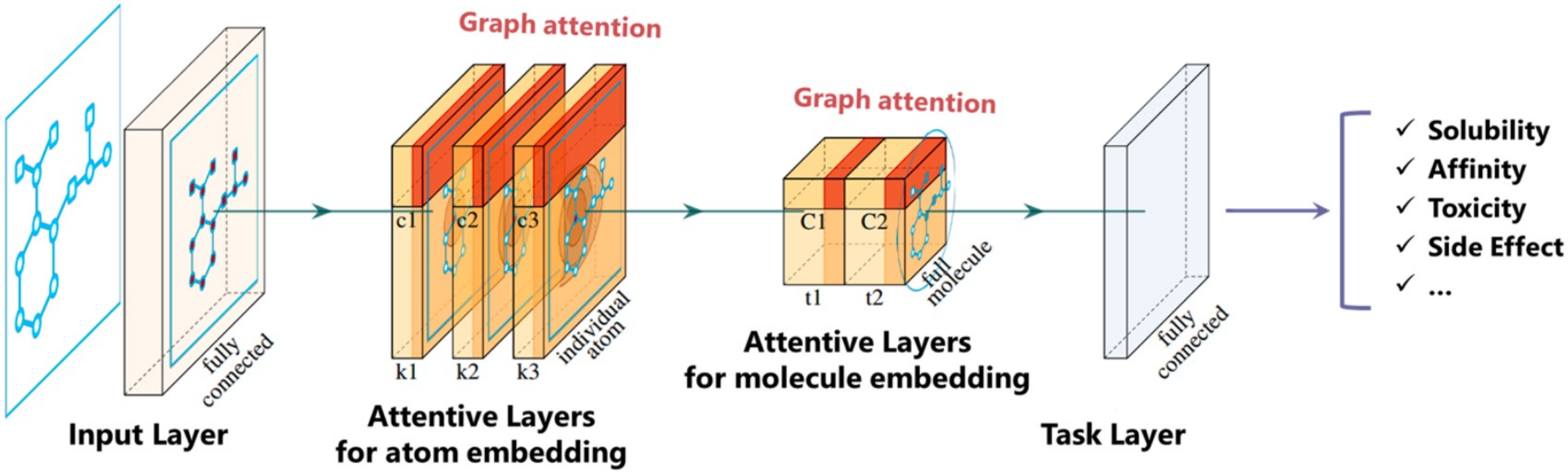
Geometry Based

- Molecular geometry information including:
  - bond length, bond angles, torsional angles
- Higher calculation costs
- Uncertainty in conformation of atoms: active conformation of a small molecule in a given binding process is usually unknown

Initial Atomic and Bond Features

atom feature	size	description
atom symbol	16	[B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, I, At, metal] (one-hot)
degree	6	number of covalent bonds [0,1,2,3,4,5] (one-hot)
formal charge	1	electrical charge (integer)
radical electrons	1	number of radical electrons (integer)
hybridization	6	[sp, sp^2, sp^3, sp^3d, sp^3d^2, other] (one-hot)
aromaticity	1	whether the atom is part of an aromatic system [0/1] (one-hot)
hydrogens	5	number of connected hydrogens [0,1,2,3,4] (one-hot)
chirality	1	whether the atom is chiral center [0/1] (one-hot)
chirality type	2	[R, S] (one-hot)
bond feature	**size**	**description**
bond type	4	[single, double, triple, aromatic] (one-hot)
conjugation	1	whether the bond is conjugated [0/1] (one-hot)
ring	1	whether the bond is in ring [0/1] (one-hot)
stereo	4	[StereoNone, StereoAny, StereoZ, StereoE] (one-hot)

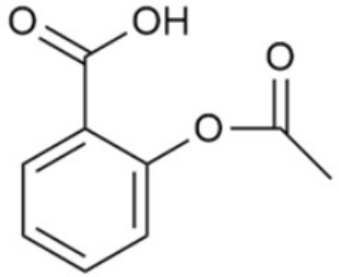
Network Structure



UNIVERSITY OF WATERLOO

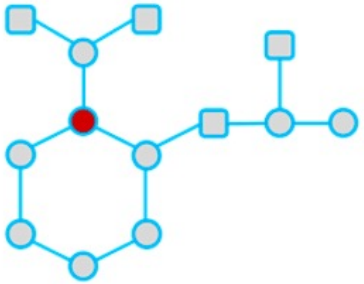
Locality Comparison

- Neural FP & GCN: Highly local
- Weave & MPNN: Averaged globally
- Attentive FP: Could be affected from any atom

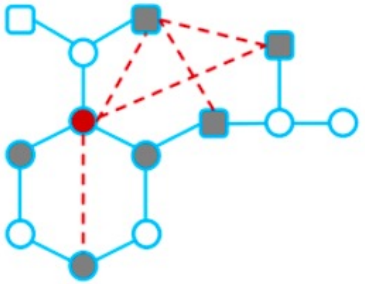


Aspirin structure

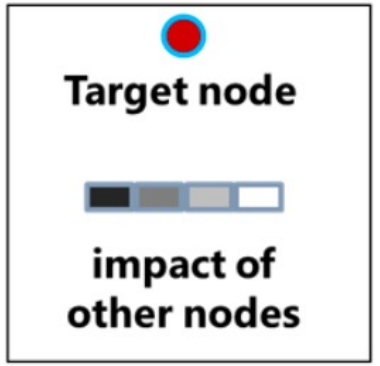
Neural FP & GCN



Weave & MPNN



Attentive FP



UNIVERSITY OF WATERLOO

Graph Neural Network

Messaging:

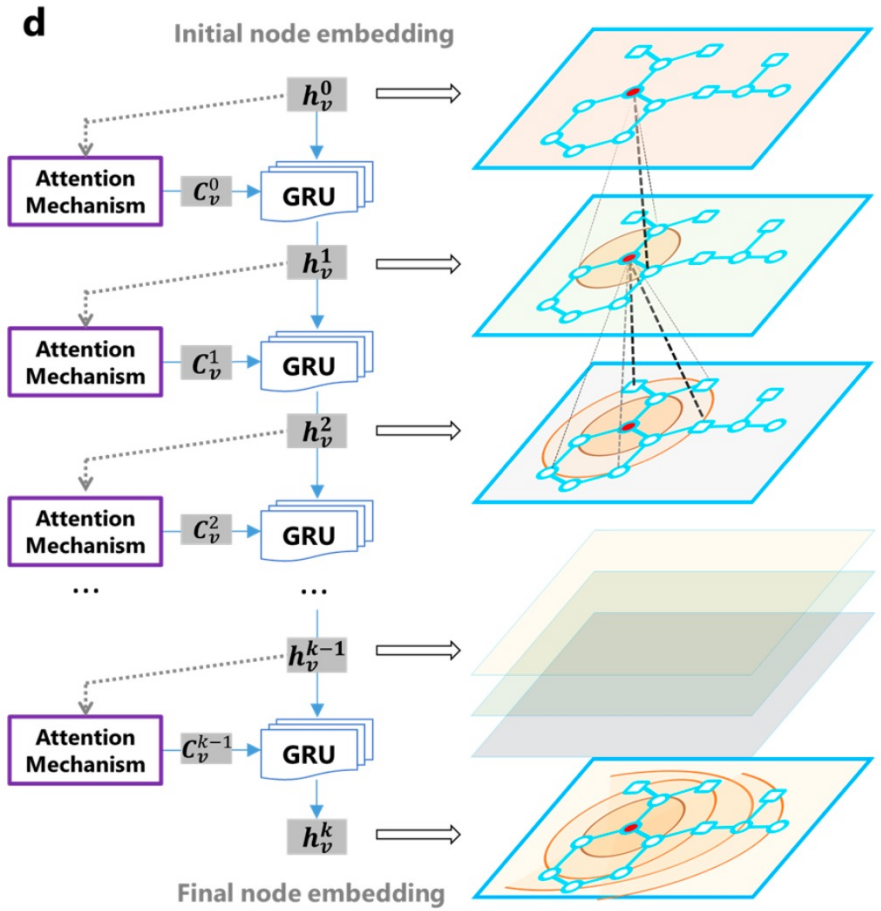
$$\mathbf{C}_v^{k-1} = \sum_{u \in N(v)} \boldsymbol{M}^{k-1}(\boldsymbol{h}_u^{k-1}, \boldsymbol{h}_v^{k-1})$$

Readout:

$$\boldsymbol{h}_v^k = \mathrm{GRU}^{k-1}(\mathbf{C}_v^{k-1}, \boldsymbol{h}_v^{k-1})$$

Generating the State Vector for Target Atom

- The framework that generates a state vector (embedding) for a target atom v
- h^k and C^k are the state vector and attention context vector at time step k for atom v, respectively
- In higher time steps, target node embedding will include information from further nodes recursively



Graph Attention Mechanism

Alignment:

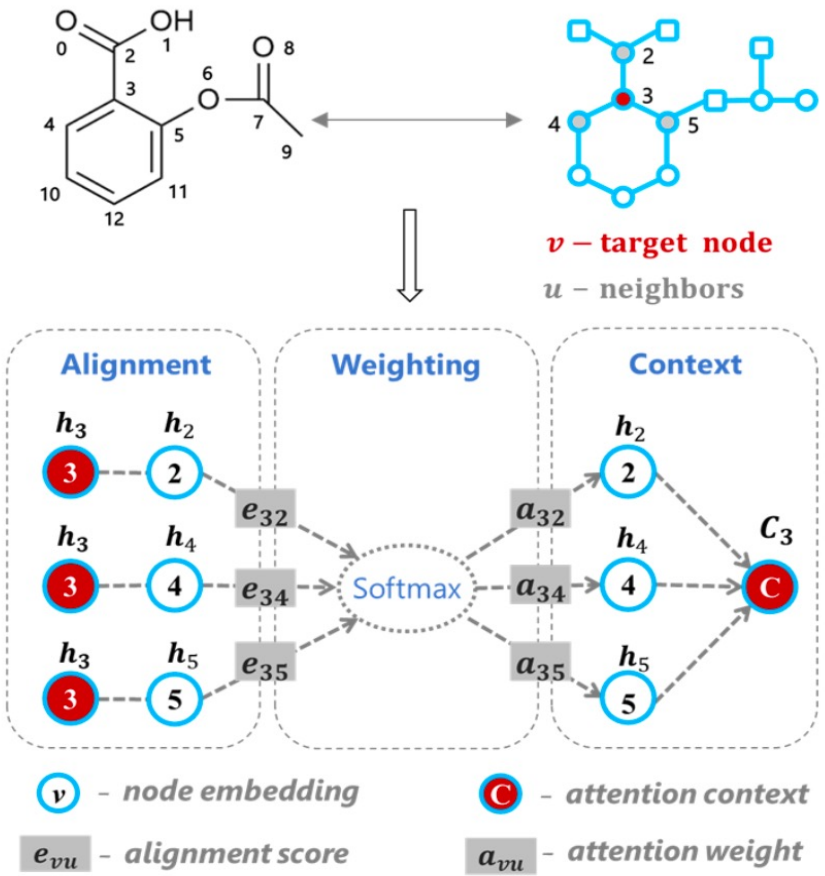
$$e_{vu} = leaky\_relu(\boldsymbol{W}\cdot[\boldsymbol{h}_v, \boldsymbol{h}_u]) \quad (1)$$

Weighting:

$$a_{vu} = softmax(e_{vu}) = \frac{\exp(e_{vu})}{\sum_{u\in N(v)} \exp(e_{vu})} \quad (2)$$

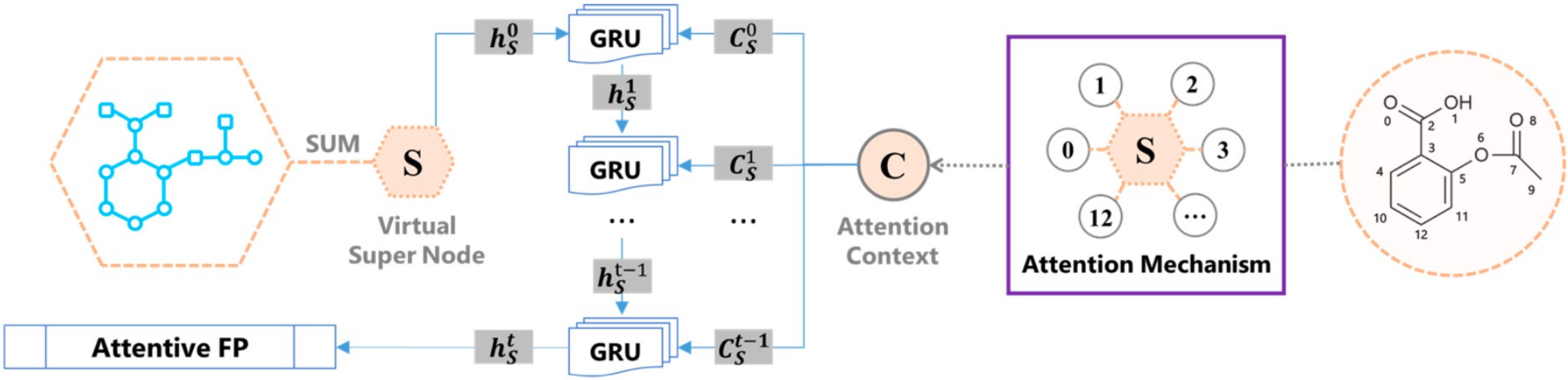
Context:

$$\mathbf{C}_v = elu\left(\sum_{u\in N(v)} a_{vu}\cdot\boldsymbol{W}\cdot\boldsymbol{h}_u\right) \quad (3)$$



Graph Attention Mechanism

Entire Molecular Graph Embedding



UNIVERSITY OF WATERLOO

Results I

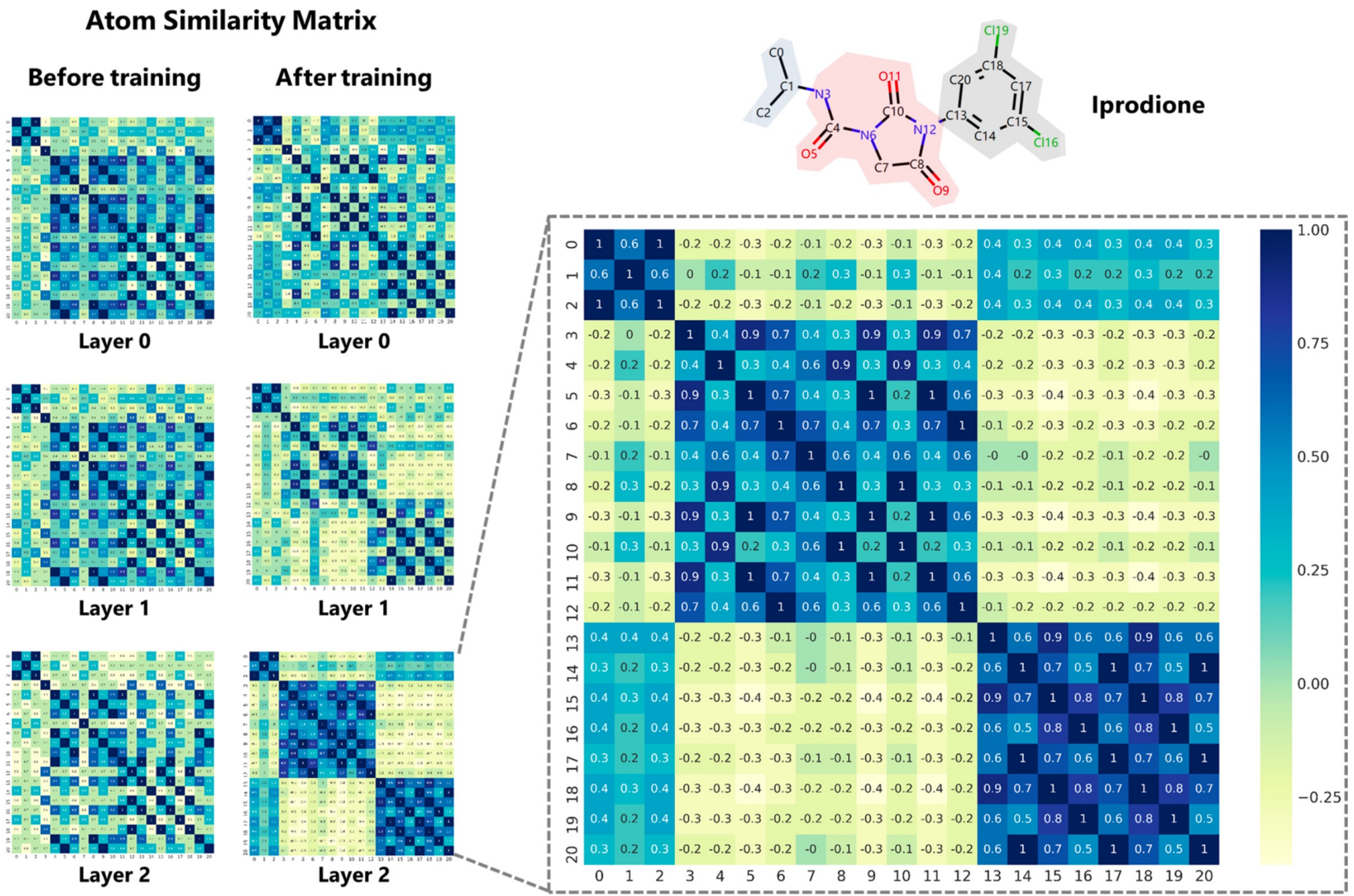
Table 3. Predictive Performances on Data Sets Relevant to Drug Discovery

category	data sets	no. of compounds	metrics	splitting	*previous best	Attentive FP
physical chemistry	ESOL	1128	RMSE	random	MPNN: 0.58	**0.503**
	FreeSolv	643	RMSE	random	MPNN: 1.15	**0.736**
	Lipop	4200	RMSE	random	Neural FP: 0.655	**0.578**
bioactivity	MUV	93,127	PRC	random	MultiTask: 0.184	**0.221**
			ROC	random	GC: 0.775	**0.843**
	HIV	41,913	ROC	scaffold	SVM: 0.792	**0.832**
	BACE	1522	ROC	scaffold	RF: **0.867**	0.850
physiology or toxicity	BBBP	2053	ROC	scaffold	SVM: 0.729	**0.920**
	Tox21	8014	ROC	random	Neural FP: 0.829	**0.858**
	ToxCast	8615	ROC	random	Weave: 0.742	**0.805**
	SIDER	1427	ROC	random	RF: **0.684**	0.637
	ClinTox	1491	ROC	random	Weave: 0.832	**0.940**

Results II

Table 4. Predictive Performances on the qm9 Data Set Quantum Properties (MAE)

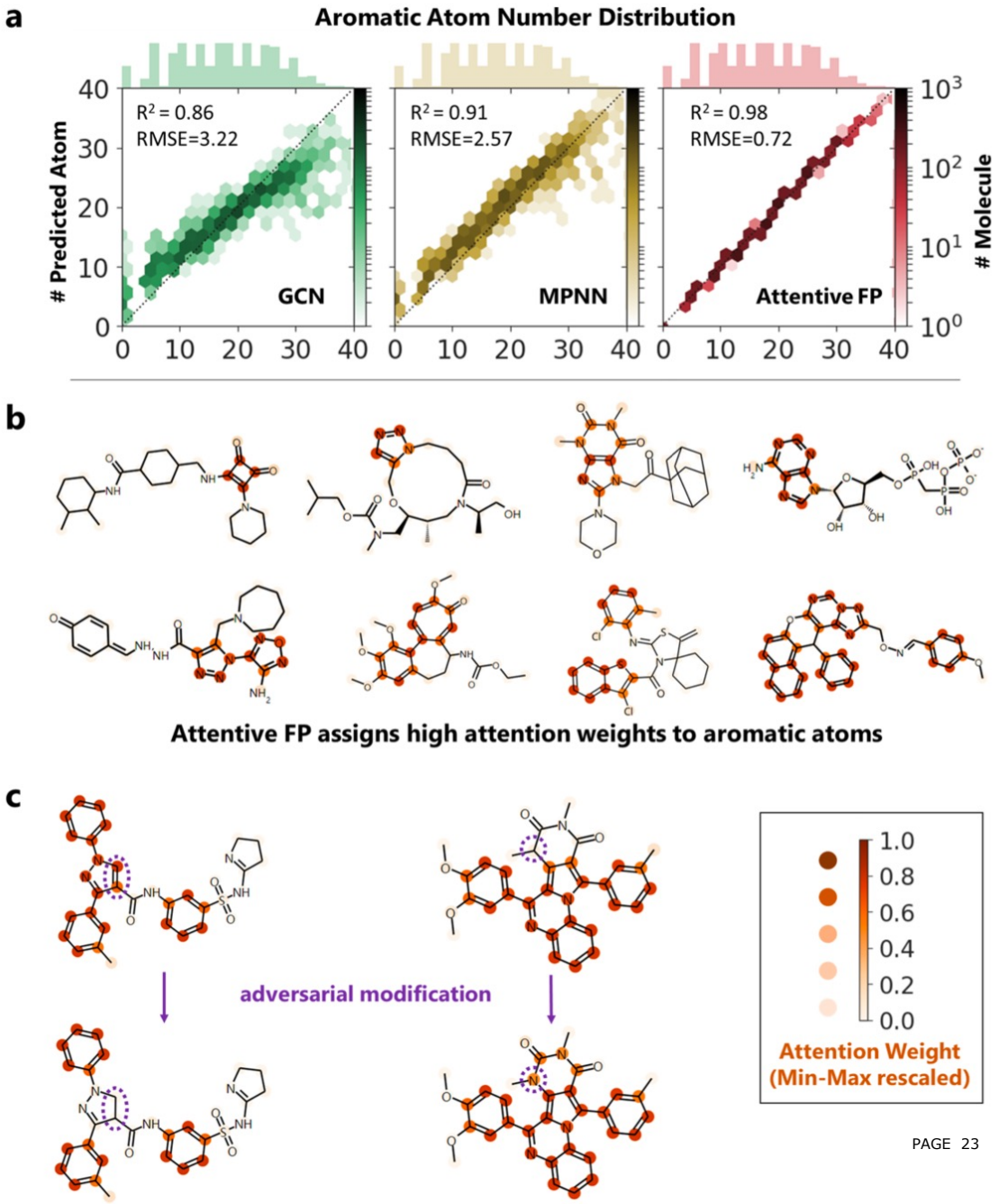
			geometry-based		graph-based			
task	unit	sample MAD	*CM	*DTNN	*ECFP	*GC	*MPNN	Attentive FP
mu	D	1.189	0.519	**0.244**	0.602	0.583	0.358	0.451
alpha	b^3	6.299	0.85	0.95	3.1	1.37	0.89	**0.492**
HOMO	hartree	0.016	0.00506	0.00388	0.0066	0.00716	0.00541	**0.00358**
LUMO	hartree	0.039	0.00645	0.00513	0.00854	0.00921	0.00623	**0.00415**
gap	hartree	0.040	0.0086	0.0066	0.01	0.0112	0.0082	**0.00528**
R2	b^2	202.017	46	**17**	125.7	35.9	28.5	26.839
ZPVE	hartree	0.026	0.00207	0.00172	0.01109	0.00299	0.00216	**0.00120**
U0	hartree	31.072	2.27	2.43	15.1	3.41	2.05	**0.898**
U	hartree	31.072	2.27	2.43	15.1	3.41	2	**0.893**
H	hartree	31.072	2.27	2.43	15.1	3.41	2.02	**0.893**
G	hartree	31.072	2.27	2.43	15.1	3.41	2.02	**0.893**
Cv	cal/mol/K	3.204	0.39	0.27	1.77	0.65	0.42	**0.252**



UNIVERSITY OF WATERLOO

Example: Aromatic Atoms Predictions

The example shows the superior performance of Attentive FP in identifying aromatic atoms



UNIVERSITY OF WATERLOO

Future Work

- The current models mostly predict the properties
- T is still not the real problem!
- There is a need for a model that can generate a suitable molecular structure directly

Questions

References

- Xiong, Z., Wang, D., Liu, X., Zhong, F., Wan, X., Li, X., ... & Zheng, M. (2019). Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry*, *63*(16), 8749-8760.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In *International conference on machine learning* (pp. 1263-1272). PMLR.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Mullard, A. (2017). The drug-maker's guide to the galaxy. *Nature*, *549*(7673), 445-447.
- Jiang, D., Wu, Z., Hsieh, C. Y., Chen, G., Liao, B., Wang, Z., ... & Hou, T. (2021). Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *Journal of cheminformatics*, *13*(1), 1-23.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- https://www.chemistryworld.com/podcasts/human-growth-hormone/7912.article