A Supervised Markovian Random Walk Model for Investigating Hepatotoxicity Signatures of Chemical Drugs with Structural Alerts

Sangsoo Lim,[1] Youngkuk Kim,[2,†] Jeonghyeon Gu,[3,†] and Sun Kim[1,2,3,4,5,*]

[1]Bioinformatics Institute, [2]Department of Computer Science and Engineering, [3]Interdisciplinary Program in Artificial Intelligence, [4]Institute of Engineering Research, Seoul National University, Seoul 08826, Korea, [5]AIGENDRUG, LTD.

[†]Equal contribution

*Corresponding Author: sunkim.bioinfo@snu.ac.kr

Introduction

Motivation

- Drug-Induced Liver Injury (DILI) is a major hurdle in drug development.
- Underlying mechanisms of DILI are mostly veiled despite great efforts of *in vivo* and *in vitro* experimental procedures of clinical trials
- Currently available in silico methods neither show high performance nor suggest important chemical substructures

Object

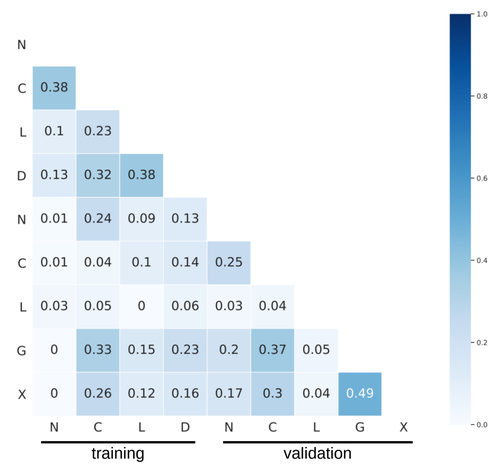
- To develop a highly accurate DILI prediction method.
- To characterize important structural alerts (SAs)

DILI Data Sets

DILI data

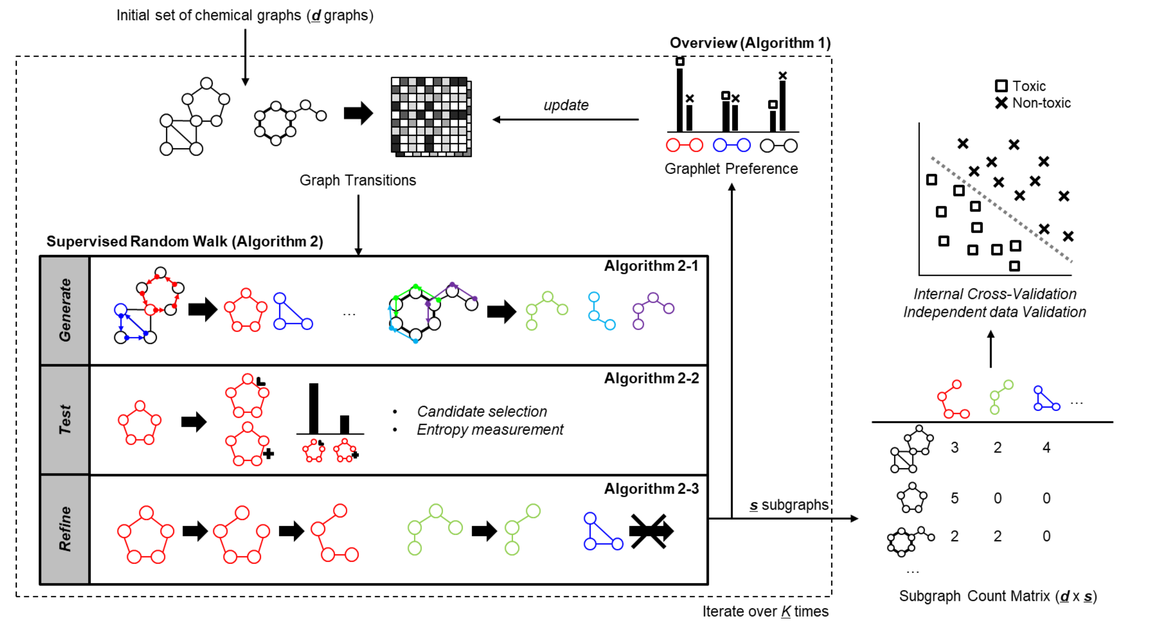
- Most data sets were retrieved from *JCIM*, 2015[1]
- DILIst data from US FDA was also used as training data[2]

Category	Name	Class labels		Total
		Toxic	Non-Toxic	
Training	NCTR	78	102	180
	Combined	232	228	460
	Liew	648	417	1,065
	DILIst	720	438	1,158
Validation	NCTR	97	87	181
	Combined	113	83	196
	Liew	70	49	119
	Greene	208	109	317
	Xu	127	106	233

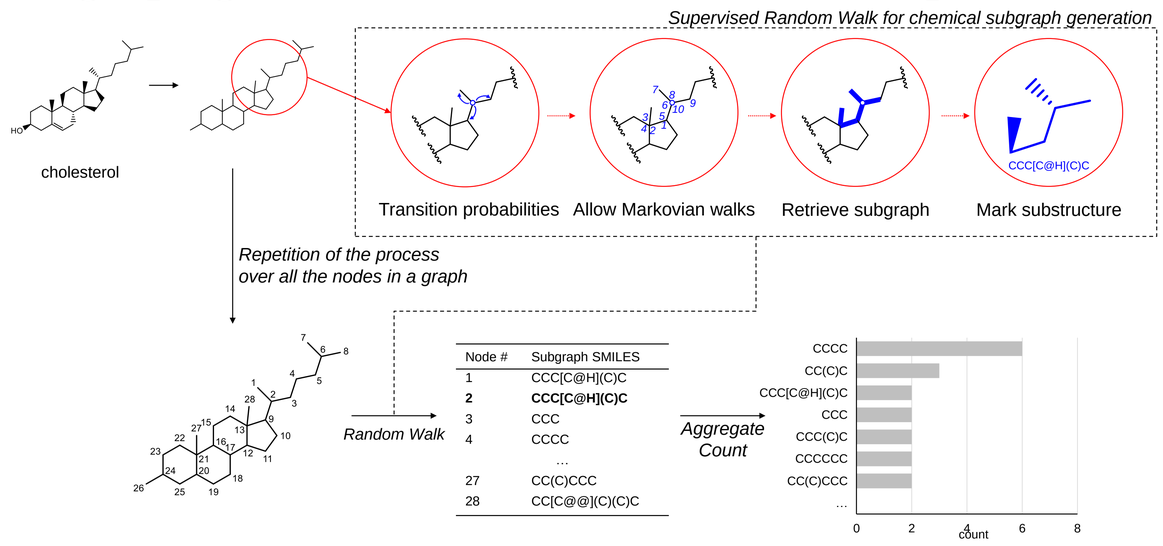


Methods

Overview of our approach



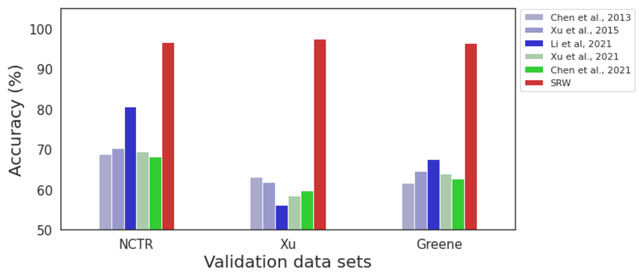
Subgraph generation: cholesterol as an example



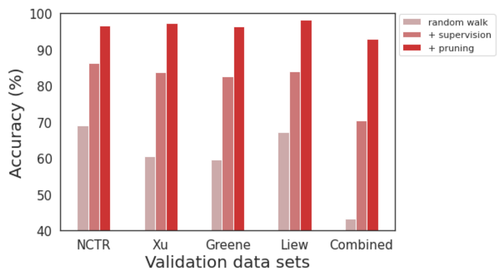
Results

1. Performance: significant improvement in accuracy

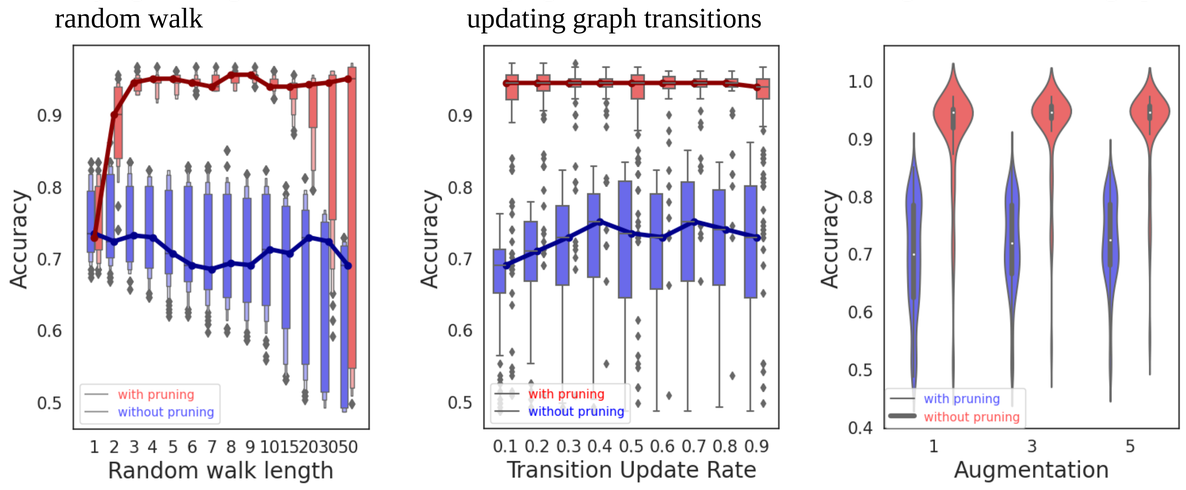
- Comparison to previous methods:
  - DILI prediction tools
  - State-of-the-art GNN methods



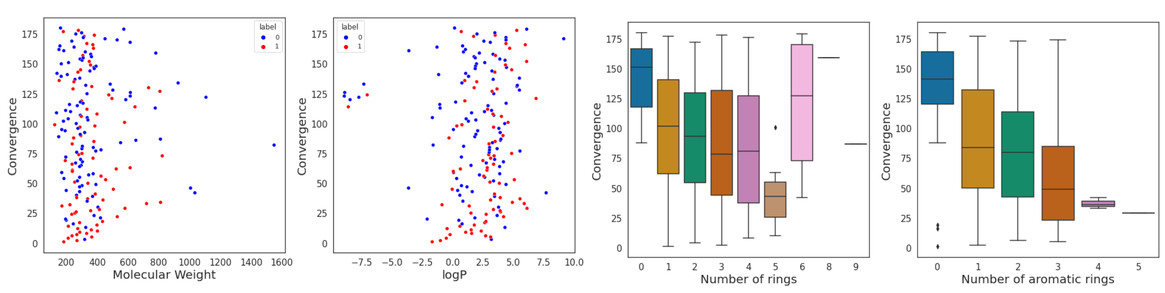
- Supervision and pruning highly improve the performance



- Varying the length of random walk
- Varying the rate of updating graph transitions
- Augmentation of subgraphs



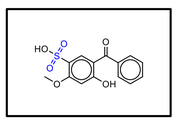
- Invariance of subgraphs to molecular properties



2. Identification of structural alerts (SAs) of DILI

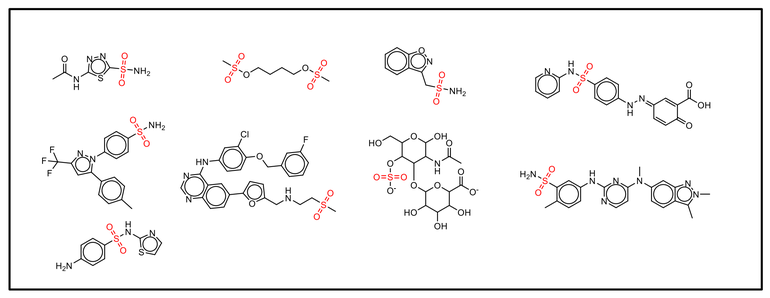
- Example SA ‘S(=O)=O’ over-represented in toxic drugs

non-toxic (1 / 102)



SMILES	S(=O)=O
F(NonTox)	1.0%
F(Tox)	11.5%
Entropy	0.396

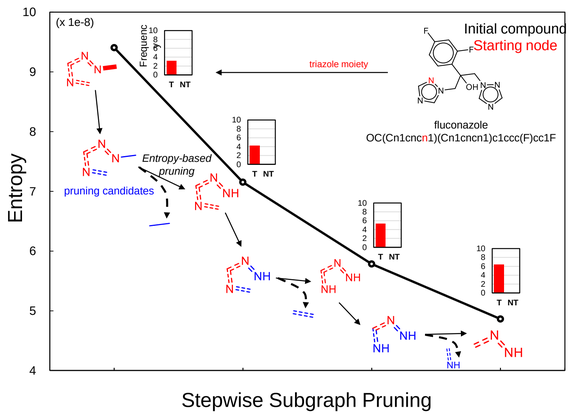
toxic (9 / 78)



- Top 10 SAs enriched to toxic drugs

Subgraph (SMILES)	Training data		Validation data	
	F(NT)	F(T)	F(NT)	F(T)
cNcc		0.090	0.011	0.064
cNc		0.090	0.011	0.064
NS=O		0.077	0.034	0.106
Cco		0.064	-	0.074
C=NN	-	0.051	-	-
cCS		0.038	-	0.021
nCO		0.038	0.034	0.085
cnCO		0.038	0.034	0.085
CNO		0.026	0.011	0.011
C#N		0.026	0.023	0.011

- Pruning subgraphs



References

1. Xu, Youjun, et al. "Deep learning for drug-induced liver injury." *Journal of chemical information and modeling* 55.10 (2015): 2085-2093.
2. Thakkar, Shraddha, et al. "Drug-induced liver injury severity and toxicity (DILIst): binary classification of 1279 drugs by human hepatotoxicity." *Drug discovery today* 25.1 (2020): 201-208.
3. Chen, Minjun, et al. "Quantitative structure-activity relationship models for predicting drug-induced liver injury based on FDA-approved drug labeling annotation and using a large collection of drugs." *toxicological sciences* 136.1 (2013): 242-249.
4. Xu, Minghao, et al. "Self-supervised Graph-level Representation Learning with Local and Global Structure." *arXiv preprint arXiv:2106.04113* (2021).
5. Chen, Fangying, Junyoung Park, and Jinkyoo Park. "A Hypergraph Convolutional Neural Network for Molecular Properties Prediction using Functional Group." *arXiv preprint arXiv:2106.01028* (2021).
6. Backstrom, Lars, and Jure Leskovec. "Supervised random walks: predicting and recommending links in social networks." *Proceedings of the fourth ACM international conference on Web search and data mining*. 2011.