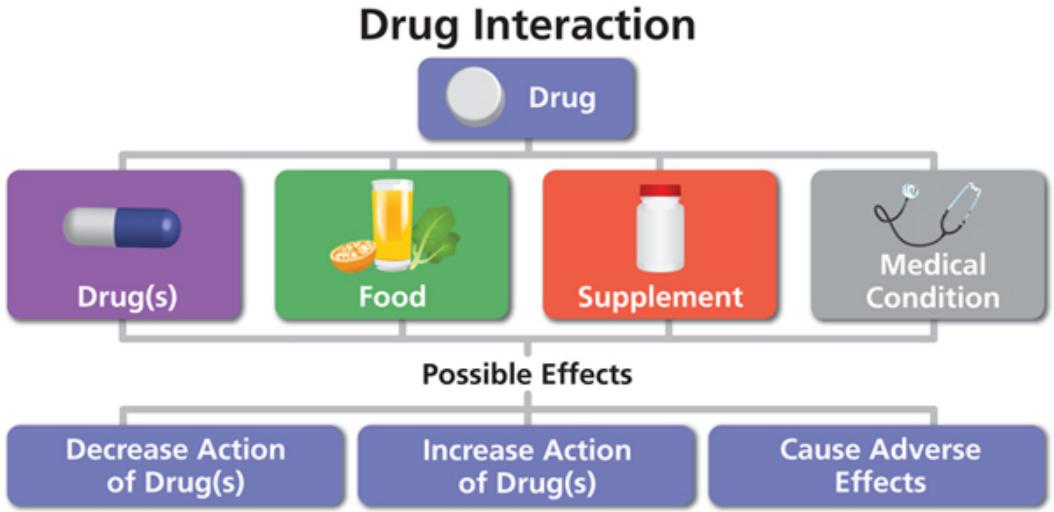
CASTER: Predicting Drug Interactions with Chemical Substructure Representation

Kexin Huang[1,2], Cao Xiao[2], Trong Nghia Hoang[3], Lucas M. Glass[2], Jimeng Sun[4,5]

[1]Harvard [2]IQVIA [3]MIT-IBM Watson AI Lab [4]UIUC CS [5]Georgia Tech

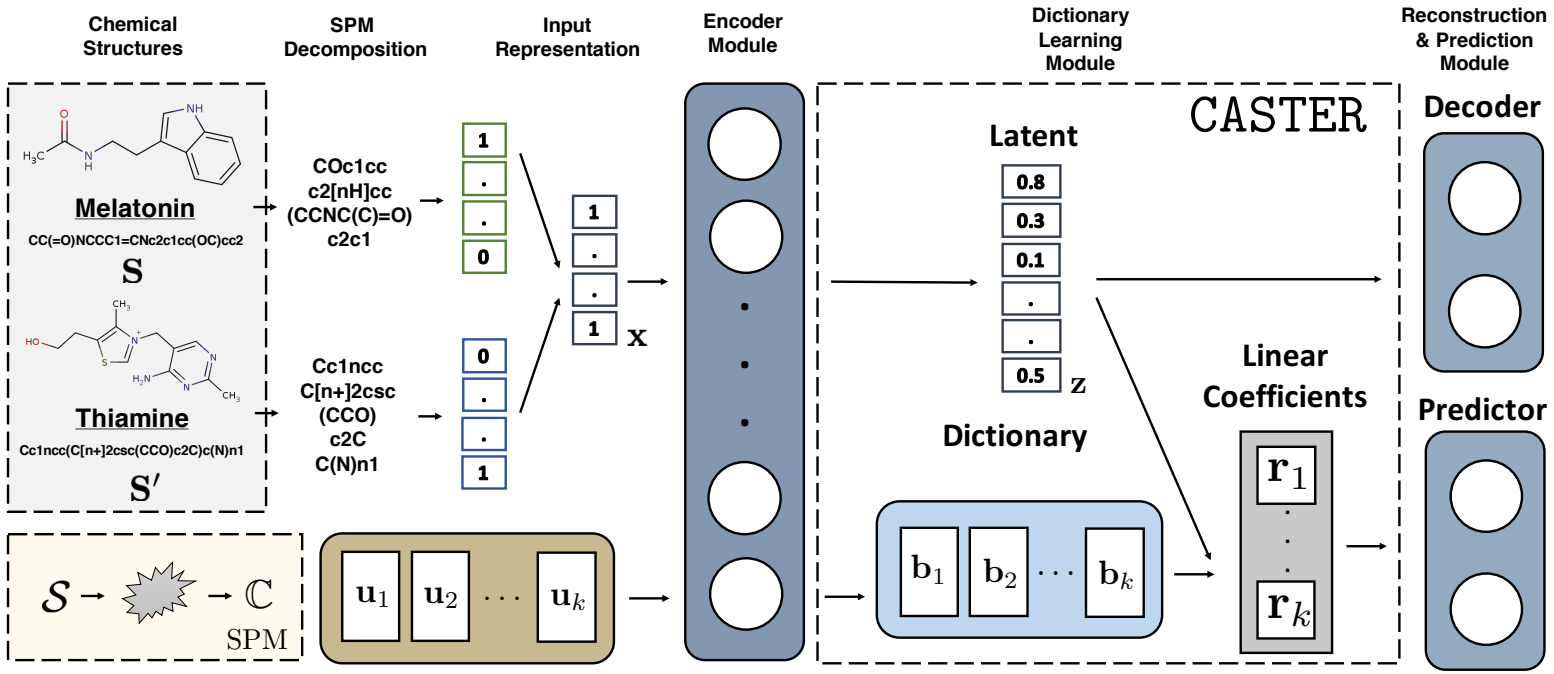


FDA reports that adverse drug interactions could kill more than **100,000** patients in the US every year.

The cost of preventable adverse drug events would be as high as **115 billion dollars**.

CASTER first automatically decomposes any drug molecules into a **sequence of sub-structures** in a data-driven manner.

Then, a **deep dictionary learning** module generates coefficients for each substructure. **Substructures contributed to the interaction outcome are marked with high coefficients.**



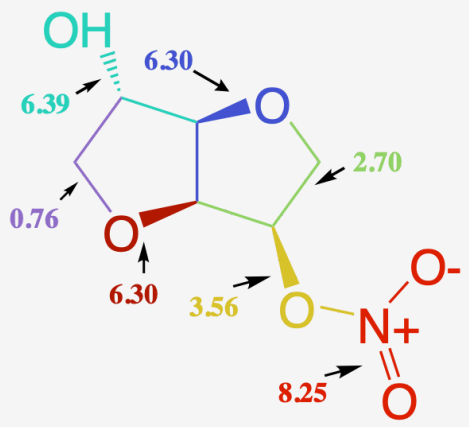
* Source: U.S. Department of Health and Human Services

+ Source: FDA, McKinsey

CASTER: Predicting Drug Interactions with Chemical Substructure Representation

Kexin Huang[1,2], Cao Xiao[2], Trong Nghia Hoang[3], Lucas M. Glass[2], Jimeng Sun[4,5]

[1]Harvard [2]IQVIA [3]MIT-IBM Watson AI Lab [4]UIUC CS [5]Georgia Tech

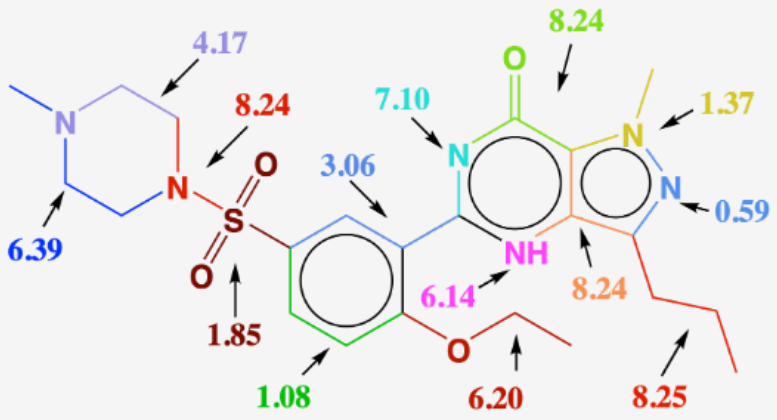


Isosorbide Mononitrate

O=[N+]([O-])O[C@@H]1CO[C@@H]2[C@@H](O)CO[C@H]12

CASTER

$p(\mathbf{x}) = 0.793$



Sildenafil

CCCc1nn(C)c2c(=O)nc(-c3cc(S(=O)(=O)N4CCN(C)CC4)ccc3OCC)[nH]c12

CASTER assigns **high** coefficients to **relevant** substructures for the interaction outcome.

The coefficients are **robust** against model random initialization. It achieves **0.767 correlation** across five different random seeds.

Model	Dataset	ROC-AUC	PR-AUC	F1
LR	BIOSNAP	0.802 ± 0.001	0.779 ± 0.001	0.741 ± 0.002
	DrugBank	0.774 ± 0.003	0.745 ± 0.005	0.719 ± 0.006
Nat.Prot	BIOSNAP	0.853 ± 0.001	0.848 ± 0.001	0.714 ± 0.001
	DrugBank	0.786 ± 0.003	0.753 ± 0.003	0.709 ± 0.004
Mol2Vec	BIOSNAP	0.879 ± 0.006	0.861 ± 0.005	0.798 ± 0.007
	DrugBank	0.849 ± 0.004	0.828 ± 0.006	0.775 ± 0.004
MolVAE	BIOSNAP	0.892 ± 0.009	0.877 ± 0.009	0.788 ± 0.033
	DrugBank	0.852 ± 0.006	0.828 ± 0.009	0.769 ± 0.031
DeepDDI	BIOSNAP	0.886 ± 0.007	0.871 ± 0.007	0.817 ± 0.007
	DrugBank	0.844 ± 0.003	0.828 ± 0.002	0.772 ± 0.006
CASTER	BIOSNAP	0.910 ± 0.005	0.887 ± 0.008	0.843 ± 0.005
	DrugBank	0.861 ± 0.005	0.829 ± 0.003	0.796 ± 0.007

CASTER has the **best** predictive performance across two datasets against state-of-the-art baselines. Meaning that it captures the mechanism of drug interactions.

Sunday, February 9, 7:30 – 9:30 PM

Applications (APP) #20 --- APP5592