UL Cheminformatics Suite

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

ToxTrack

ULReachacross.com

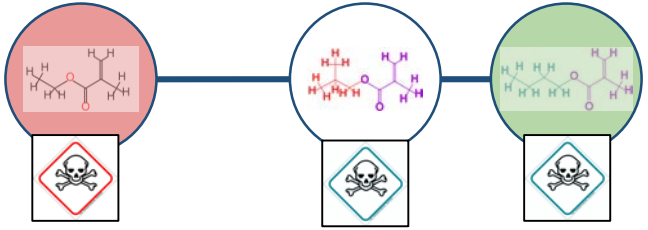
Jan 2016 | Dec 2016 | SOT 2017

Legacy

ECHA database
NLP collection of ECHA C&L

Hopkins Publications
Skin sens., Eye irrit., Oral models

Production tool release
https://www.ulreachacross.com/



Acute Oral?

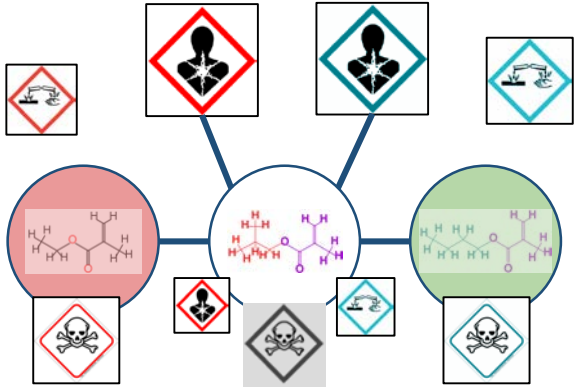
Mid 2017 | Fall 2018 | Feb 2018

Production | Development

Algorithm Upgrade
data fusion / potency

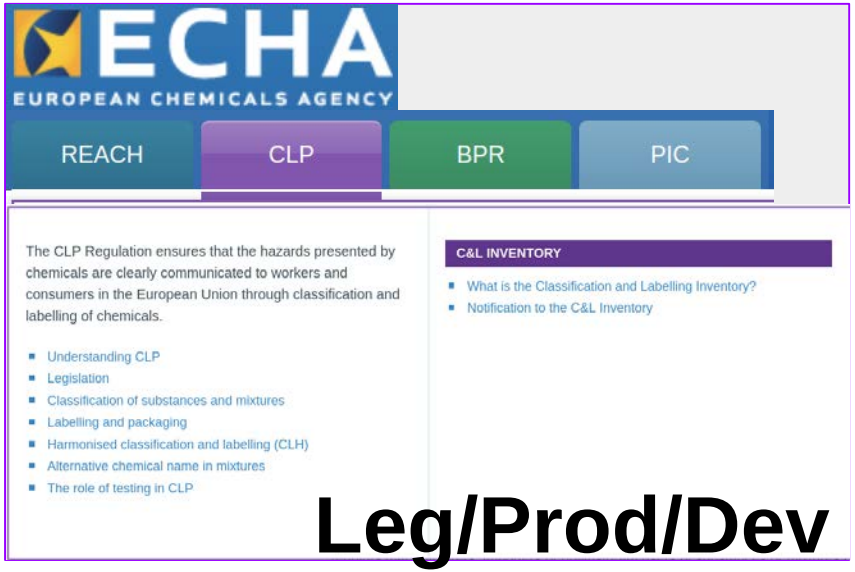
Datasource Integration
pubchem / integration pipeline

Validation / Iteration
Dev Cycle and NTP Challenge



Acute Oral?

Data source



Leg/Prod/Dev

Dev

Dev

Data source

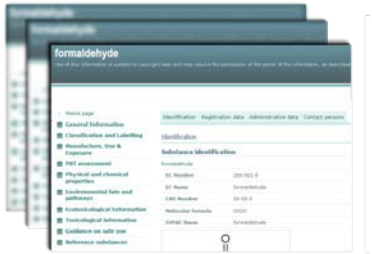
Leg/Prod/Dev

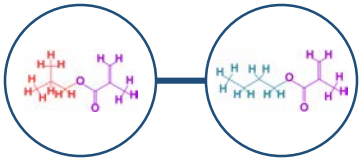
833844 Chemical Endpoints

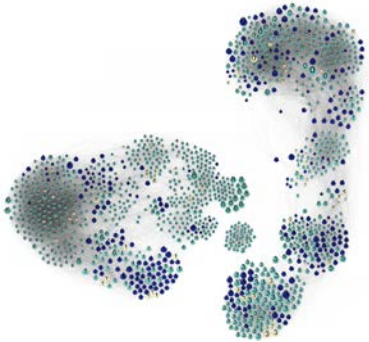
Rows 20

smiles	endpoint	inchi	value
O=C(CCl)CC	H314	AALRHBLMAV...	-1
O=C(O)C1=CC=C2C(=O)N(C(=O)...	H302	AAOFNSJIPAZH...	-1
O=C1C=CC=CC1=CNC=2C=CC...	H319	AAPPQBJWIDZ...	1
O=P(C=1C=CC=CC1)(C2=CC=C(...	H315	AAYLOGMTTM...	1
O=C(O)CCCC1=CC=C(C(=C1)C)C	H303	ABMVUAWFTZ...	-1
O=C(NCC)CC1N=C(C=2C=CC(Cl...	H220	AAAQFGUYHFJ...	-1
O=C1C=CC=CC1=CNNC(=O)C...	H402	AABQOXLSWQ...	-1
O=C(OC1=CC=CC(C=NNC(=O)C...	H412	AAGVMZPRDN...	-1
C#CC1=CC=CC(=C1)NC=2N=C...	H272	AAKJLRGGTJKA...	-1
N#CC(F)(C(F)(F)F)C(F)(F)F	H420	AASDJASZOZG...	-1
O=P(OCC)(OCC)CC	NTPAcuteOralChallenge_nonToxic	AATNZNJRDOV...	1
[I-].ClC=1C=CC=C[N+]1C	H319	ABFPKTQEQNI...	1
Cl.O=C(OCC)CNC(C)C	H314	ABTRDXFEQPO...	-1
NC1=CC=C(C=C1)[Sn](C=2C=C...	H315	ABVNDIYOGM...	1
OC(COC=1C=CC=CC1C2CCCC...	NTPAcuteOralChallenge_LD50	ABXHHEZNIJU...	1850
IC=1C=CC=2C3=CC=C(I)C=C3C(...	H410	ABZISBKAAVQ...	-1
O=C(N)C1=CC(NC1(C)C)(C)C	H314	ACFYUJLIWIDS...	-1
S=C(NN)NC1=CC=C(Br)C=C1	H318	ACKSCWQUPJX...	-1
O=C(N1C=C(C=2C=CC=CC21)C...	H413	ACZZUIXOIFNC...	1
O=C=NC1=CC=C(Cl)C=C1	H260	ADAKRBAJFHTI...	-1

Similarity





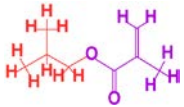


Data Source

Similarity

Graph Algorithms

Fingerprinter



0 1 0 0 1 1 0

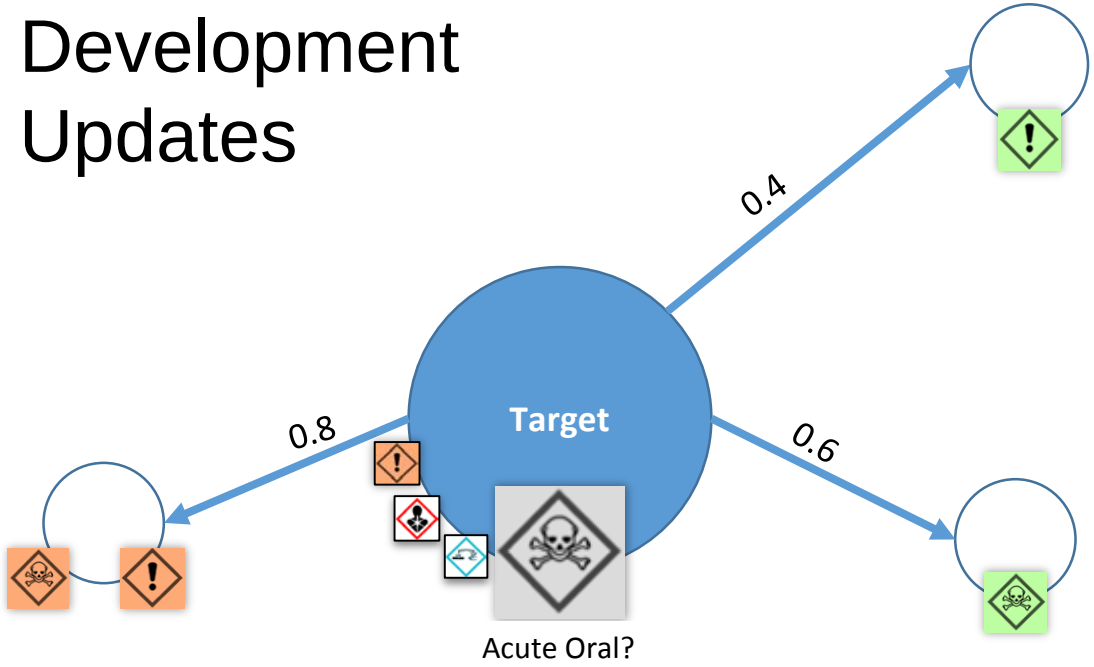
Metric

$$M(0\,1\,0\,0\ldots,0\,0\,1\,0\ldots) = -\infty \leq x \leq \infty$$

$$\frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Tanimoto - Heuristic

Development Updates



Hazard/properties:

79 (eg H225 - flammable liquid)

Features

79 x 3 = 237 (target & pos & neg)

Database:

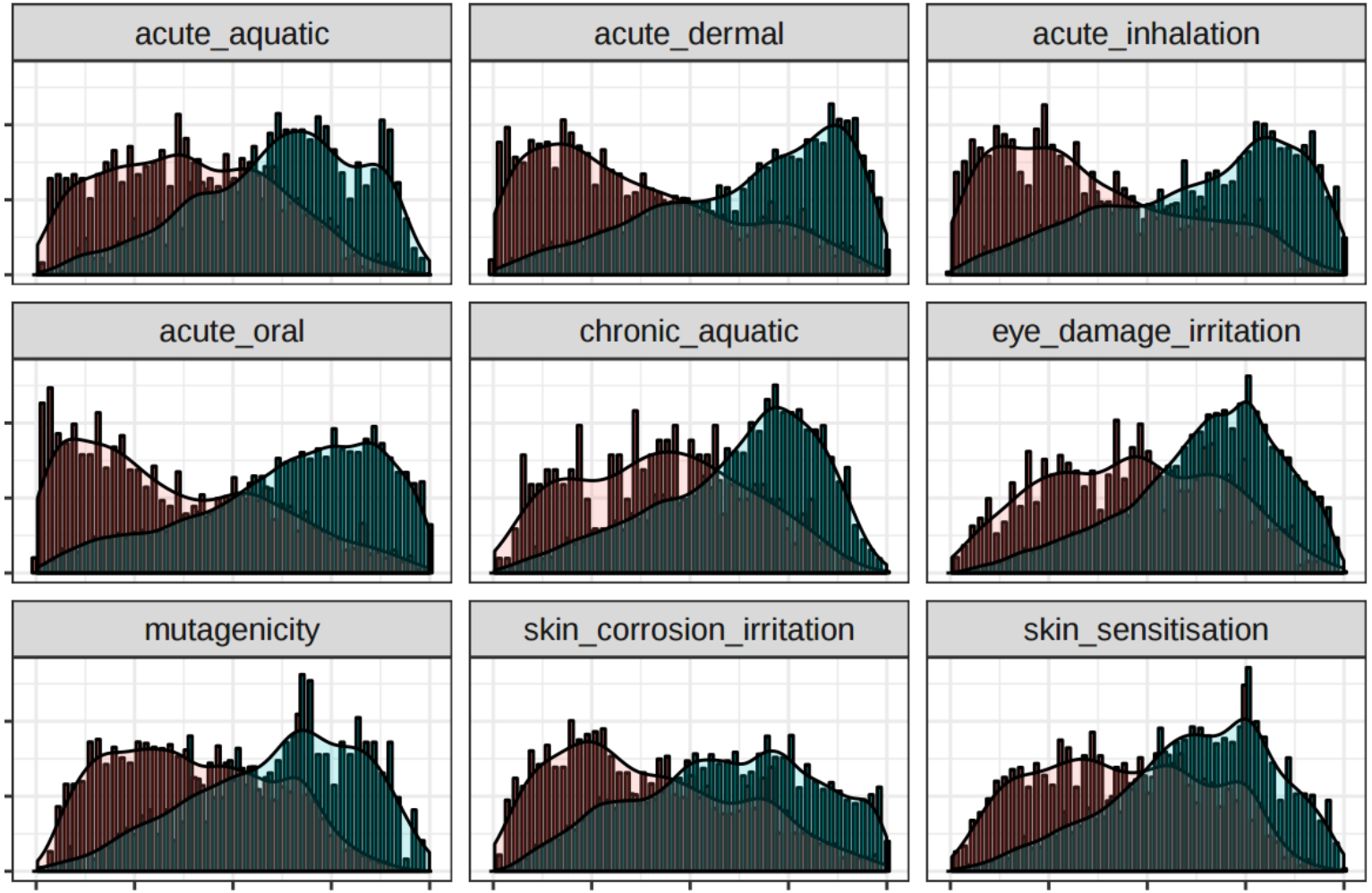
ECHA C&L + Pubchem + NTP

Learning:

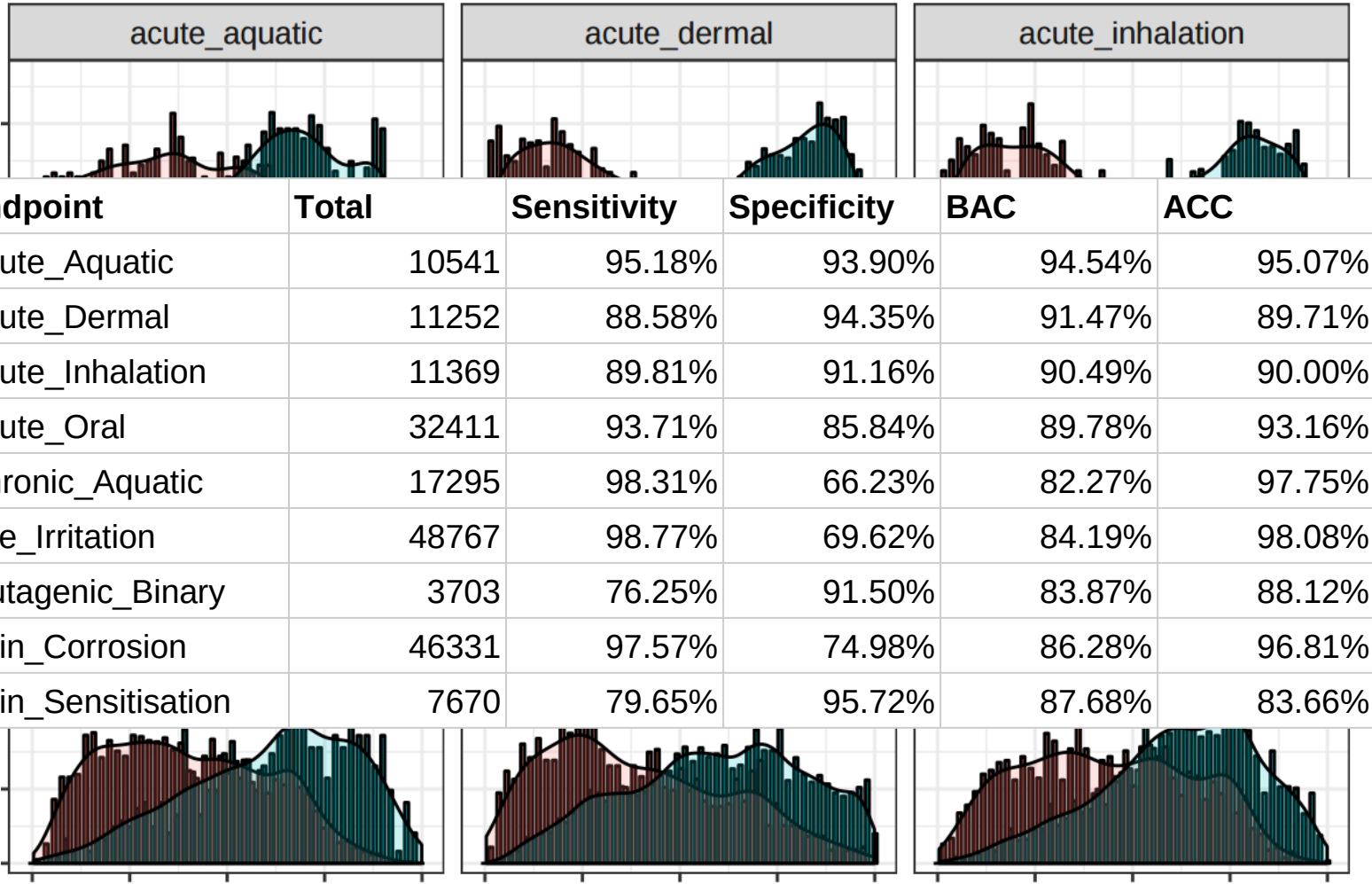
Random Forest / grad. Boost trees
Multilayer Perceptron

Target			Source Pos		Source Neg	
acid	muta.	corro.	Oral	Acid	Oral	Acid
T	T	F	0.8	0.8	0.6	0.4

Binary Hazard

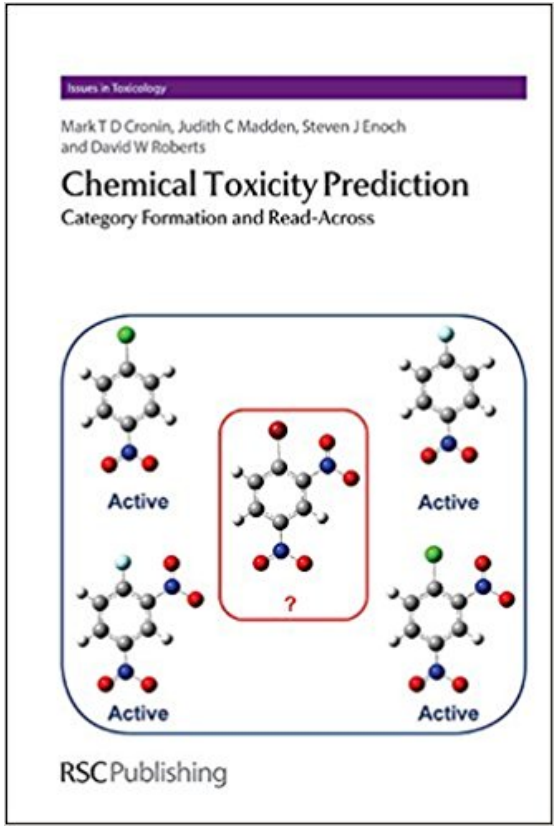


Binary Hazard



Endpoint	Total	Sensitivity	Specificity	BAC	ACC
Acute_Aquatic	10541	95.18%	93.90%	94.54%	95.07%
Acute_Dermal	11252	88.58%	94.35%	91.47%	89.71%
Acute_Inhalation	11369	89.81%	91.16%	90.49%	90.00%
Acute_Oral	32411	93.71%	85.84%	89.78%	93.16%
Chronic_Aquatic	17295	98.31%	66.23%	82.27%	97.75%
Eye_Irritation	48767	98.77%	69.62%	84.19%	98.08%
Mutagenic_Binary	3703	76.25%	91.50%	83.87%	88.12%
Skin_Corrosion	46331	97.57%	74.98%	86.28%	96.81%
Skin_Sensitisation	7670	79.65%	95.72%	87.68%	83.66%

Advantages



Familiar Concept

endpoint	pos	neg	total
skin_sensitisation	2865	1886	4751
eye_damage_irritation	14778	944	15722
acute_oral	10225	1932	12157
mutagenicity	600	2795	3395
skin_corrosion_irritation	13758	1348	15106
acute_dermal	4334	1980	6314
acute_aquatic	1122	921	2043
chronic_aquatic	2554	251	2805
acute_inhalation	4812	1372	6184

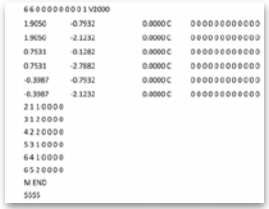
More Data

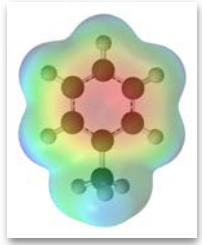
Issues

Representation

= 4 H
1
>
= 8 H
0
>
= 16 H
0
>
= 32 H
0

CC1=CC=CC=C1





Fingerprinters | **SMILES** | **Mol file** | **Electron Density**

Heuristic Similarity

= 1 Li
0
>
= 2 Li

$$\frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Tanimoto - Heuristic

Issues

Representation

= 4 H
1
>
= 8 H
0
>
= 16 H
0
>
= 32 H
0

Fingerprinters

= 1 Li
0
>
= 2 Li

Heuristic Similarity

$$\frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

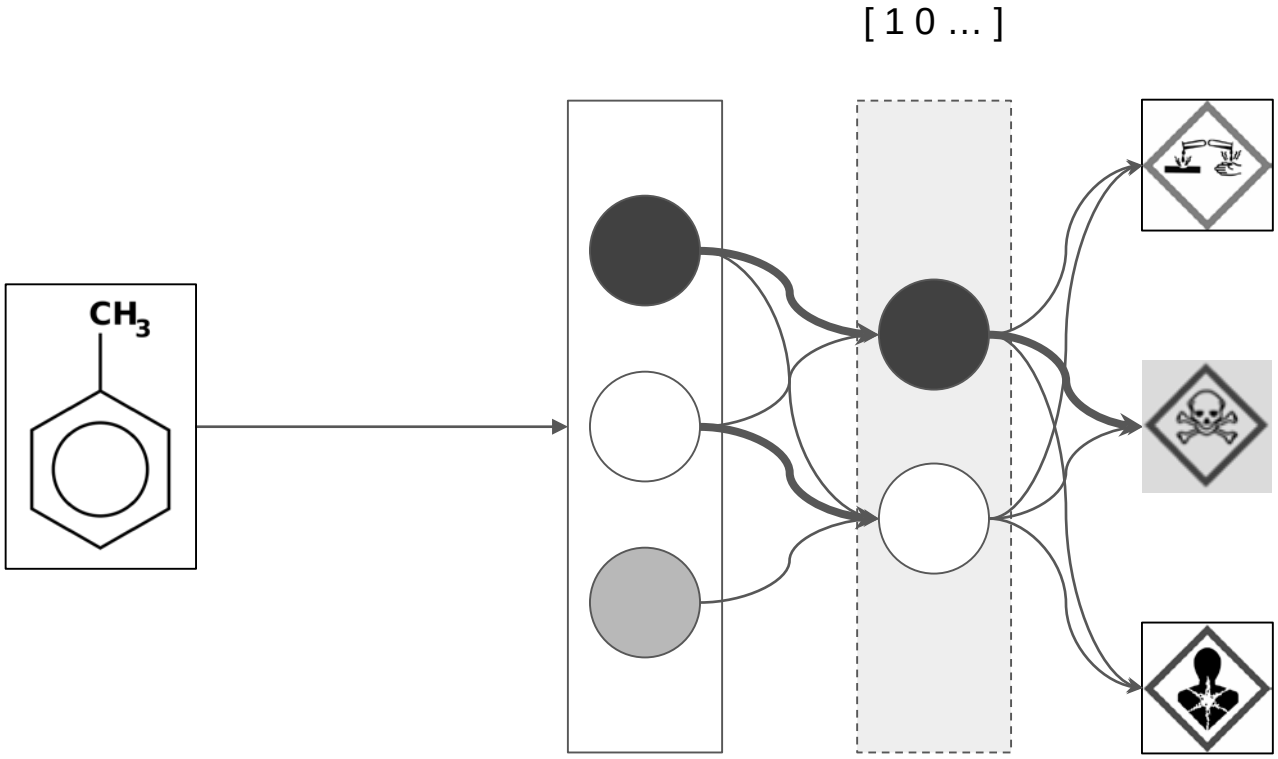
Tanimoto - Heuristic

Rapid Data Integration

Rows 20

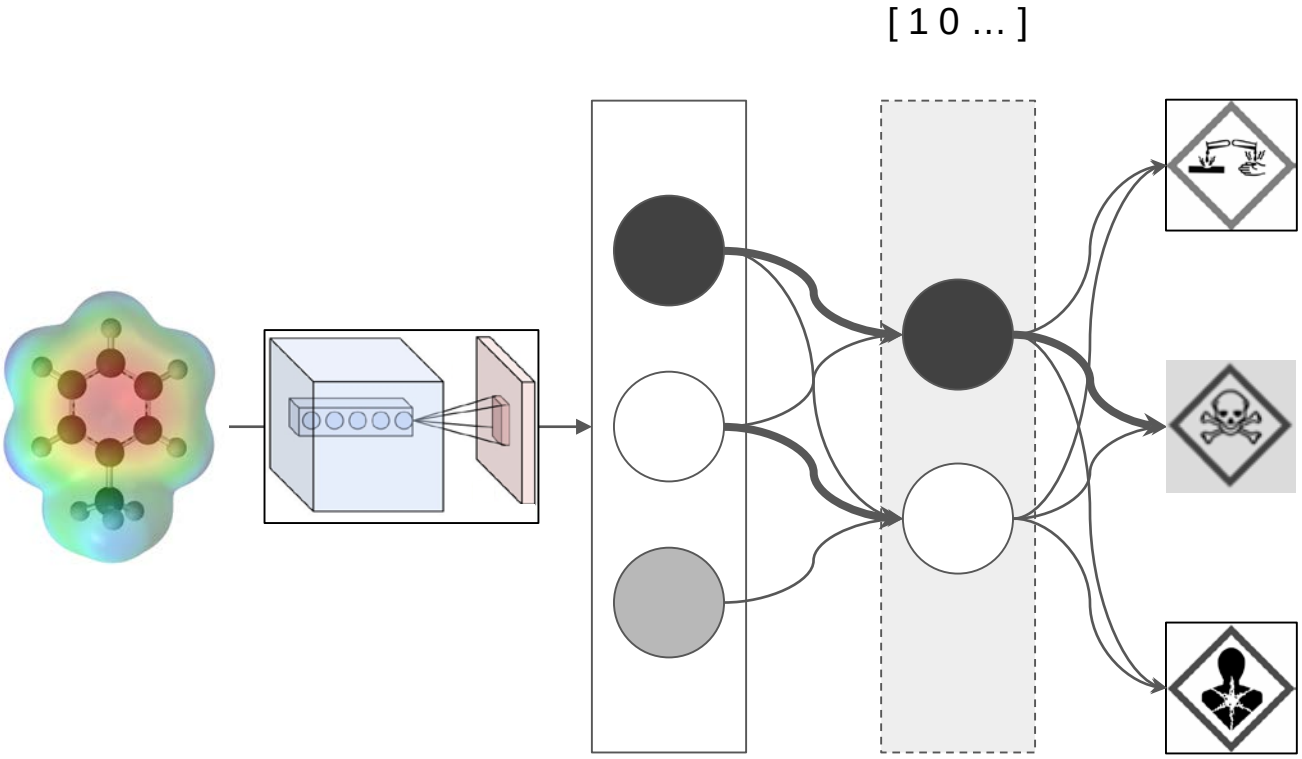
smiles	endpoint	inchi	value
O=C(CCl)CC	H314	AALRHBLMAV...	-1
O=C(O)C1=CC=C2C(=O)N(C(=O)...	H302	AAOFNSJIPAZH...	-1
O=C1C=CC=CC1=CNC=2C=CC...	H319	AAPPQBJWIDZ...	1
O=P(C=1C=CC=CC1)(C2=CC=C(...	H315	AAYLOGMTTM...	1
O=C(O)CCCC1=CC=C(C(=C1)C)C	H303	ABMVUAWFTZ...	-1
O=C(NCC)CC1N=C(C=2C=CC(Cl...	H220	AAAQFGUYHFJ...	-1
O=C1C=CC=CC1=CNNC(=O)C...	H402	AABQOXLSWQ...	-1
O=C(OC1=CC=CC(C=NNC(=O)C...	H412	AAGVMZPRDN...	-1
C#CC1=CC=CC(=C1)NC=2N=C...	H272	AAKJLRGGTJKA...	-1
N#CC(F)(C(F)(F)F)C(F)(F)F	H420	AASDJASZOZG...	-1
O=P(OCC)(OCC)CC	NTPAcuteOralChallenge_nonToxic	AATNZNJRDOV...	1
[I-].ClC=1C=CC=C[N+]1C	H319	ABFPKTQEQNI...	1
Cl.O=C(OCC)CNC(C)C	H314	ABTRDXFEQPO...	-1
NC1=CC=C(C=C1)[Sn](C=2C=C...	H315	ABVNDIYOGM...	1
OC(COC=1C=CC=CC1C2CCCC...	NTPAcuteOralChallenge_LD50	ABXHHEZNIJU...	1850
IC=1C=CC=2C3=CC=C(I)C=C3C(...	H410	ABZISBKAAVQ...	-1
O=C(N)C1=CC(NC1(C)C)(C)C	H314	ACFYUJLIWIDS...	-1
S=C(NN)NC1=CC=C(Br)C=C1	H318	ACKSCWQUPJX...	-1
O=C(N1C=C(C=2C=CC=CC21)C...	H413	ACZZUIXOIFNC...	1
O=C=NC1=CC=C(Cl)C=C1	H260	ADAKRBAJFHTI...	-1

Multi-task learning



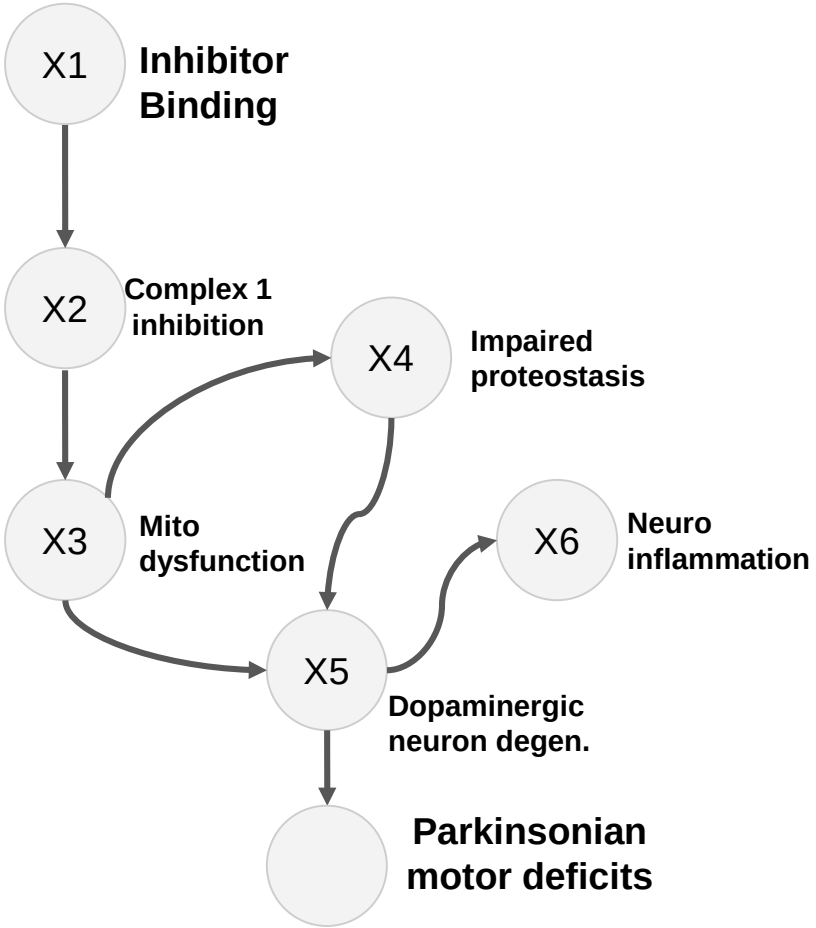
Deep Chemchart http://dev.chemchart.com:5006/

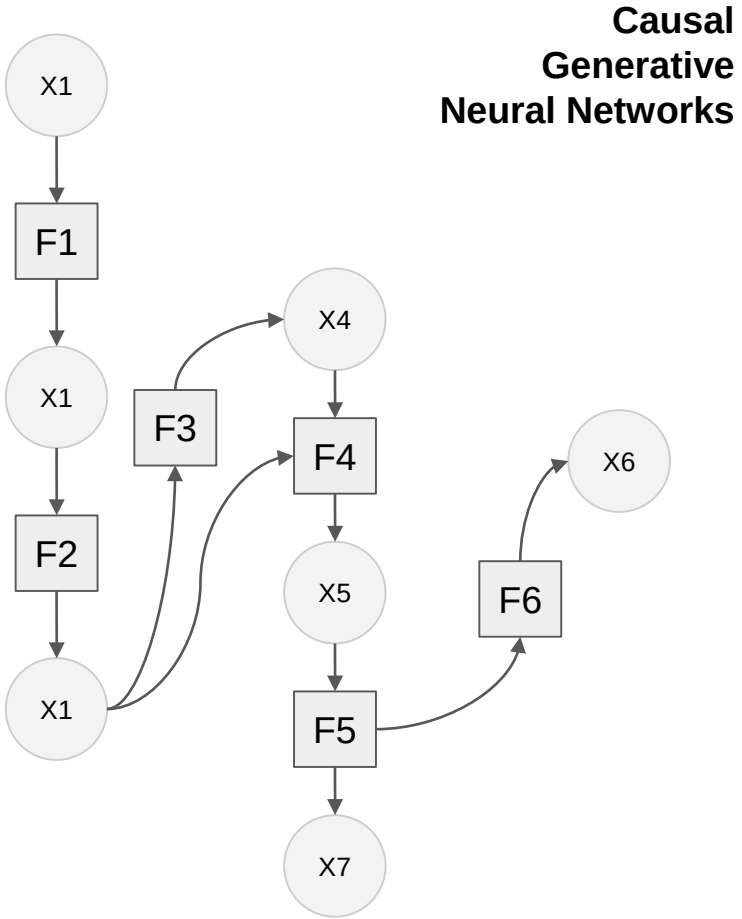
Multi-task learning



Deep Chemchart http://dev.chemchart.com:5006/

Adverse Outcome Pathways





Conclusions

ULReachAcross.com

Read Across + Transfer Learning

Multi-task neural networks

Adverse Outcome Pathway + Causal Generative Neural Networks