

UL Cheminformatics Suite



JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

TOXTRACK

ULReachacross.com

Jan 2016

Dec 2016

SOT 2017

Legacy

ECHA database

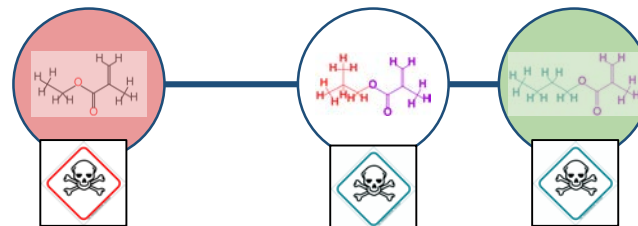
NLP collection of ECHA C&L

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

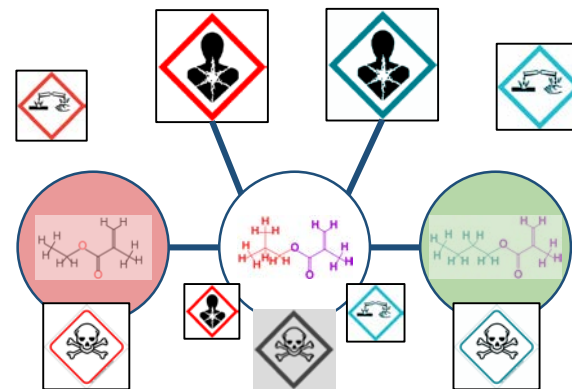
data fusion / potency

Datasource Integration

pubchem / integration pipeline

Validation / Iteration

Dev Cycle and NTP Challenge



Acute Oral?

Data source



ECHA
EUROPEAN CHEMICALS AGENCY

REACH | **CLP** | BPR | PIC

The CLP Regulation ensures that the hazards presented by chemicals are clearly communicated to workers and consumers in the European Union through classification and labelling of chemicals.

C&L INVENTORY

- What is the Classification and Labelling Inventory?
- Notification to the C&L Inventory

- Understanding CLP
- Legislation
- Classification of substances and mixtures
- Labelling and packaging
- Harmonised classification and labelling (CLH)
- Alternative chemical name in mixtures
- The role of testing in CLP

Leg/Prod/Dev



PubChem

12 Safety and Hazards

12.1 Hazards Identification

12.1.1 GHS Classification



Dev



National Toxicology Program
U.S. Department of Health and Human Services

Dev

Data source

ECHA
EUROPEAN CHEMICALS AGENCY

REACH | **CLP** | BPR | PIC

The CLP Regulation ensures that the hazards presented by chemicals are clearly communicated to workers and consumers in the European Union through classification and labelling of chemicals.

- Understanding CLP
- Legislation
- Classification of substances and mixtures
- Labelling and packaging
- Harmonised classification and labelling (CLH)
- Alternative chemical name in mixtures
- The role of testing in CLP

C&L INVENTORY

- What is the Classification and Labelling Inventory?
- Notification to the C&L Inventory

Leg/Prod/Dev

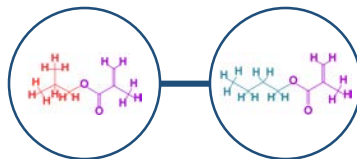
833844
Chemical
Endpoints

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

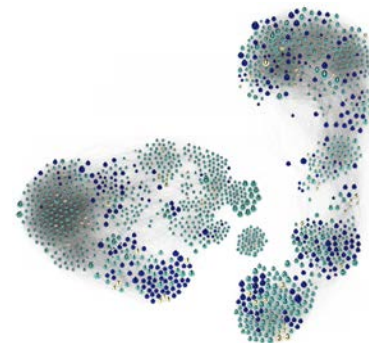
Similarity



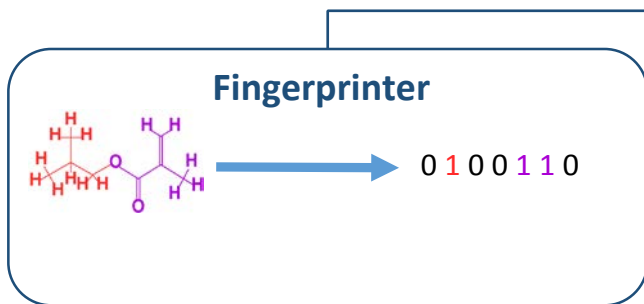
Data Source



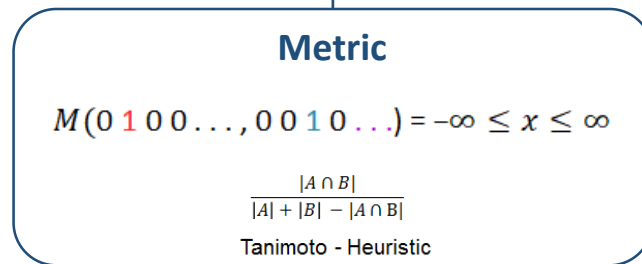
Similarity



Graph Algorithms

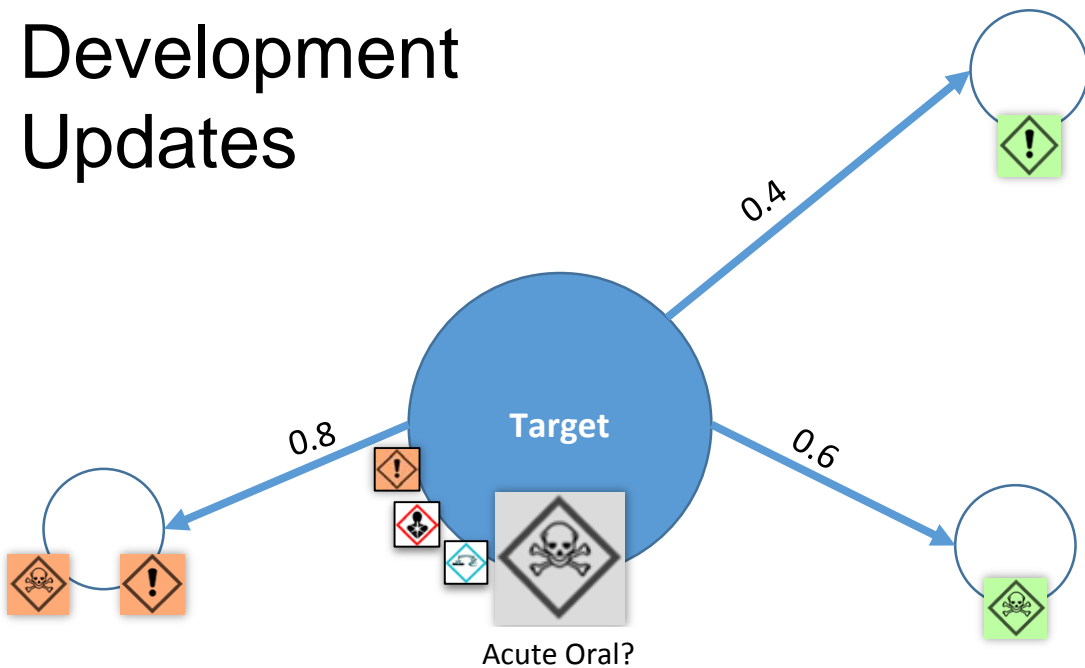


Fingerprint



Metric

Development Updates



Hazard/properties:

79 (eg H225 - flammable liquid)

Features

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

Database:

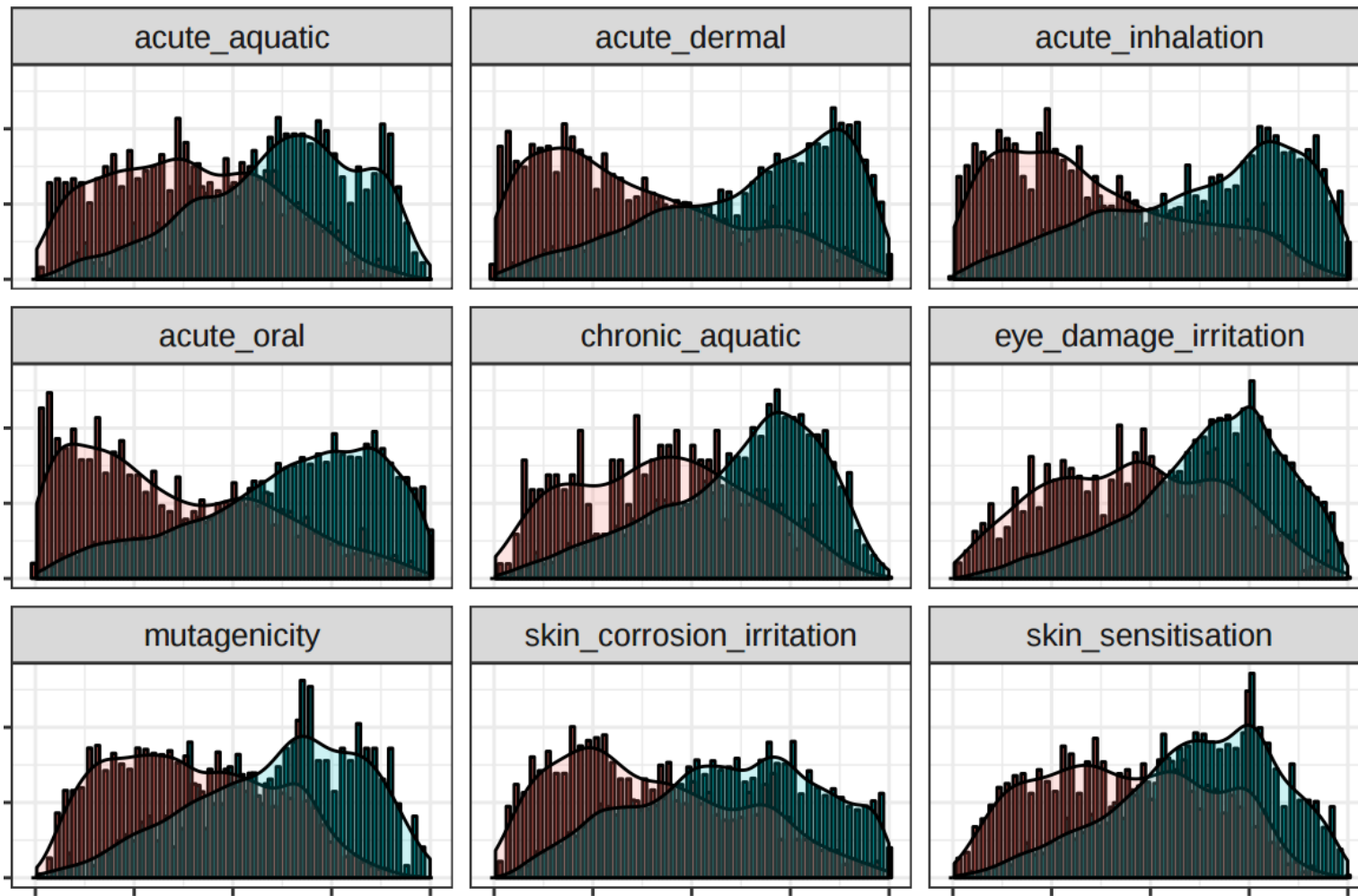
ECHA C&L + Pubchem + NTP

Learning:

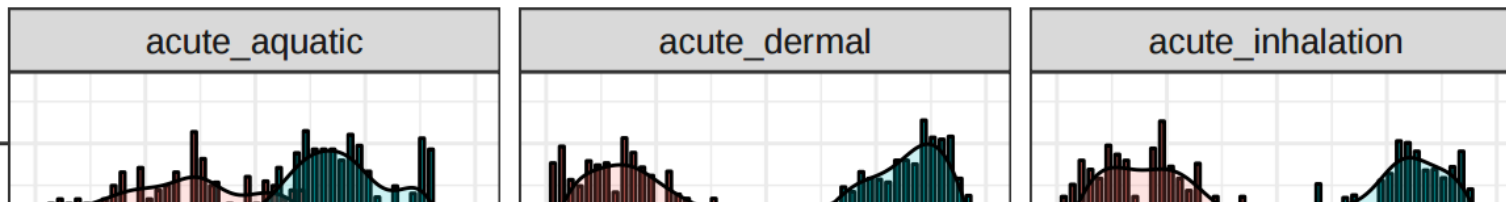
Random Forest / grad. Boost trees
Multilayer Perceptron

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

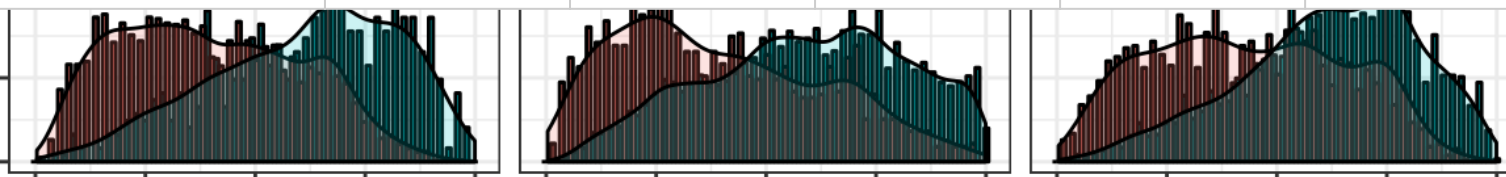
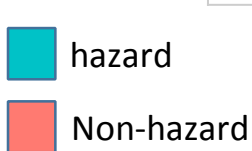
Binary Hazard



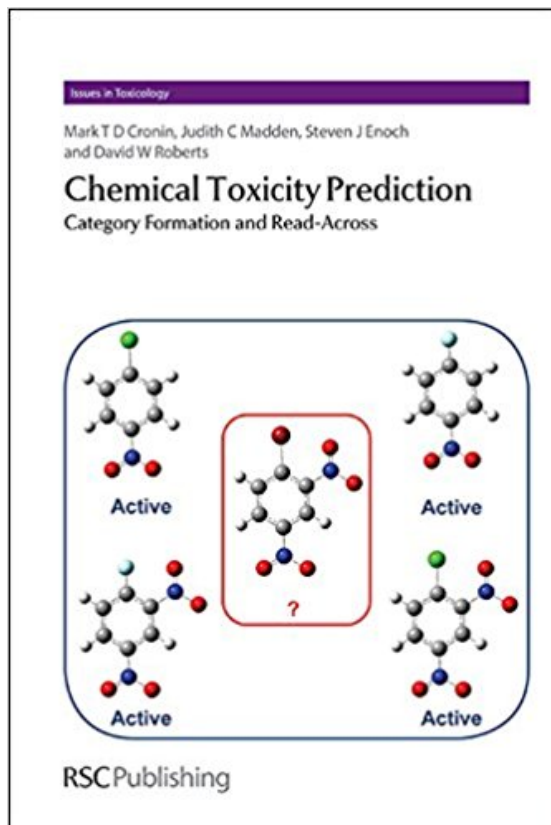
Binary Hazard



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
Fingerprints

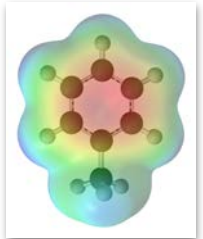


SMILES

```

6600000000110000
19050 -0.7931 0.0000C 00000000000
19050 -2.1282 0.0000C 00000000000
0.7531 -0.1282 0.0000C 00000000000
0.7531 -2.7882 0.0000C 00000000000
-0.3987 -0.7931 0.0000C 00000000000
-0.3987 -2.1282 0.0000C 00000000000
22110000
3120000
4220000
5310000
6410000
65210000
M END
5055
    
```

Mol file



Electron Density

Heuristic Similarity

= 2 Li
>

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

Tanimoto - Heuristic



Issues

= 4 H

1

>

= 8 H

0

>

= 16 H

0

>

= 32 H

0

Fingerprints

= 1 Li

0

= 2 Li

>

Representation

Heuristic Similarity

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

Tanimoto - Heuristic

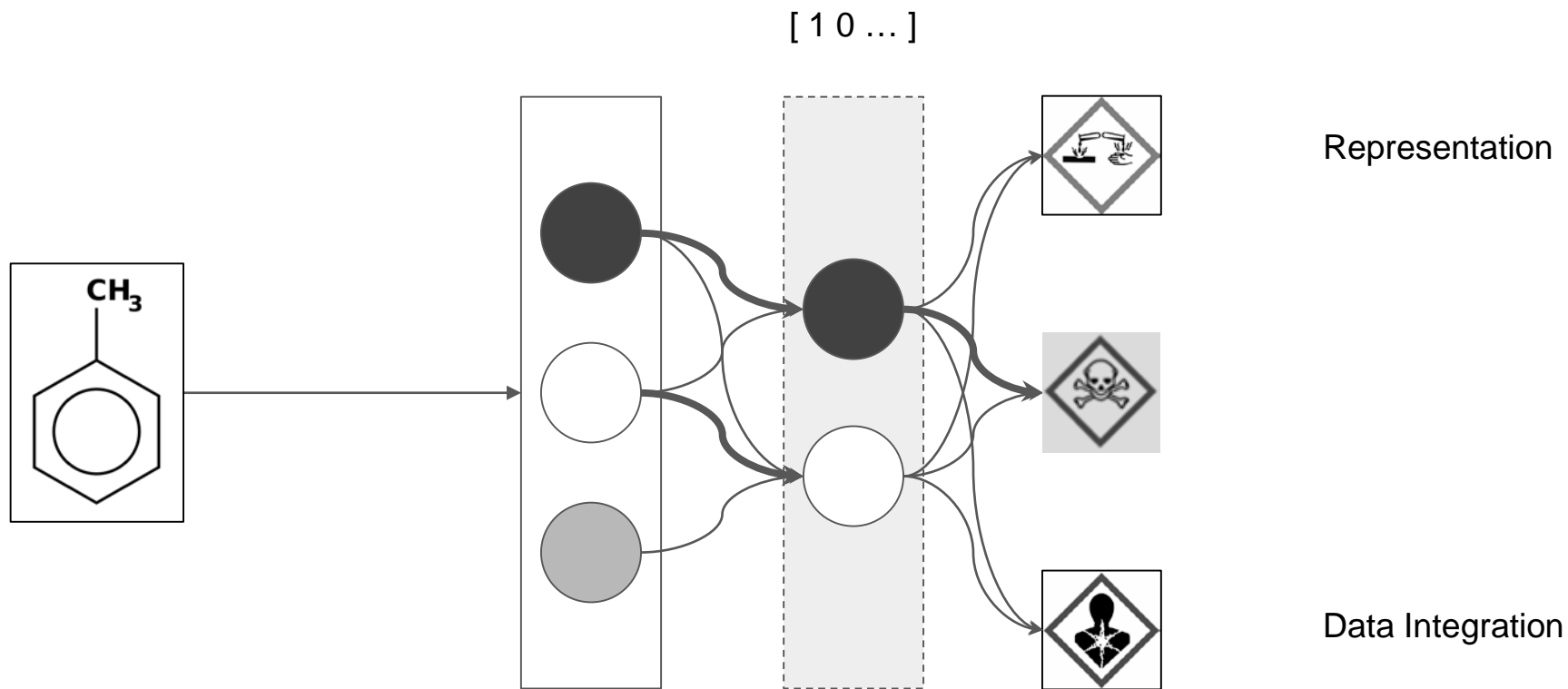


Rapid Data Integration

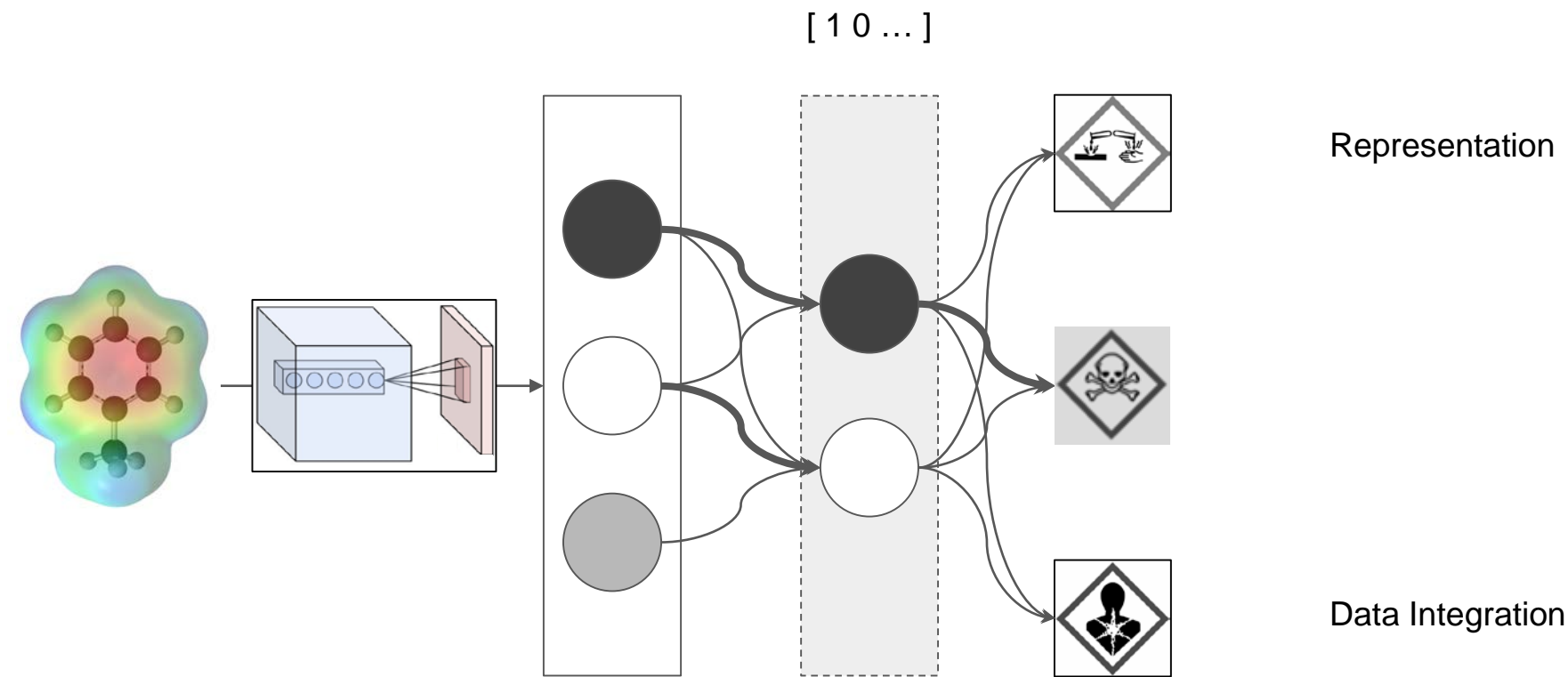
Rows 20

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

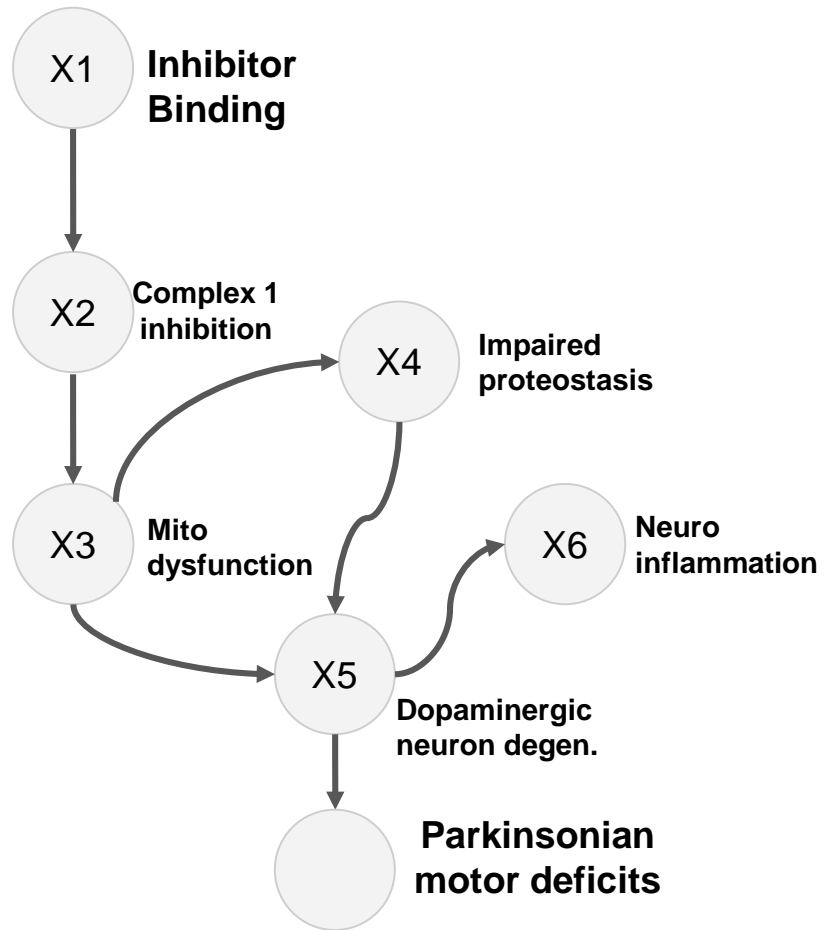
Multi-task learning



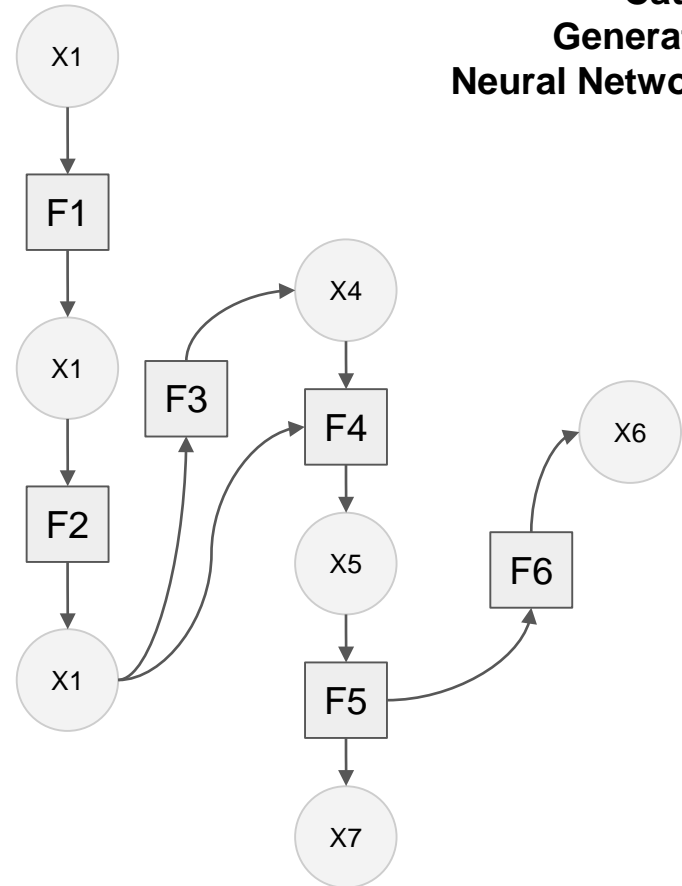
Multi-task learning



Adverse Outcome Pathways



Causal Generative Neural Networks



Conclusions

ULReachAcross.com

Read Across + Transfer Learning

Multi-task neural networks

Adverse Outcome Pathway + Causal Generative Neural Networks