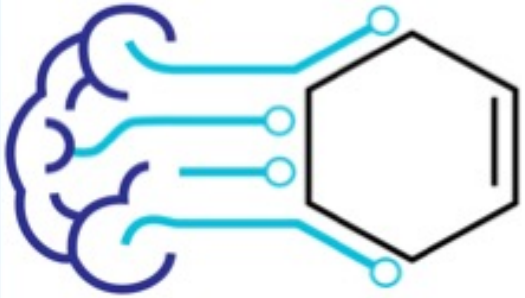


Institute of Structural Biology

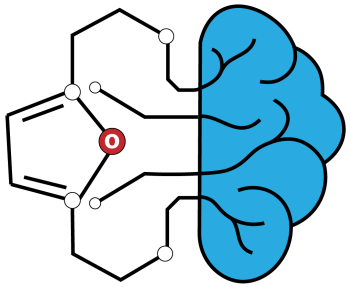
Advanced Machine Learning in Drug Discovery

Igor V. Tetko

Helmholtz Munich and BIGCHEM GmbH

Katholieke Universiteit te Leuven, 2 September 2024

HELMHOLTZ MUNICH




HELMHOLTZ
MUNICH
German Research Center for Environmental Health

Agenda

Computational predictions in drug discovery

OCHEM platform

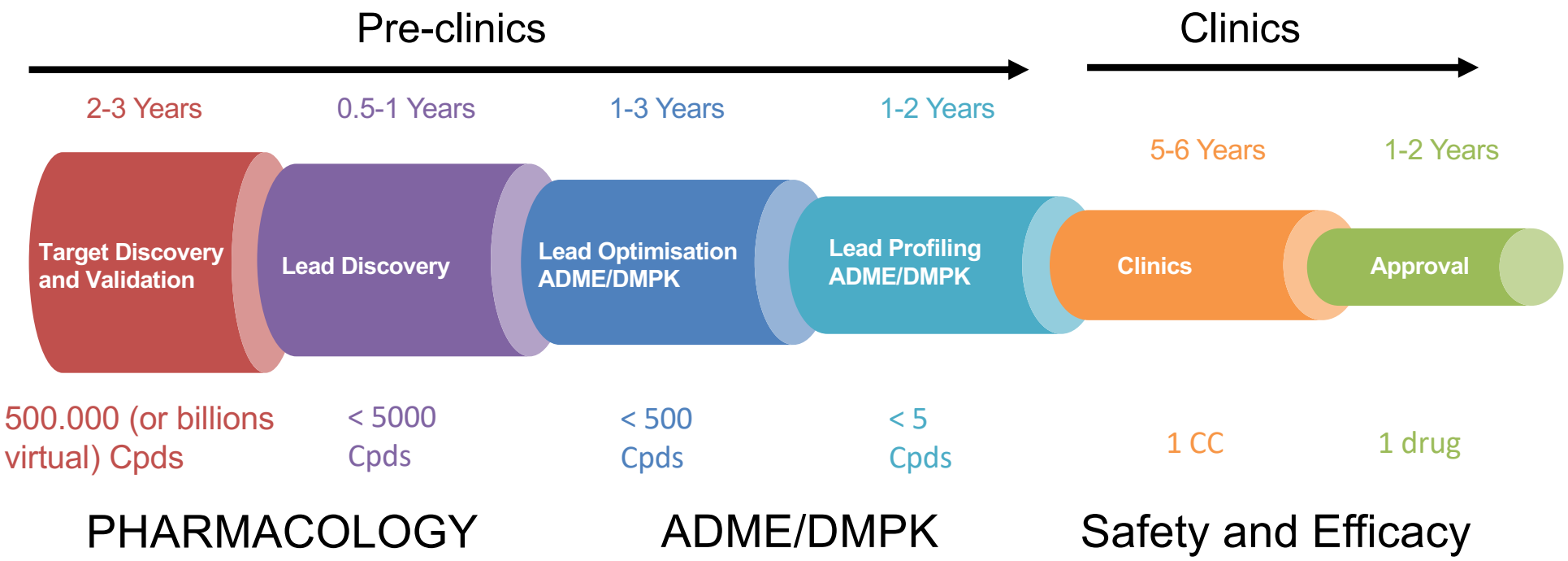
Consensus modelling as strategy to develop best models

Tox24 Challenge results

Uncertainty estimation and Applicability domain

Explainable AI

Traditional Process of Drug Discovery



- Profiling and screening in the virtual space helps to identify the most promising candidates

Slide courtesy of Dr. C. Höfer, Merck

ADMETox filters in Bayer

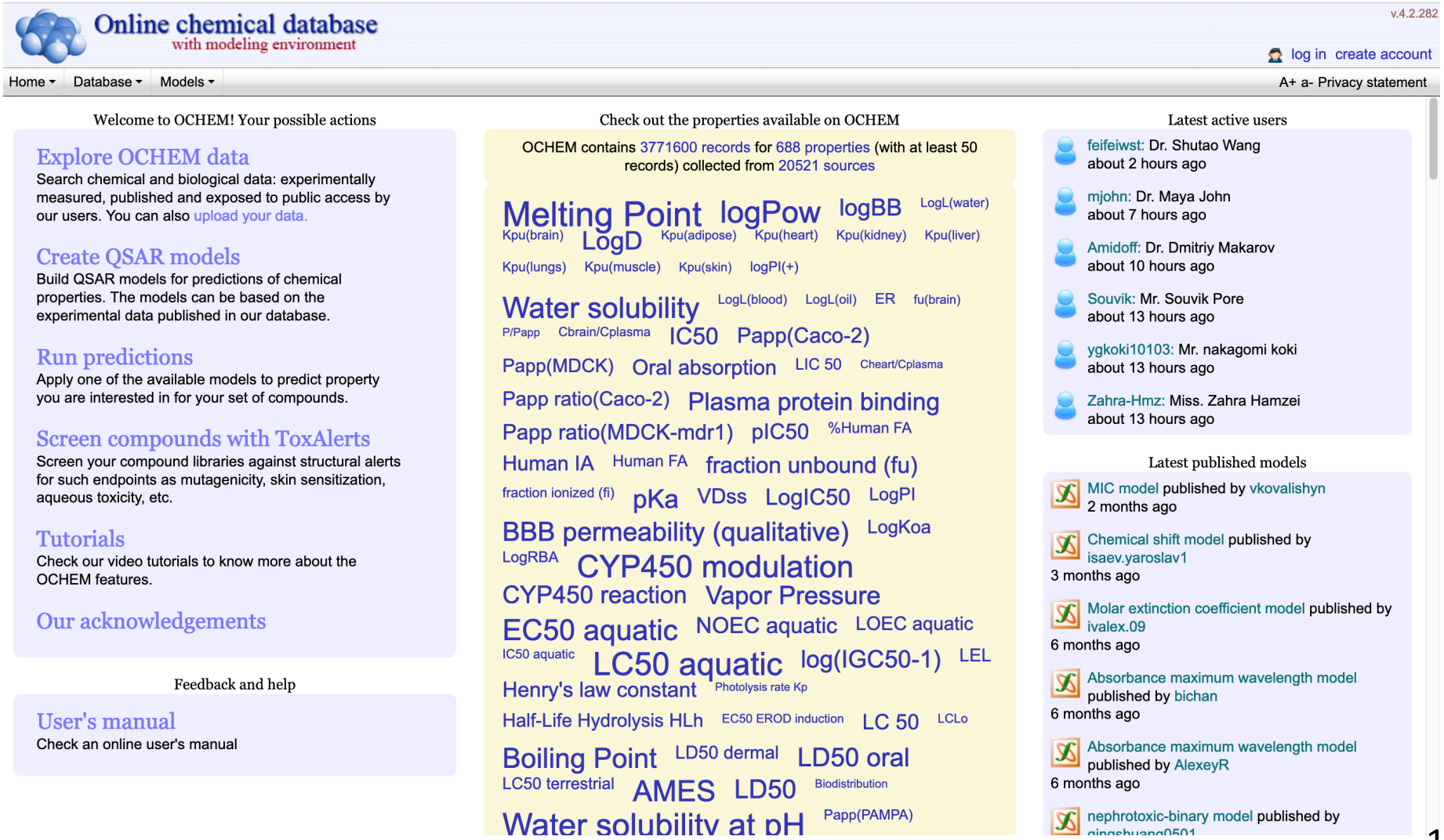
Insufficient quality | First approach | Medium model | Good model | Robust model

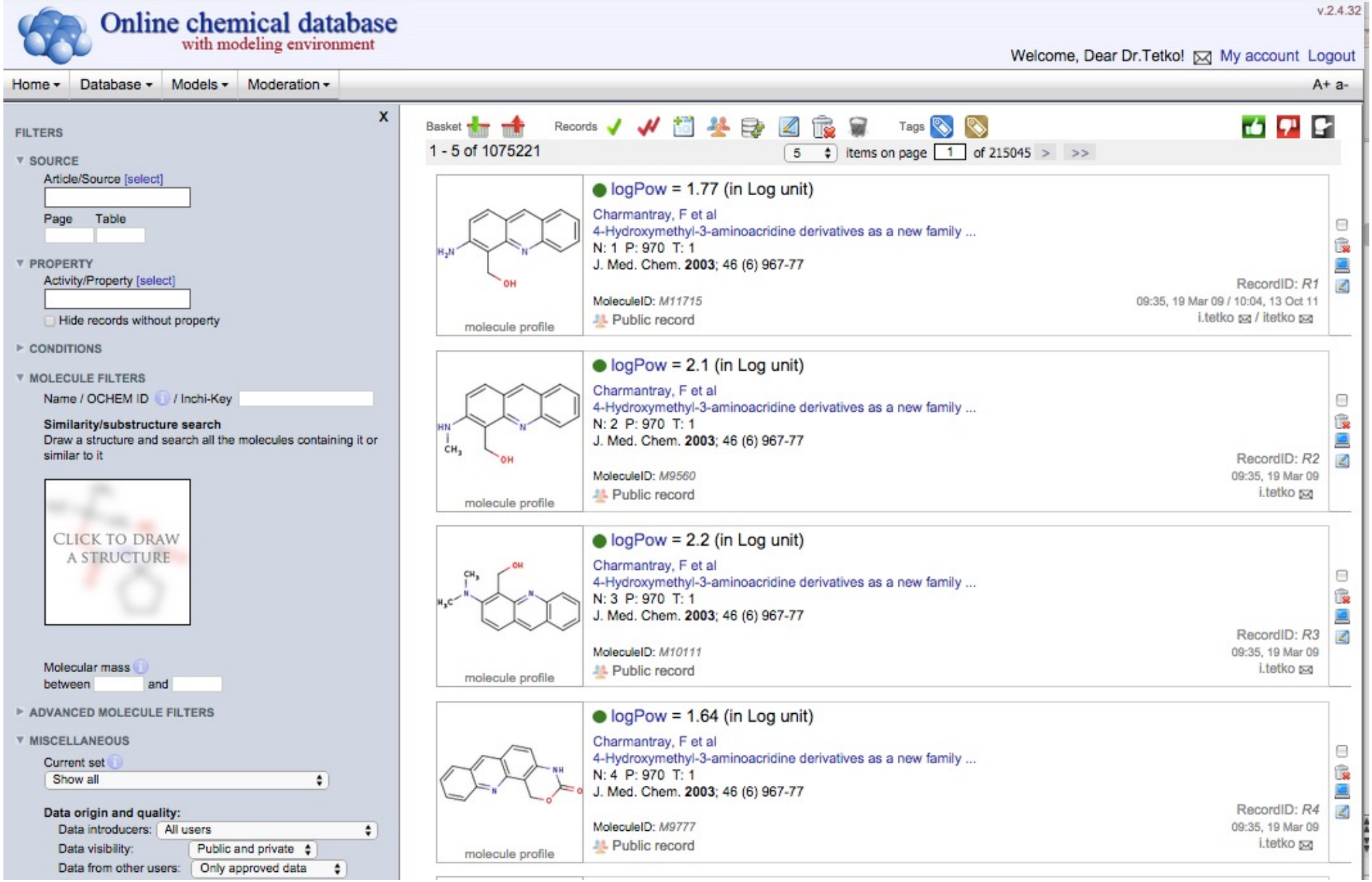
Endpoint		Model type	Data set size	2005	2009	2014	2019	Retraining
Absorption	Caco-2 permeation	C (N)	>10 000			RF	SVR	Weekly
	Caco-2 efflux	C (N)	>10 000			RF	SVR	Weekly
	Bioavailability (rat)	C	~2000				RF	On demand
Distribution	Human serum albumin	N	>30 000			PLS	MTNN	On demand
	Fraction unbound	N	>1000			PLS	MTNN	On demand
Metabolism	Microsomal stability (hum)	C (N)	>10 000			RF	RF	Weekly
	Microsomal stability (mouse)	C (N)	>10 000			RF	RF	Weekly
	Microsomal stability (rat)	C (N)	>10 000			RF	RF	Weekly
	Hepatocyte stability (rat)	C (N)	>30 000			RF	RF	Weekly
Toxicity	hERG inhibition	C	>10 000			RF	SVM	Weekly
	Ames mutagenicity	C	>10 000			RF	RF	On demand
	CYP inhibition isoforms	C	>10 000			RF	RF	On demand
	Phospholipidosis	C	<1000			SVM	SVM	On demand
	Structure filter tool	Score	n.a.	-	-	-	-	On demand
PhysChem	Solubility (DMSO)	N	>30 ,000			PLS	MTNN	On demand
	Solubility (Powder)	N	<10 000				MTNN	On demand
	logD @ pH 7.5	N	>70 000			PLS	MTNN	On demand
	Membrane affinity	N	<10 000			PLS	MTNN	On demand
	pKa	N	>10 000			ANN	ANN	On demand
	Oral PhysChem score	Score	n.a.	-	-	-	-	On demand
	i.v. PhysChem score	Score	n.a.	-	-	-	-	On demand

Drug Discovery Today

Göller, A.H. et al *Drug Discov. Today* **2020,** *25* (9), 1702-1709.

OCHEM https://ochem.eu




Online chemical database
with modeling environment
v.2.4.32
Welcome, Dear Dr.Tetko! My account Logout
Home Database Models Moderation
A+ a-
FILTERS
SOURCE
Article/Source [select]
Page Table
PROPERTY
Activity/Property [select]
Hide records without property
CONDITIONS
MOLECULE FILTERS
Name / OCHEM ID / Inchi-Key
Similarity/substructure search
Draw a structure and search all the molecules containing it or similar to it
CLICK TO DRAW A STRUCTURE
Molecular mass between and
ADVANCED MOLECULE FILTERS
MISCELLANEOUS
Current set
Show all
Data origin and quality:
Data introducers: All users
Data visibility: Public and private
Data from other users: Only approved data
Basket
Records
Tags
1 - 5 of 1075221
5 items on page 1 of 215045 > >>
molecule profile
logPow = 1.77 (in Log unit)
Charmantray, F et al
4-Hydroxymethyl-3-aminoacridine derivatives as a new family ...
N: 1 P: 970 T: 1
J. Med. Chem. 2003; 46 (6) 967-77
MoleculeID: M11715
Public record
RecordID: R1
09:35, 19 Mar 09 / 10:04, 13 Oct 11
i.tetko / itetko
molecule profile
logPow = 2.1 (in Log unit)
Charmantray, F et al
4-Hydroxymethyl-3-aminoacridine derivatives as a new family ...
N: 2 P: 970 T: 1
J. Med. Chem. 2003; 46 (6) 967-77
MoleculeID: M9560
Public record
RecordID: R2
09:35, 19 Mar 09
i.tetko
molecule profile
logPow = 2.2 (in Log unit)
Charmantray, F et al
4-Hydroxymethyl-3-aminoacridine derivatives as a new family ...
N: 3 P: 970 T: 1
J. Med. Chem. 2003; 46 (6) 967-77
MoleculeID: M10111
Public record
RecordID: R3
09:35, 19 Mar 09
i.tetko
molecule profile
logPow = 1.64 (in Log unit)
Charmantray, F et al
4-Hydroxymethyl-3-aminoacridine derivatives as a new family ...
N: 4 P: 970 T: 1
J. Med. Chem. 2003; 46 (6) 967-77
MoleculeID: M9777
Public record
RecordID: R4
09:35, 19 Mar 09
i.tetko

Traditional representation of chemical structures

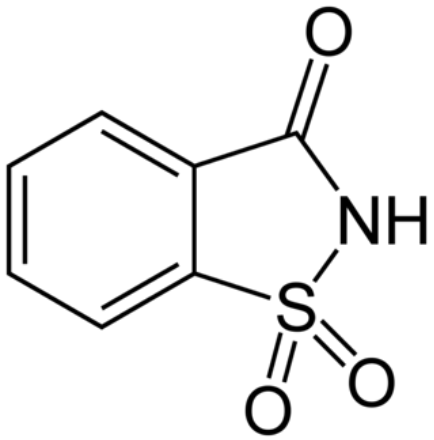
Saccharin

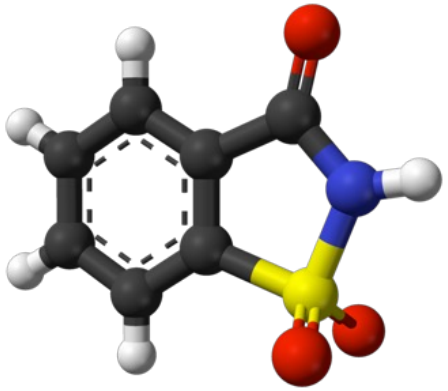
1D

2D

3D

$C_7H_5NO_3S$





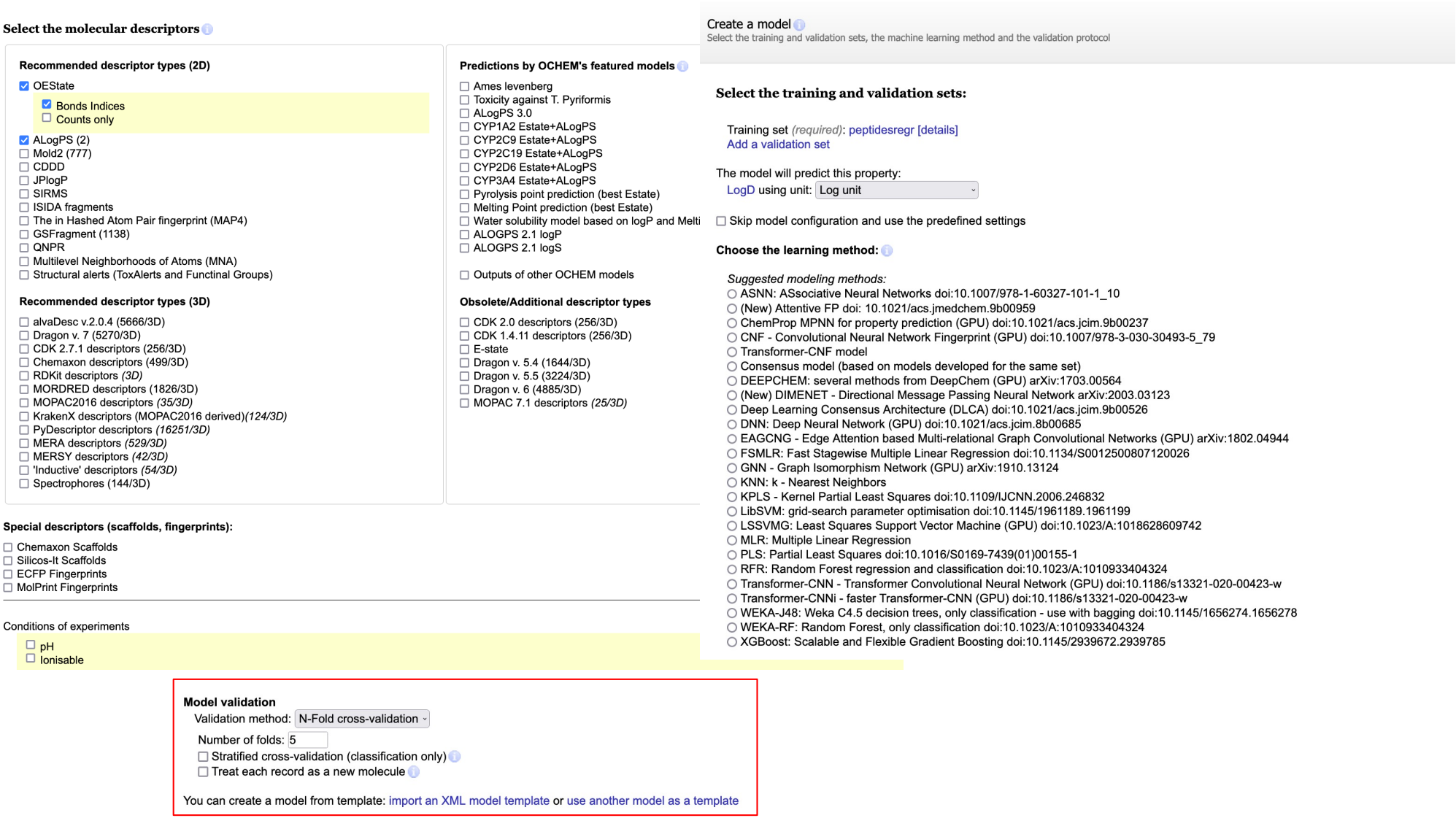
Examples of descriptors

- [x] alvaDesc v.2.0.4 (5666/3D)

[select all] [select none] [select 3D] [unselect 3D]

- [x] Constitutional descriptors (50)
- [x] Topological indices (79)
- [x] Connectivity indices (37)
- [x] 2D matrix-based descriptors (608)
- [x] Burden eigenvalues (96)
- [x] ETA indices (40)
- [x] Geometrical descriptors (3D, 38)
- [x] 3D autocorrelations (3D, 80)
- [x] 3D-MoRSE descriptors (3D, 224)
- [x] GETAWAY descriptors (3D, 273)
- [x] Functional group counts (3D, 154)
- [x] Atom-type E-state indices (346)
- [x] 2D Atom Pairs (1596)
- [x] Charge descriptors (3D, 15)
- [x] Drug-like indices (30)
- [x] WHALES (3D, 33)
- [x] Chirality (70)
- [x] Ring descriptors (35)
- [x] Walk and path counts (46)
- [x] Information indices (51)
- [x] 2D autocorrelations (213)
- [x] P_VSA-like descriptors (69)
- [x] Edge adjacency indices (324)
- [x] 3D matrix-based descriptors (3D, 132)
- [x] RDF descriptors (3D, 210)
- [x] WHIM descriptors (3D, 114)
- [x] Randic molecular profiles (3D, 41)
- [x] Atom-centred fragments (115)
- [x] Pharmacophore descriptors (165)
- [x] 3D Atom Pairs (3D, 36)
- [x] Molecular properties (3D, 27)
- [x] CATS 3D (3D, 300)
- [x] MDE (19)

QSPR/QSAR modelling in OCHEM



Each descriptor re-presentation sees only part of molecules

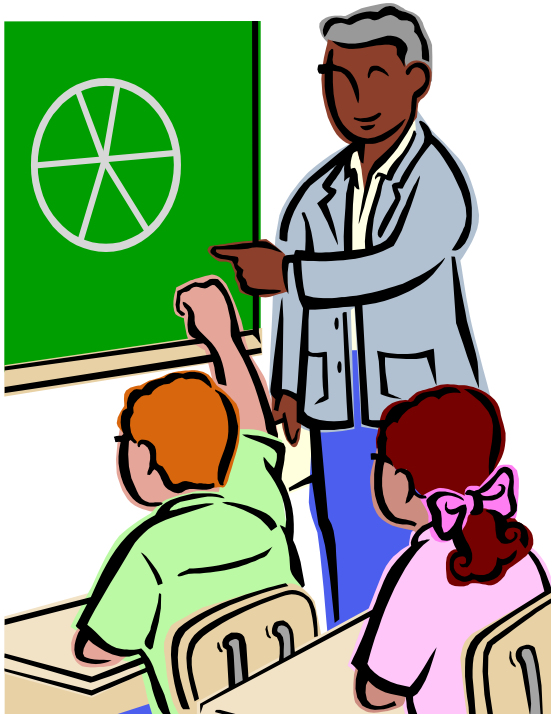
Blind monks examining an elephant, an ukiyo-e print by Hanabusa Itchō (1652–1724).
https://en.wikipedia.org/wiki/Blind_men_and_an_elephant

Consensus modelling

Best method(s) are defined

Average prediction of models is used

The consensus prediction is more accurate and stable



Key Links

CompTox Home
Basic Information
Organization
EPA Exposure Research
Research Projects
Chemical Databases
ToxCast Stakeholder Events
EPA Chemical Safety Research
Research Publications
Scientific Reviews
Communities of Practice
ToxCast Data Challenges
Staff Profiles
CompTox Partners
Jobs and Opportunities

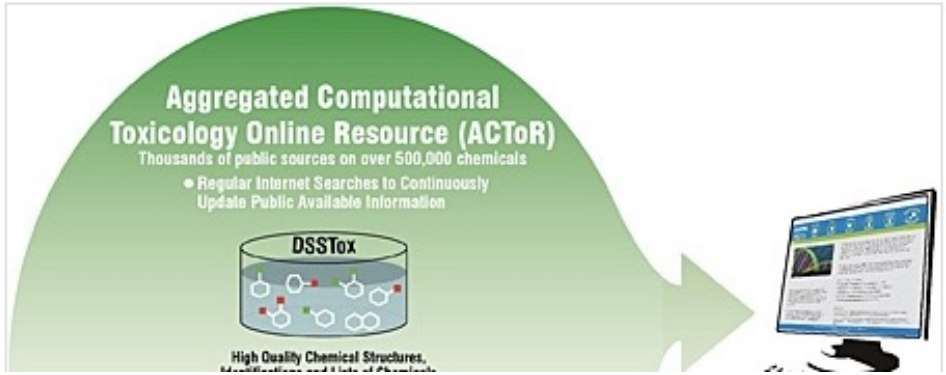
ToxCast Chemical Data Challenges and Release

EPA's high-throughput screening data on 1,800 chemicals is accessible through the interactive Chemical Safety for Sustainability Dashboards (iCSS dashboard). The iCSS dashboard provides user-friendly and customizable access to toxicity data from ToxCast and Tox21 high-throughput chemical screening technologies.

Using the TopCoder and InnoCentive crowd-sourcing platform, EPA invited the science and technology community to work with the data and provide solutions for how the new toxicity data can be used to predict potential health effects. The ToxCast data challenges focused on using this data and other publicly available data to predict the lowest effect level from traditional toxicity studies using laboratory animals. Challenge winners received awards for solving this challenge.

Key Links

- Lowest Effect Level Challenge Results (PDF, 497KB, 18pp)
- Chemical Safety for Sustainability Dashboards
- Complete ToxCast Phase II Data & Files
- TopCoder Challenge
- InnoCentive Challenge
- Stakeholder Workshops



ToxCast Challenge

- OCHEM model (by Dr. S. Novotarskyi) got the the first position for the US Environmental Protection Agency (EPA) challenge in May 2014

- Prediction of systemic Lowest Effect Level (LEL)
  - lowest dose that shows adverse effects in animal toxicity tests

- Challenge: build a prediction model using data from high-throughput ***in vitro*** assays provided by EPA to quantitatively predict a chemical's systemic LEL.

http://www.epa.gov/ncct/challenges.html

Novotarskyi, S. et al. *Chem. Res. Toxicol*. 2016, 29, 768-75.

ToxCast Challenge SetUp

- Training set of 483 compounds
- Test set of 1371
  - Leader board: 63
  - Private set: 80

EPA in vitro assays data were also provided (not used in Top I solution)

Model: Associative Neural Network (training with descriptor set optimization)

Novotarskyi, S. et al. *Chem. Res. Toxicol*. 2016, 29, 768-75.

TOC

Lowest Effect Level (LEL)
ToxCast EPA prediction challenge

in vitro + *in silico* → *in vivo* ≈ *in silico* → *in vivo*

data upload, descriptors calculation, modeling, consensus, Rank-I submission, on-line available

Novotarskyi, S. et al. *Chem. Res. Toxicol*. 2016, 29, 768-75.

Table 1. Number of Descriptors and Models' Accuracy for the Prediction of the Test Set Compounds

		RMSE		
descriptor set	number of selected descriptors	whole test set (n = 143)	inside of AD[a] (n = 136)	outside of AD (n = 7)
CDK	159	1.13	1.01	2.4
Dragon	1824	1.15	1.05	2.4
Fragmentor	631	1.18	1.04	2.7
GSFrag	202	1.1	0.97	2.5
Mera, Mersy	242	1.04	0.96	2.1
Chemaxon	97	1.16	1.06	2.4
Inductive	39	1.17	1.03	2.7
Adriana	133	1.14	1.01	2.5
QNPR	381	1.12	1.02	2.7
E-state	185	1.16	1	2.8
in vitro	143	1.21	1.11	2.5
Consensus	4036	1.08	0.96	2.5

[a]AD is the applicability domain of the model as defined by OCHEM[8] (see also ref 20).

Statistical uncertainty

Table 2. Summary of the Performance of the Top-Ranked Models of the EPA ToxCast Challenge

			test set					
	training set (n = 483)[a]		provisional subset (n = 63)		final subset (n = 80)			full, n = 143
model	RMSE	R^2	RMSE	rank	RMSE	R^2	rank	RMSE
novserj	0.88 ± 0.04	0.27 ± 0.04	1.03 ± 0.08[b]	8	1.12 ± 0.08[b]	0.31	1	1.08 ± 0.07
NobuMiu			1.03	9	1.13	0.30	2	1.09
a9108tc			1.05	16	1.13	0.29	3	1.10
klo86 min			1.09	27	1.14	0.29	4	1.12
in vitro assays[c]	0.97 ± 0.04	0.11 ± 0.03						1.24 ± 0.09
MW + NC[d]	0.97 ± 0.04	0.11 ± 0.03						1.18 ± 0.08

[a]Prediction accuracy for the "out-of-the-bag" samples. [b]Confidence intervals were estimated using the subsets, which were sampled from the training set, and each had the same size as the respective test set (see for more details ref 23). [c]Best model based on the *in vitro* assay descriptors developed using the LibSVM method (see also Table S1). [d]Model based on molecular weight (MW) and number of carbon atoms (NC) developed using the same approach as the above *in vitro* model.

Novotarskyi, S. et al. *Chem. Res. Toxicol*. 2016, 29, 768-75.

National Center for Advancing Translational Sciences

Tox21 Data Challenge 2014

Login

» Home About Registration Data/Resources Submissions Discussion Leaderboard Contact Us Survey NEW

Open call ends: November 14, 2014

About the Data

The Challenge

The 2014 Tox21 data challenge is designed to help scientists understand the potential of the chemicals and compounds being tested through the Toxicology in the 21st Century initiative to disrupt biological pathways in ways that may result in toxic effects.

The goal of the challenge is to "crowdsource"

All challenge winners will receive the opportunity to submit a paper for publication in a special thematic issue of *Frontiers in Environmental Science* and recognition on the NCATS website and via social media.

Best Balanced accuracy - Abdelaziz, A. et al. *Front. Environ. Sci.* 2016, 4, 2.

Challenge setup

Subchallenge Overview

Subchallenges 1-12

Predict the compound activity outcome (active or inactive) in one or more of the 12 pathway assays based on the chemical structure information for the following assays:

- estrogen receptor alpha, LBD (ER, LBD)
- estrogen receptor alpha, full (ER, full)
- aromatase
- aryl hydrocarbon receptor (AhR)
- androgen receptor, full (AR, full)
- androgen receptor, LBD (AR, LBD)
- peroxisome proliferator-activated receptor gamma (PPAR-gamma)
- nuclear factor (erythroid-derived 2)-like 2/antioxidant responsive element (Nrf2/ARE)
- heat shock factor response element (HSE)
- ATAD5
- mitochondrial membrane potential (MMP)
- p53

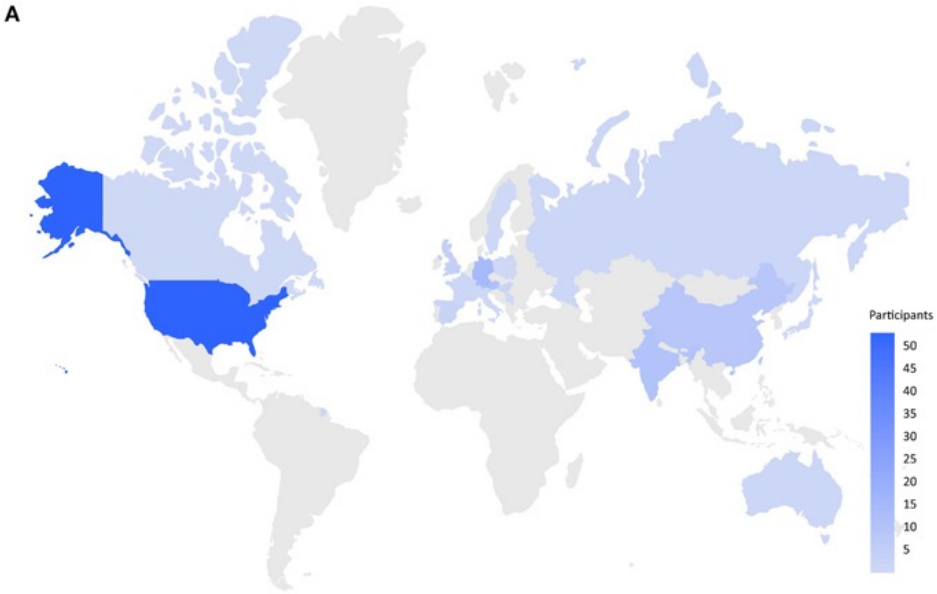
Grand Challenge (All 12)

Subchallenge 13 (all nuclear receptor)

Subchallenge 14 (all stress response pathways)

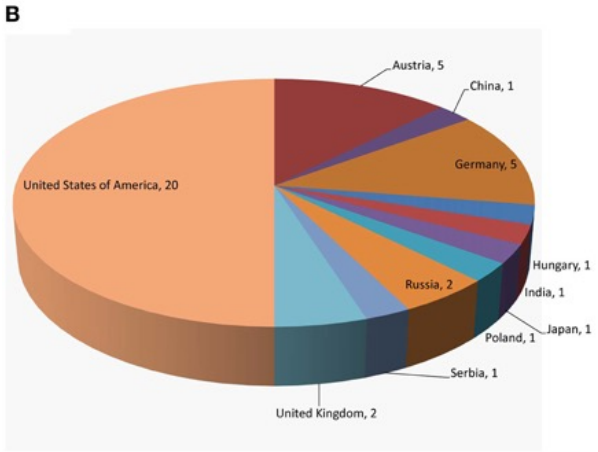
Set	Class	AhR	AR	AR-LBD	ARE	Aromatase	ATAD5	ER	ER-LBD	HSE	MMP	p53	PPAR.g
Train	Inactive	7219	8982	8296	6069	6866	8753	6760	8307	7722	6178	8097	7962
Train	Active	950	380	303	1098	360	338	937	446	428	1142	537	222
Leader	Inactive	241	289	249	186	196	247	238	277	257	200	241	252
Leader	Active	31	3	4	48	18	25	27	10	10	38	28	15

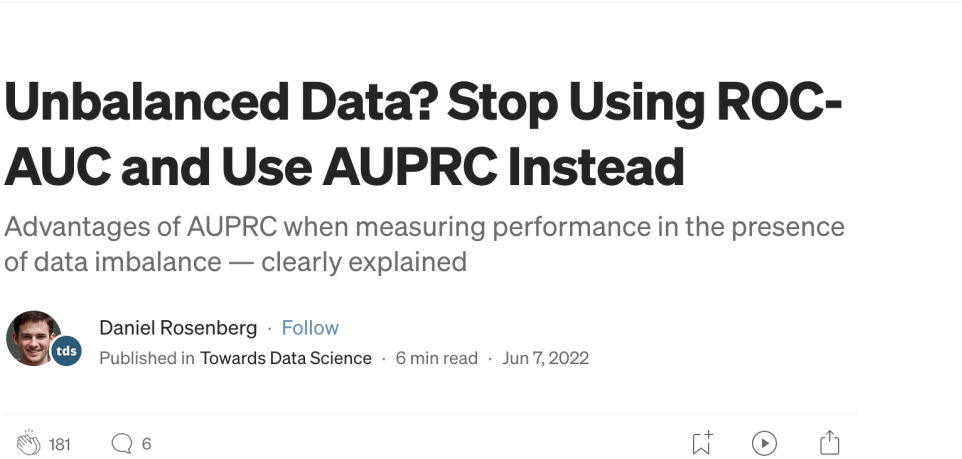
Tox21 Challenge winners



ROC-AUC:
Mayr et al
DeepTox: Multi-task deep neural network
https://doi.org/10.3389/fenvs.2015.00080

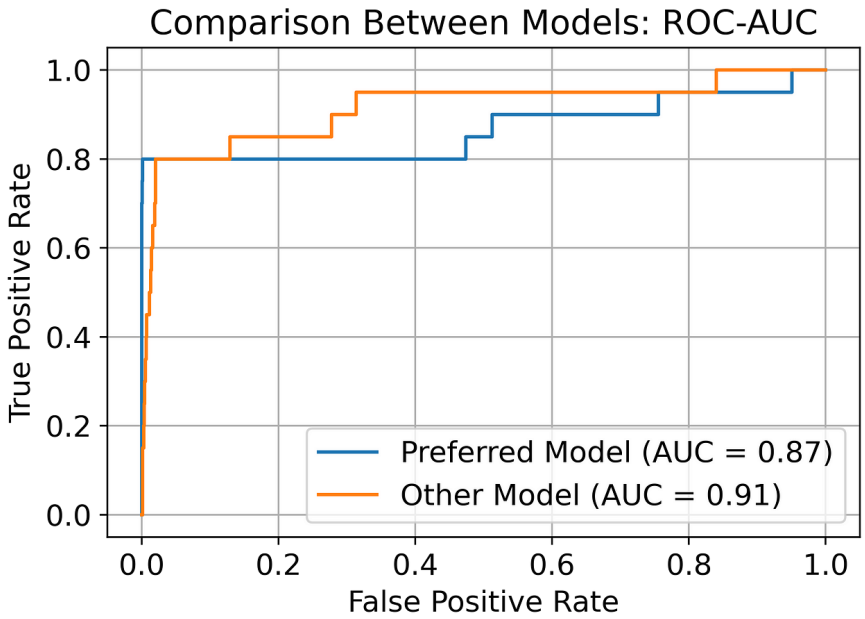
Best balanced accuray:
Abdelaziz et al.
ASNN: Associative neural network
https://doi.org/10.3389/fenvs.2016.00002

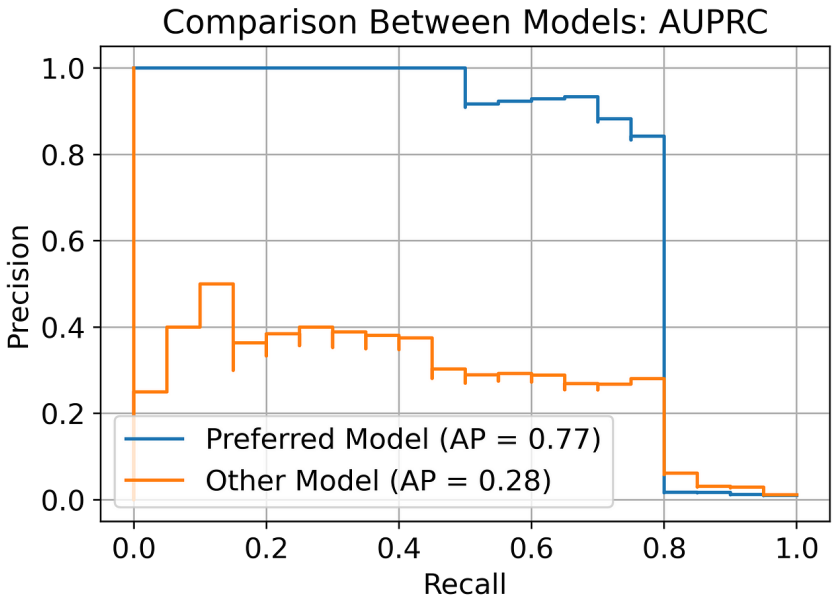




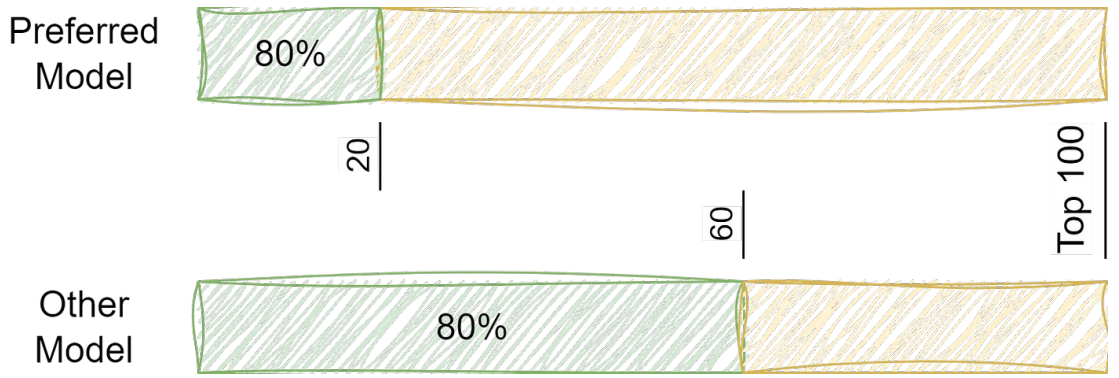
https://towardsdatascience.com/imbalanced-data-stop-using-roc-auc-and-use-auprc-instead-46af4910a494

Area Under the Precision-Recall Curve (AUPRC)





20 positives
200 negatives



https://towardsdatascience.com/imbalanced-data-stop-using-roc-auc-and-use-auprc-instead-46af4910a494

Winning model: OCHEM-generated consensus model

Andrea Kopp (Hunklinger)

SLAS Europe 2023

25.05.2023

HELMHOLTZ MUNICH

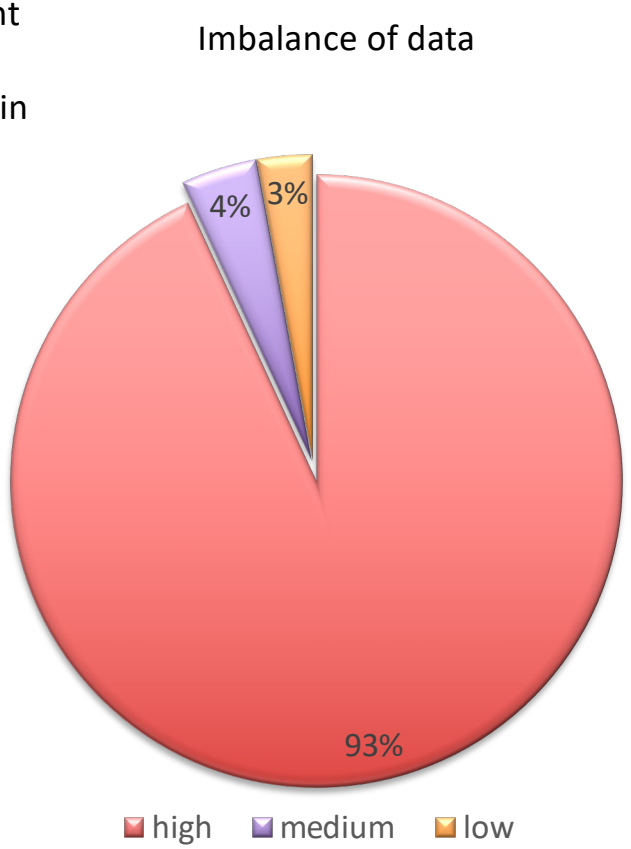
Together with Peter Hartog, Martin Šícho and Guillaume Godin

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

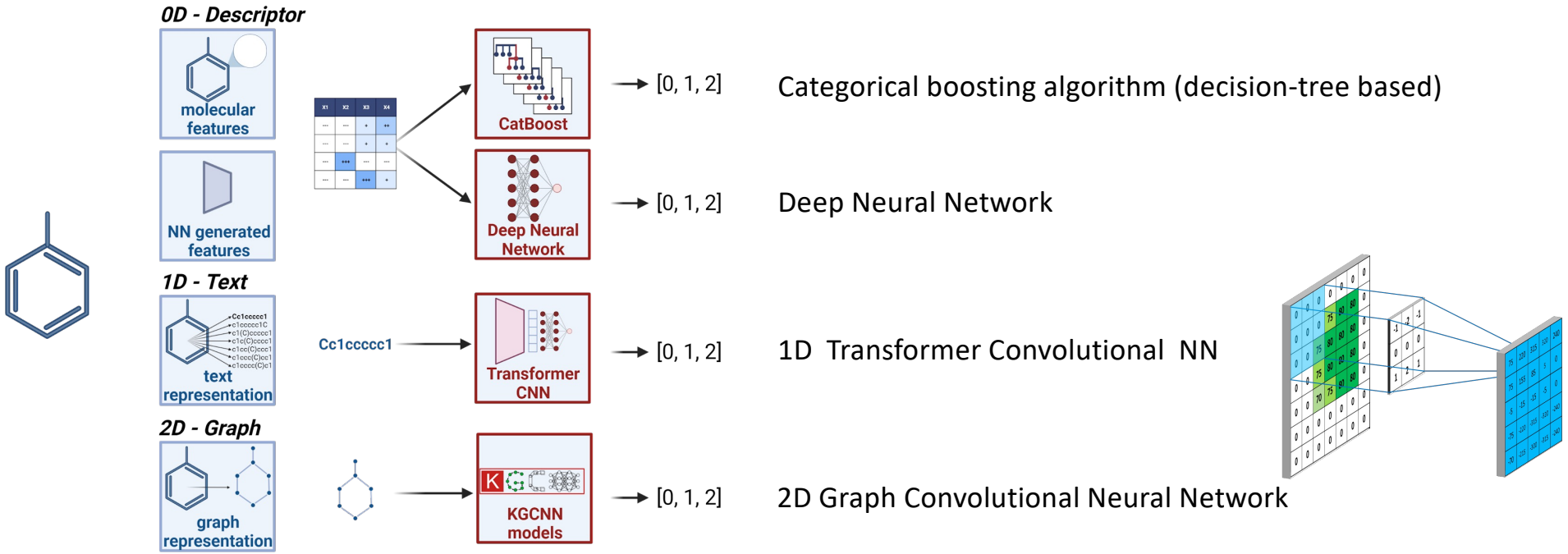
Hunklinger at al, DOI: 10.1016/j.slasd.2024.01.005

Solubility challenge set-up

- Experimentally: Nephelometer measures undissolved sediment
- Classification into *low*, *medium* and *high* soluble with phenytoin and amiodarone as thresholds
- 70k training datapoints, 15k public leaderboard, 15k private leaderboard
- Stratified random sampling

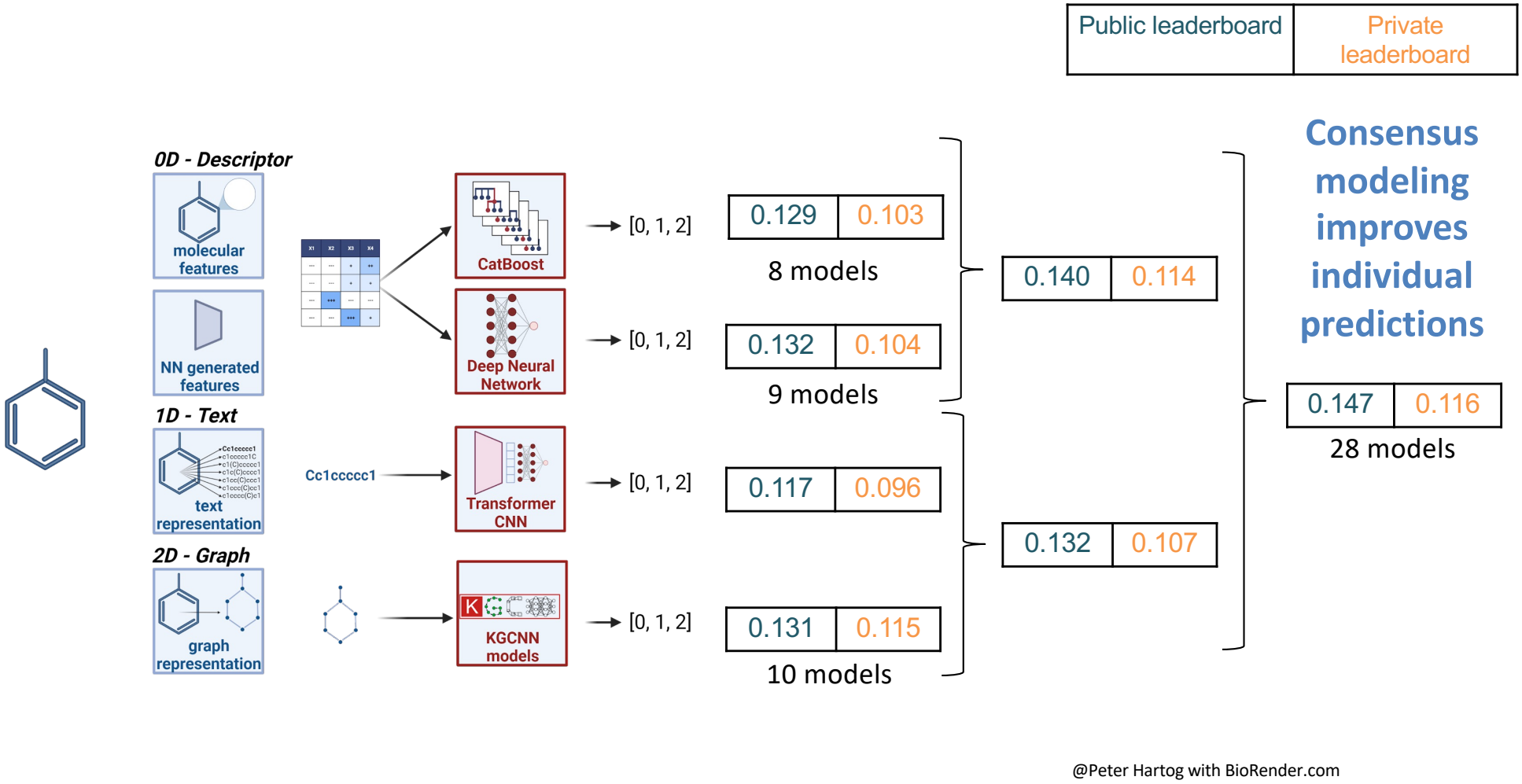


Molecular representation



Hunklinger at al, DOI: 10.1016/j.slasd.2024.01.005

Quadratic kappa metric scores



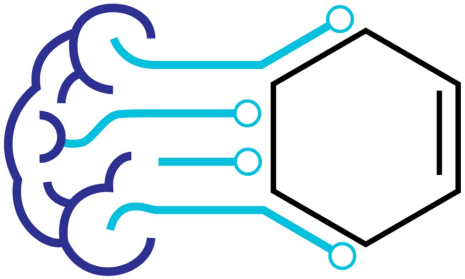
Hunklinger at al, DOI: 10.1016/j.slasd.2024.01.005

ICANN24

33rd International Conference on Artificial Neural Networks

Tox24 Challenge: How accurately can we predict binding to transthyretin?

Start:17/05 » Submit:31/08 » Winner:18/09 » Article:31/12



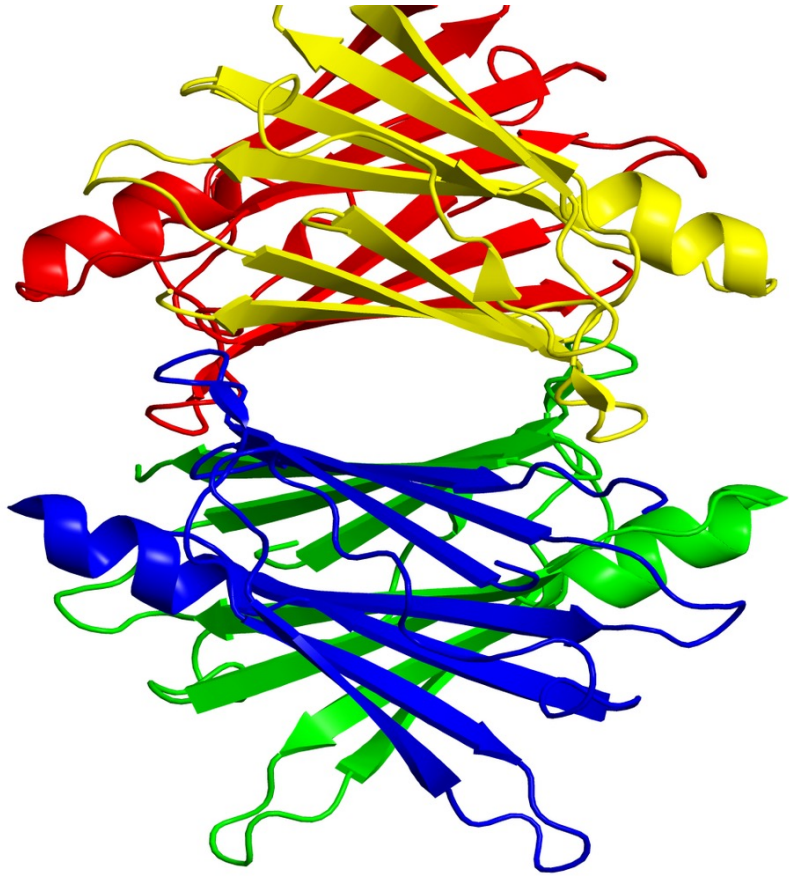
https://ochem.eu

https://e-nns.org/icann2024

Transthyretin binding and EDC

TTR is one of the serum binding proteins responsible for delivering thyroid hormones (THs) to target tissues and maintaining the balance of free versus bound THs.

The binding of compounds to TTR and subsequent displacement of TH is important to identify potential interference of the thyroid system which are endocrine disrupting chemicals (EDCs).



https://en.wikipedia.org/wiki/Transthyretin

Tox24 Challenge

EPA submitted an article in February with screening of TTR compounds

March-April: negotiation with EPA, ChemResTox, ICANN2024, AIDD to organize the challenge

May 17th – data are publicly available

- 1012 training set
- 200 LeaderBoard set
- 300 Blind Test set

August 15th

- LeaderBoard set is available

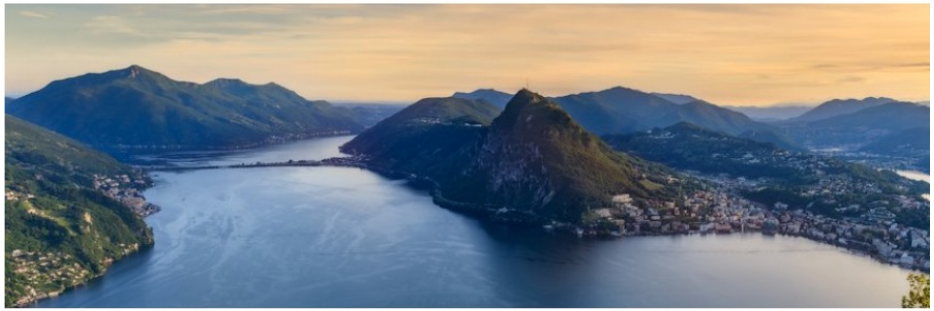
September 1st

- Results announced

ICANN24

33rd International Conference on Artificial Neural Networks

Home **Conference** Registration Contributors Organisation T



Tox24 Challenge

The **Tox24** challenge is designed to assess the progress in computational methods for predicting in vitro activity of compounds. All ML experts are strongly invited to participate in it and compete for a prize of **1000€**, to be awarded for the winning model.

Tetko, I. V. Tox24 Challenge. *Chem. Res. Toxicol.* **2024**, *37* (6), 825–826. https://doi.org/10.1021/acs.chemrestox.4c00192.

Tox24 Challenge results

Tox24 final ranking based on the blind set - congratulations to the winning team Amidoff!
In case of equal performances the earlier entry was used. Models with lower RMSE (Root Mean Squared Error) are better.
You can access the winning model here

rank	User	submission id	*** Blind test set RMSE ***	Leaderboard set RMSE	Submission time (GMT-12)
1	Amidoff	1148	20.5	7.8	25-08-2024 06:13:02
2	tcirino	1098	20.7	4.1	21-08-2024 08:53:04
3	znavoyan	1309	20.7	5.7	30-08-2024 07:16:19
4	uesawa	1149	20.8	21.2	25-08-2024 14:03:09
5	YingkaiZhangLab	1388	20.8	20.7	31-08-2024 06:49:33
6	AntonijaBoss	1400	21.2	20.6	31-08-2024 09:07:16
7	SankalpJain	1284	21.3	5.1	30-08-2024 02:46:16
8	alx.dga	1097	21.4	0.0	21-08-2024 08:45:46
9	luispintoc	1209	21.4	15.0	28-08-2024 06:00:04
10	GAT_Wang	1253	21.4	6.0	29-08-2024 15:26:28
11	vchupakhin	1404	21.4	3.2	31-08-2024 13:34:03

Winning model (Makarov, D.; Ksenofontov, A.)

Model name: Consensus TTR binding activity , published in Tox24 Challenge
Public ID is 1149

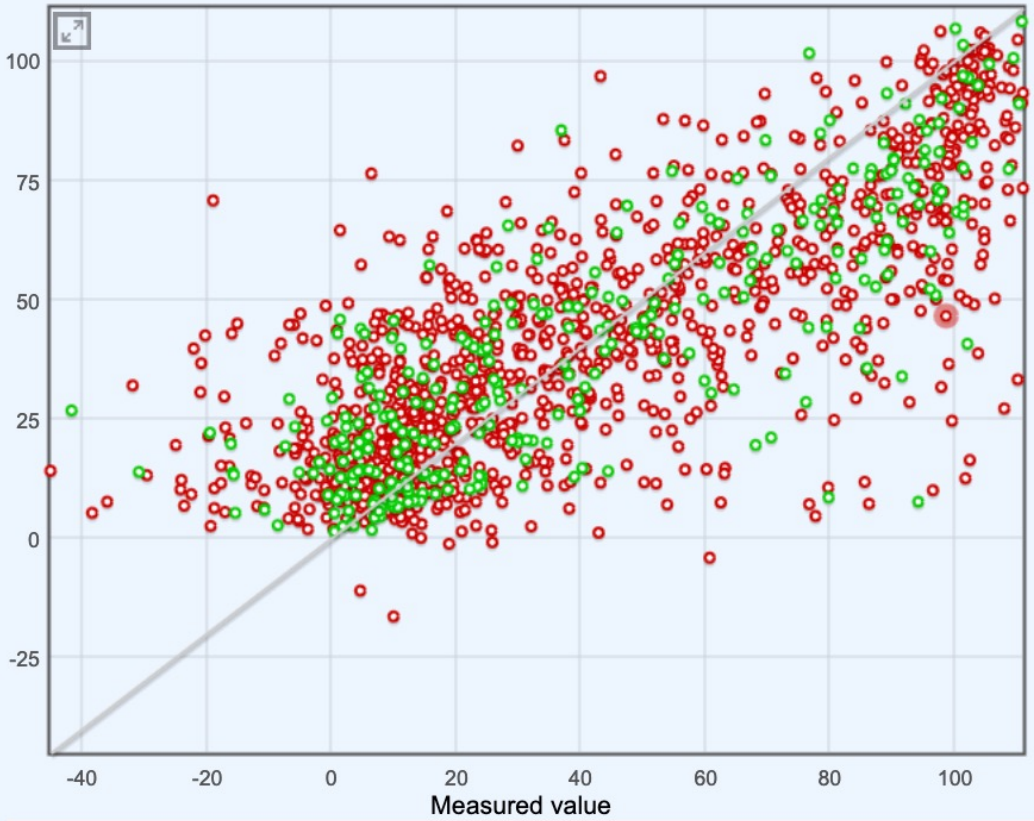
Predicted property: **TTR binding activity** modeled in %
Training method: Consensus

4 individual models:
ModelID: 1145
ModelID: 1146
ModelID: 1147
ModelID: 1148
AVERAGE

No validation

Size: 1 Kb

Data Set	#	R2	q2	RMSE	MAE
Training set: TTR_train	1208 records	0.62 ± 0.02	0.62 ± 0.02	22.3 ± 0.6	16.6 ± 0.4
Test set: tox24_challenge_test_data [x]	300 records	0.67 ± 0.04	0.67 ± 0.03	21 ± 1	15.6 ± 0.8



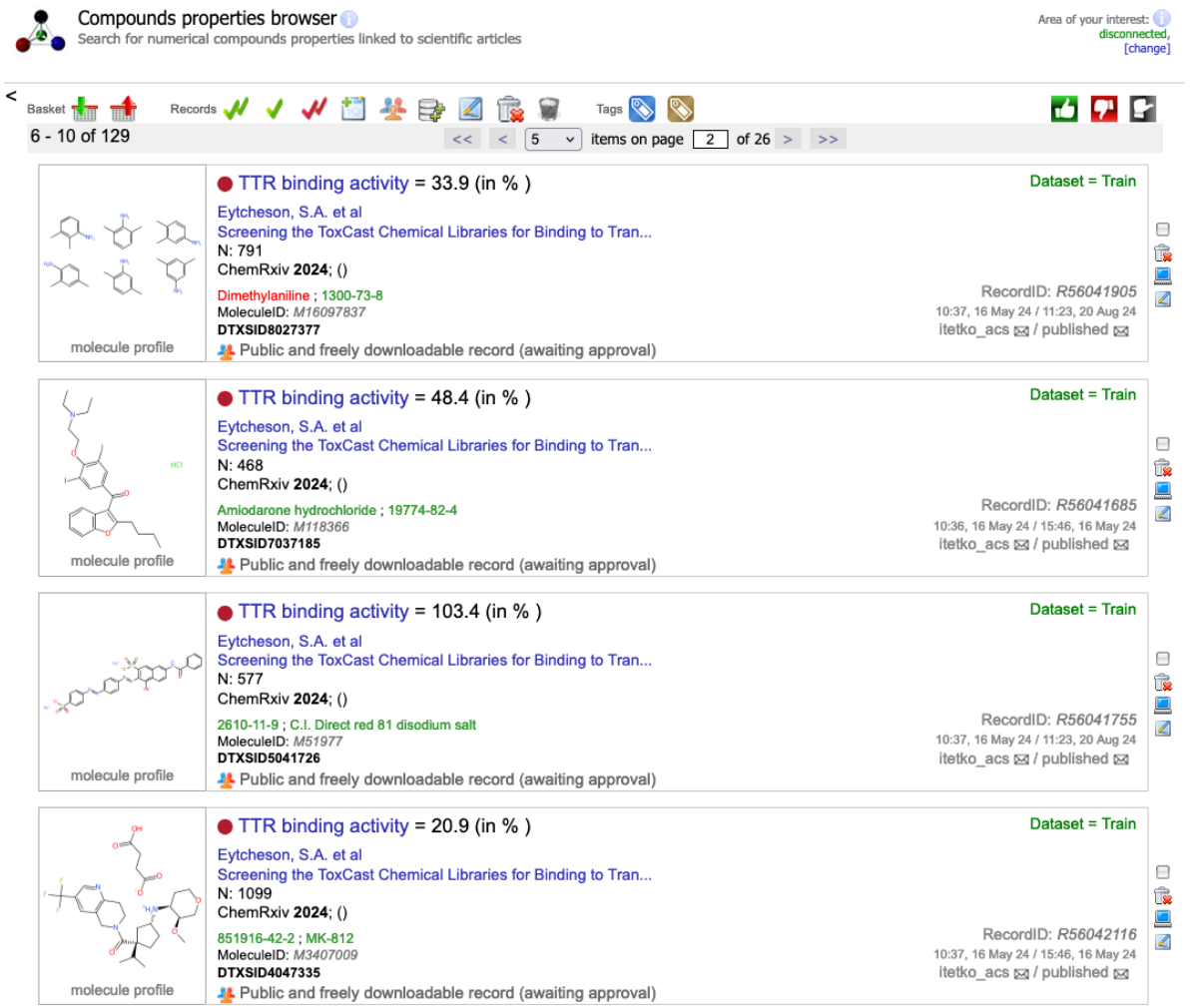
Number of compounds ignored because of errors in original model = 4

Winning model (Makarov, D.; Ksenofontov, A.)

Consensus model:

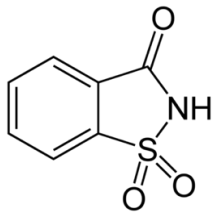
- Transformer CNN
- Transformer CNF2
- Cat Boost based on Mold2 descriptors
- Cat Boost based on ALOGPS + OESTATE descriptors

And using mixture descriptors



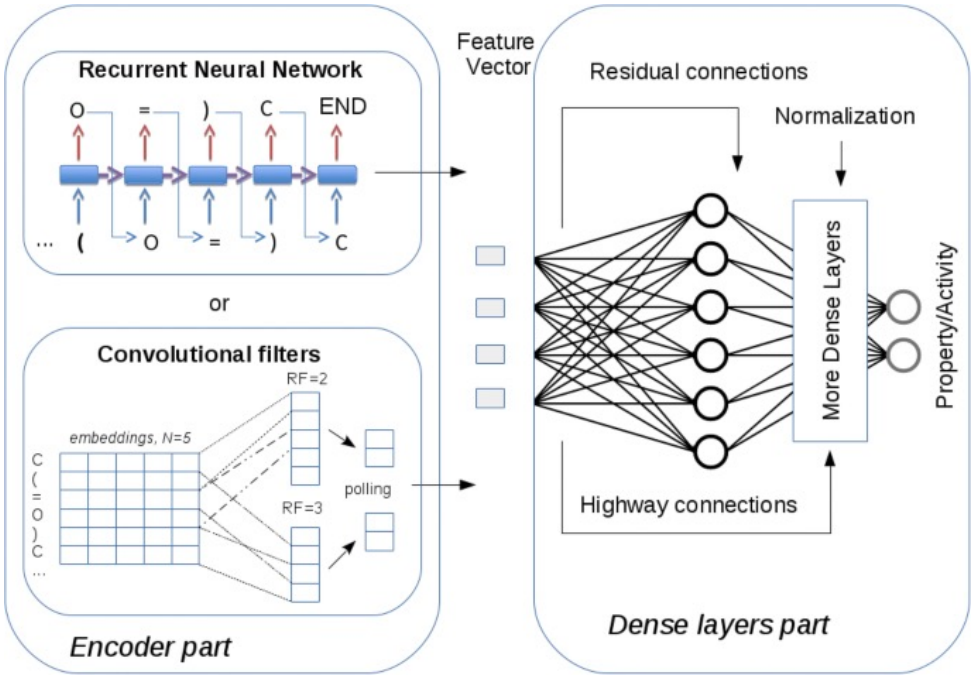
Machine Learning directly from chemical structures

Saccharin: c1ccc2c(c1)C(=O)NS2(=O)=O

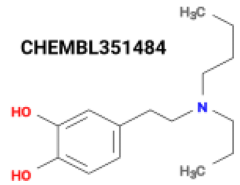


Text processing: convolutional neural networks, transformers, LSTM

Graph processing: message passing neural networks



non-canonical		canonical
c1c(ccc(O)c1O)CCN(CCCC)CCC	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
c1(ccc(c(c1)O)O)CCN(CCC)CCCC	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
c1c(c(cc(c1)CCN(CCCC)CCC)O)O	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
c1(CCN(CCCC)CCC)ccc(O)c(O)c1	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
CCCCN(CCc1cc(c(cc1)O)O)CCC	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
CCCCN(CCc1ccc(c(c1)O)O)CCC	>>	**CCCCN(CCc1ccc(c(c1)O)O)CCC**
N(CCCC)(CCc1ccc(c(O)c1)O)CCC	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
C(N(CCc1ccc(O)c(c1)O)CCCC)CC	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
N(CCc1ccc(c(O)c1)O)(CCCC)CCC	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
c1c(O)c(ccc1CCN(CCC)CCCC)O	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC
c1(c(cc(cc1)CCN(CCC)CCCC)O)O	>>	CCCCN(CCc1ccc(c(c1)O)O)CCC

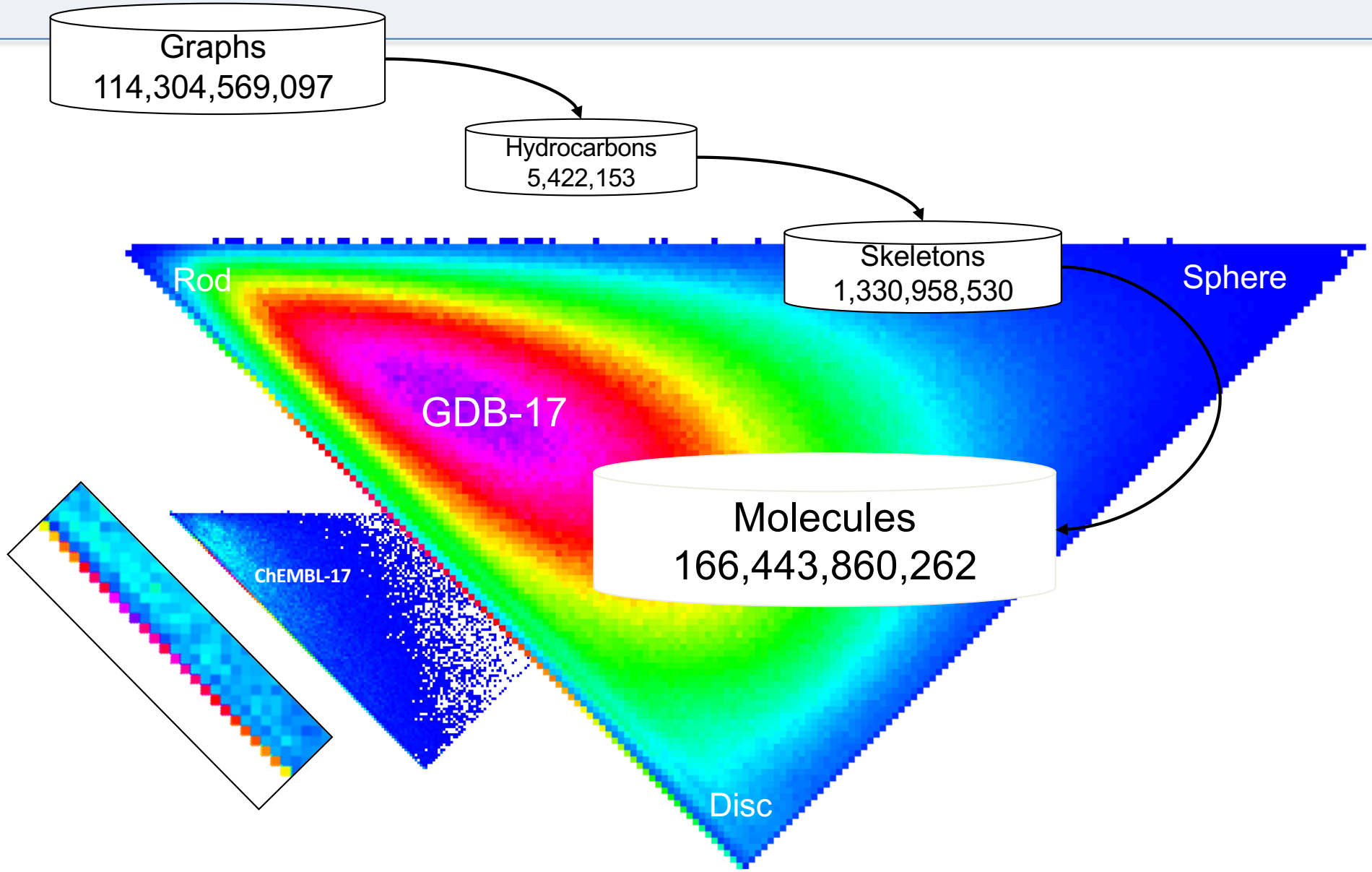


CheMBL database was used for pretraining

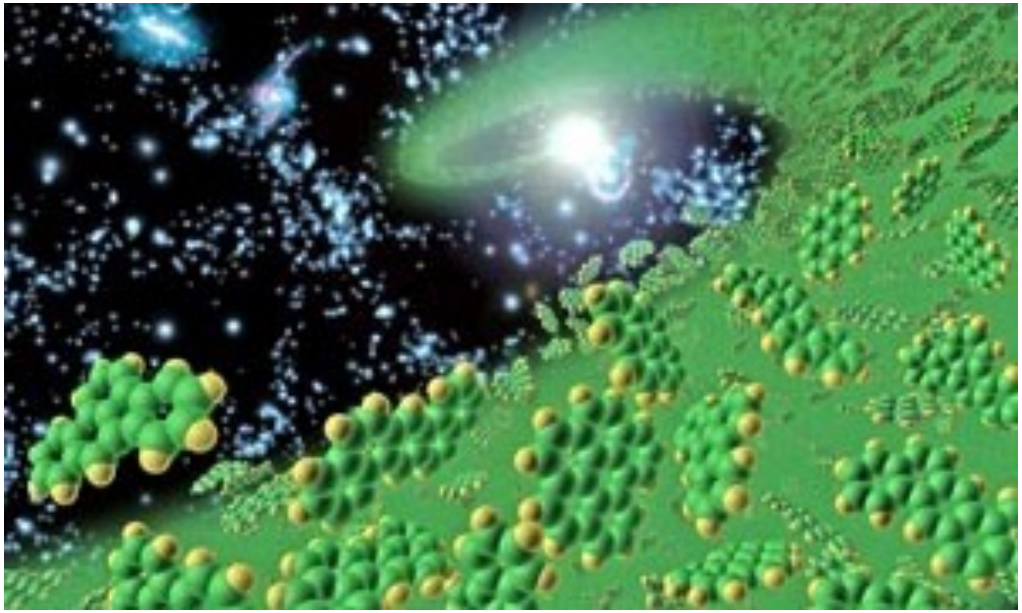
Image augmentation

https://github.com/aleju/imgaug

GDB-17



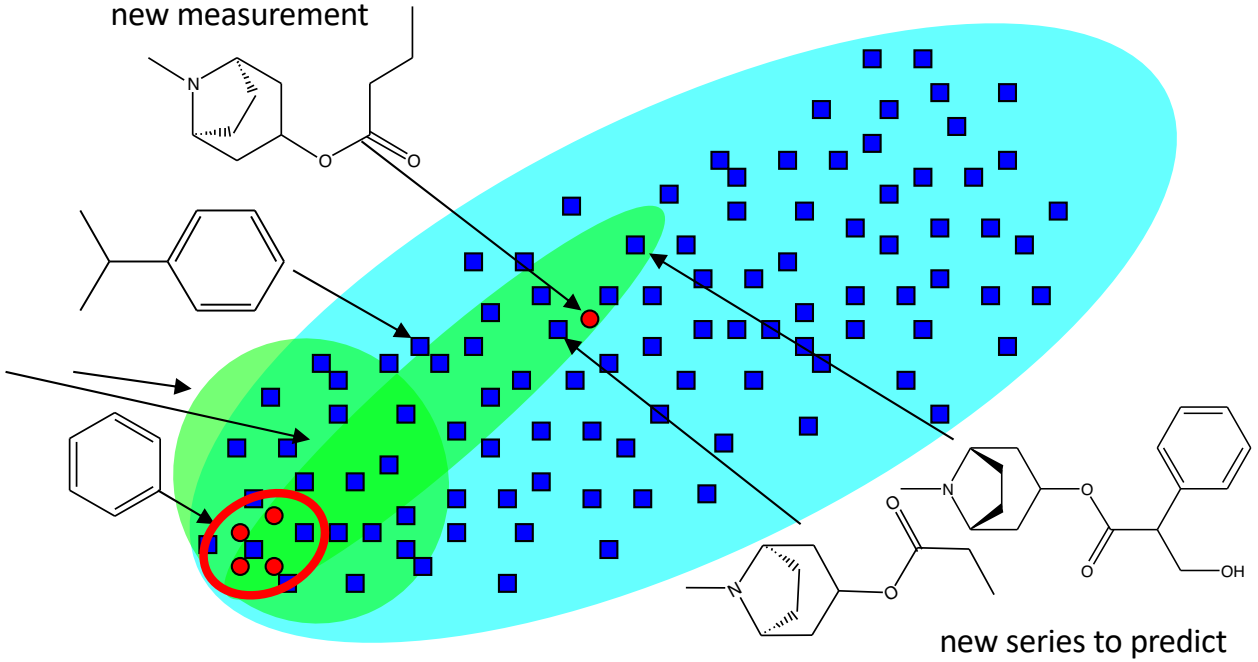
L. Ruddigkeit et al., *J. Chem. Inf. Model.* 2012, *52*, 2864-2875 (GDB-17)



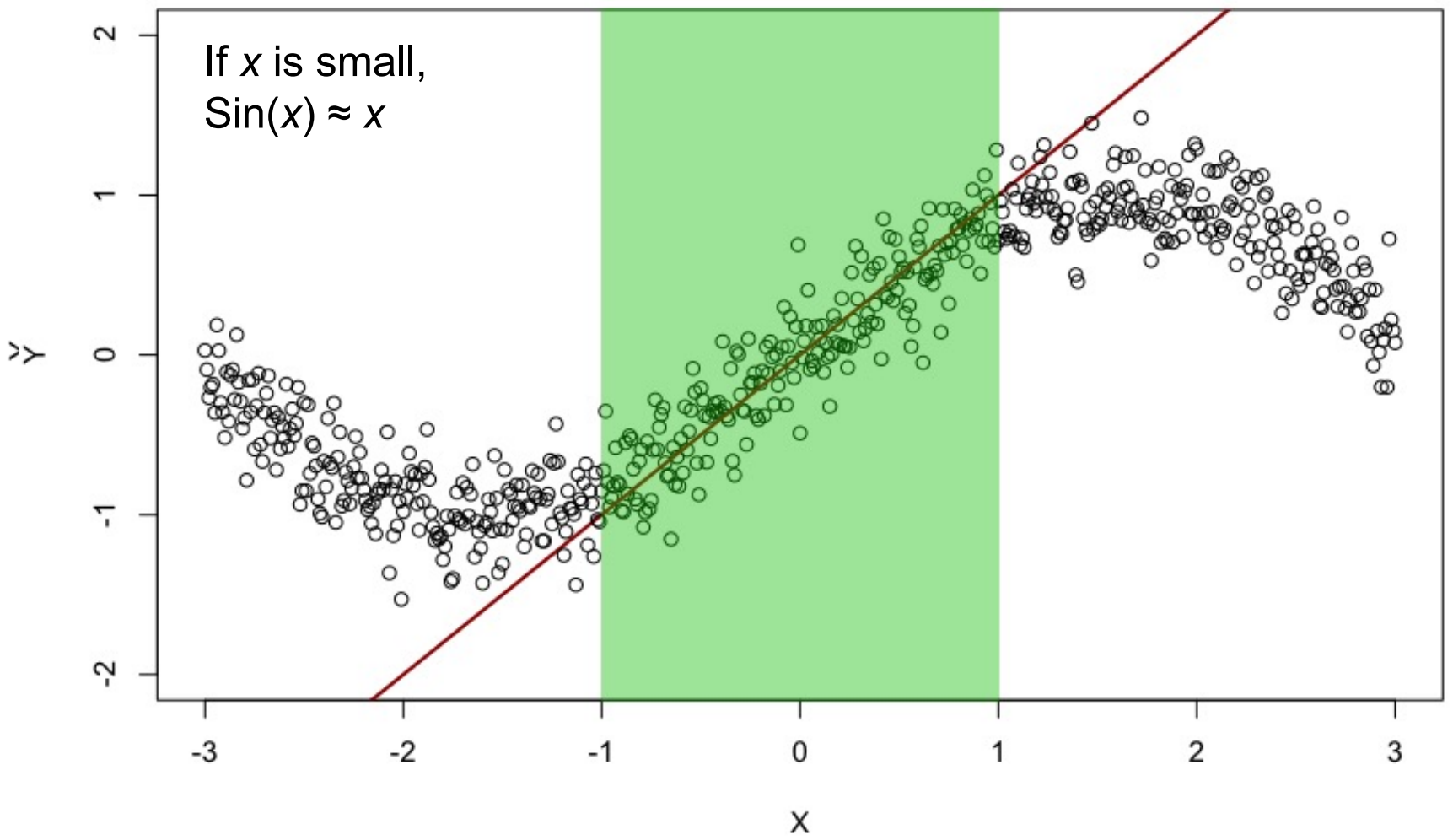
But: we can't predict unpredictable!

10^5 (Soulbility set) → 10^{11} (GDB17)

1 → 1,000,000



Accuracy of prediction



Overview of analyzed distances to models (DMs)

EUCLID

$$EU_m = \frac{\sum_{j=1}^{k} d_j}{k}$$

$$EUCLID = \overline{EU}_m$$

k is number of nearest neighbors, m index of model

TANIMOTO

$$Tanimoto(a,b) = \frac{\sum x_{a,i}x_{b,i}}{\sum x_{a,i}x_{a,i} + \sum x_{b,i}x_{b,i} - \sum x_{a,i}x_{b,i}}$$

$x_{a,i}$ and $x_{b,i}$ are fragment counts

LEVERAGE

$$LEVERAGE = \mathbf{x}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}$$

PLSEU (DModX)

Error in approximation (restoration) of the vector of input variables from the latent variables and PLS weights.

STD

$$STD = \frac{1}{N-1}\sum(y_i - \bar{y})^2$$

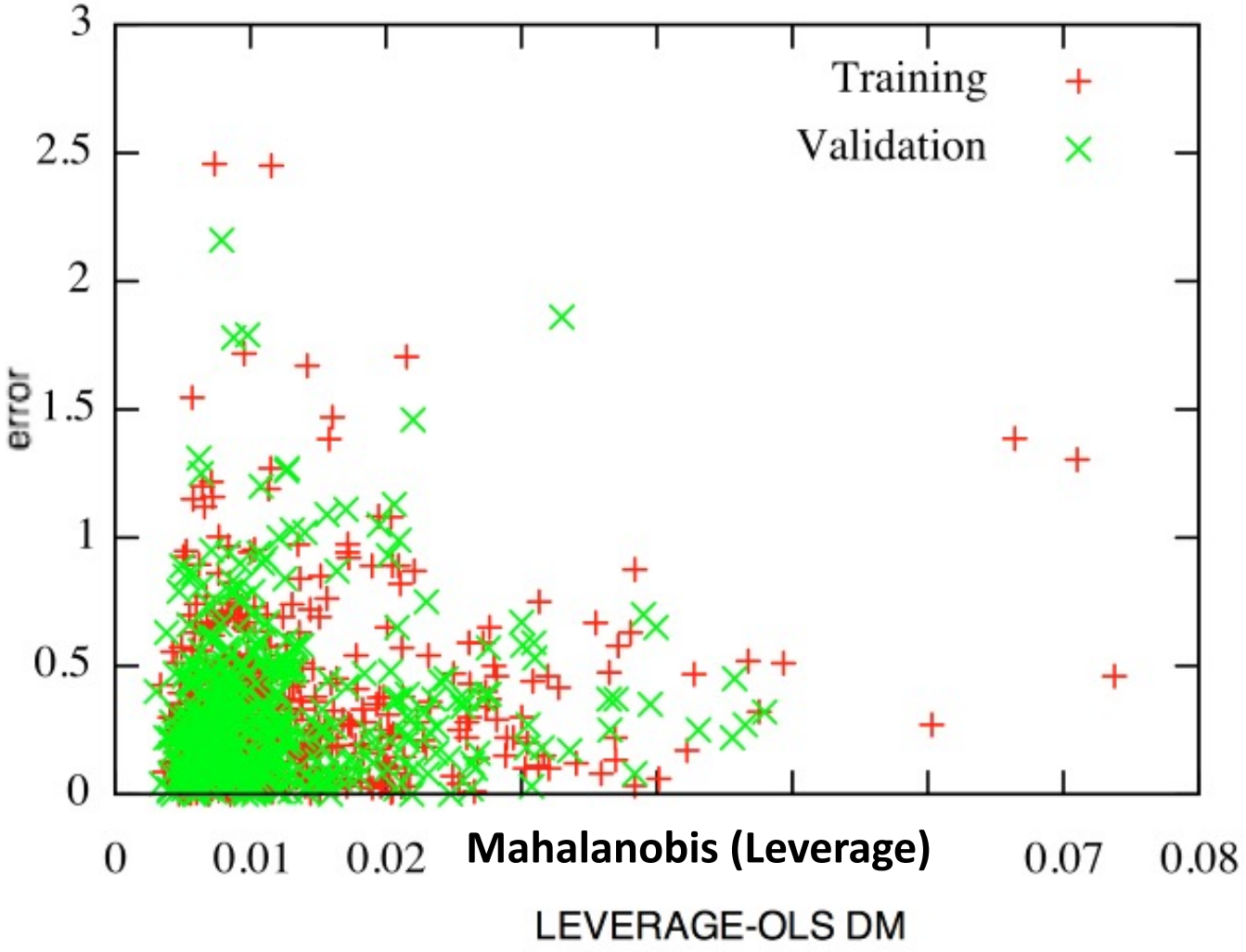
y_i is value calculated with model i and $\bar{y}$ is average value

CORREL

$$CORREL(a) = \max_j CORREL(a,j) = R^2(\mathbf{Y}^a_{calc}, \mathbf{Y}^j_{calc})$$

$\mathbf{Y}^a = (y_1, \ldots, y_N)$ is vector of predictions of molecule i

Descriptor space, ASNN model: DM does not work



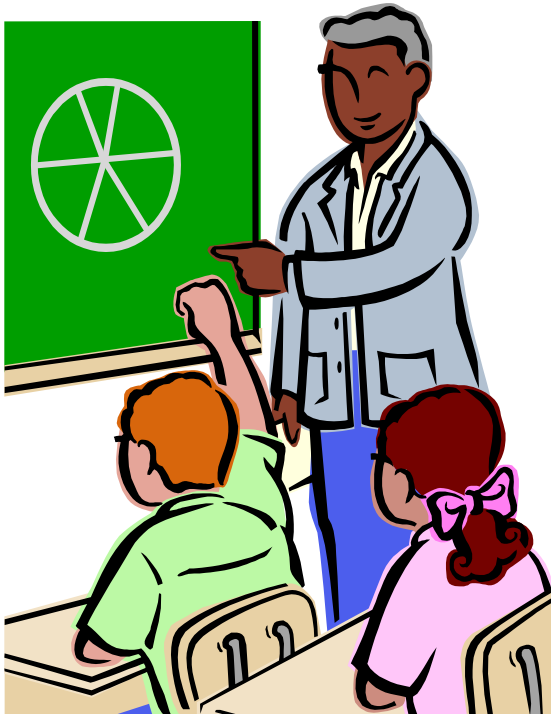
Tetko et al, *J. Chem. Inf. Model*, **2008,** *48*, 1733-46.

Consensus modelling

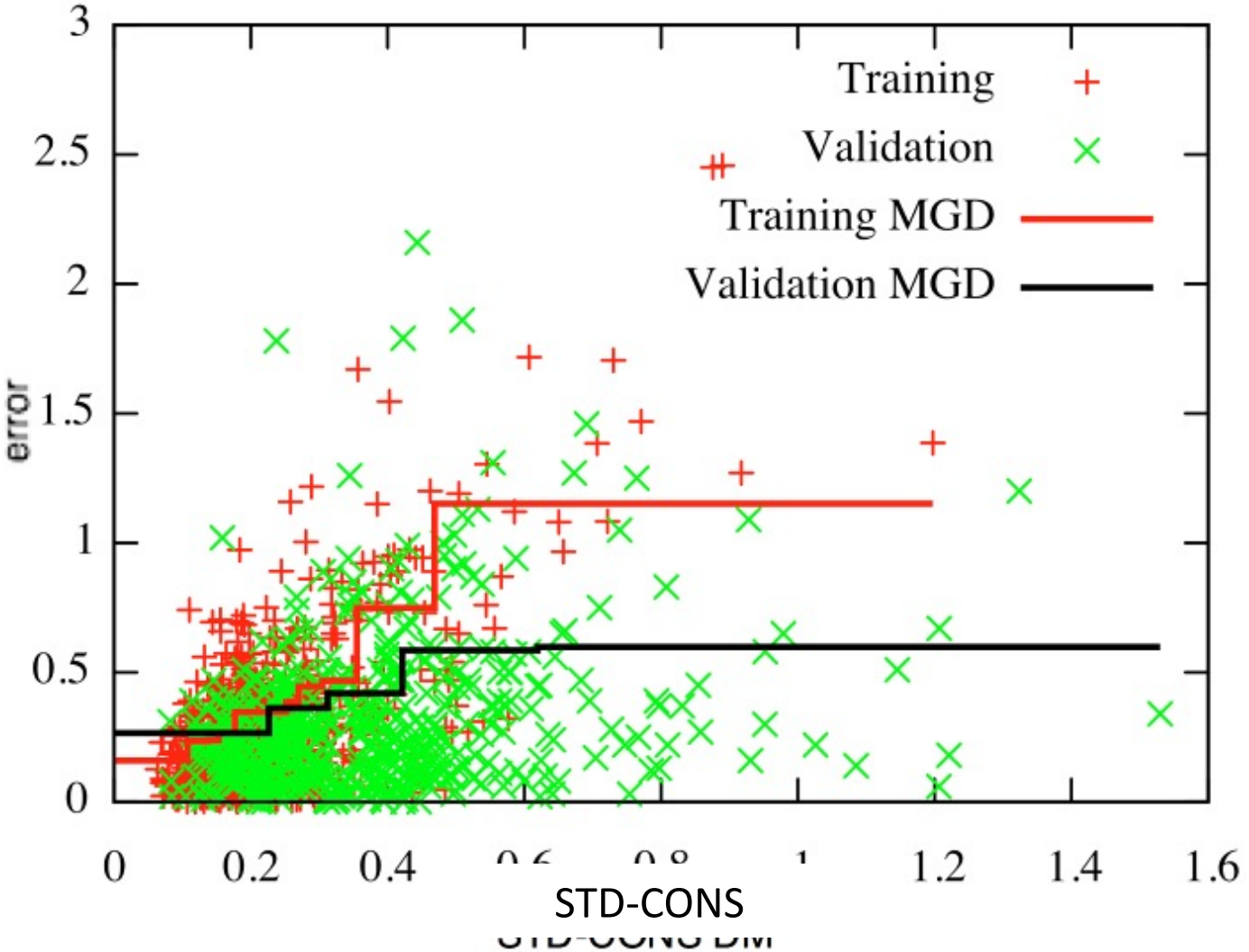
Best method(s) are defined

Average prediction of models is used

The consensus prediction is more accurate and stable

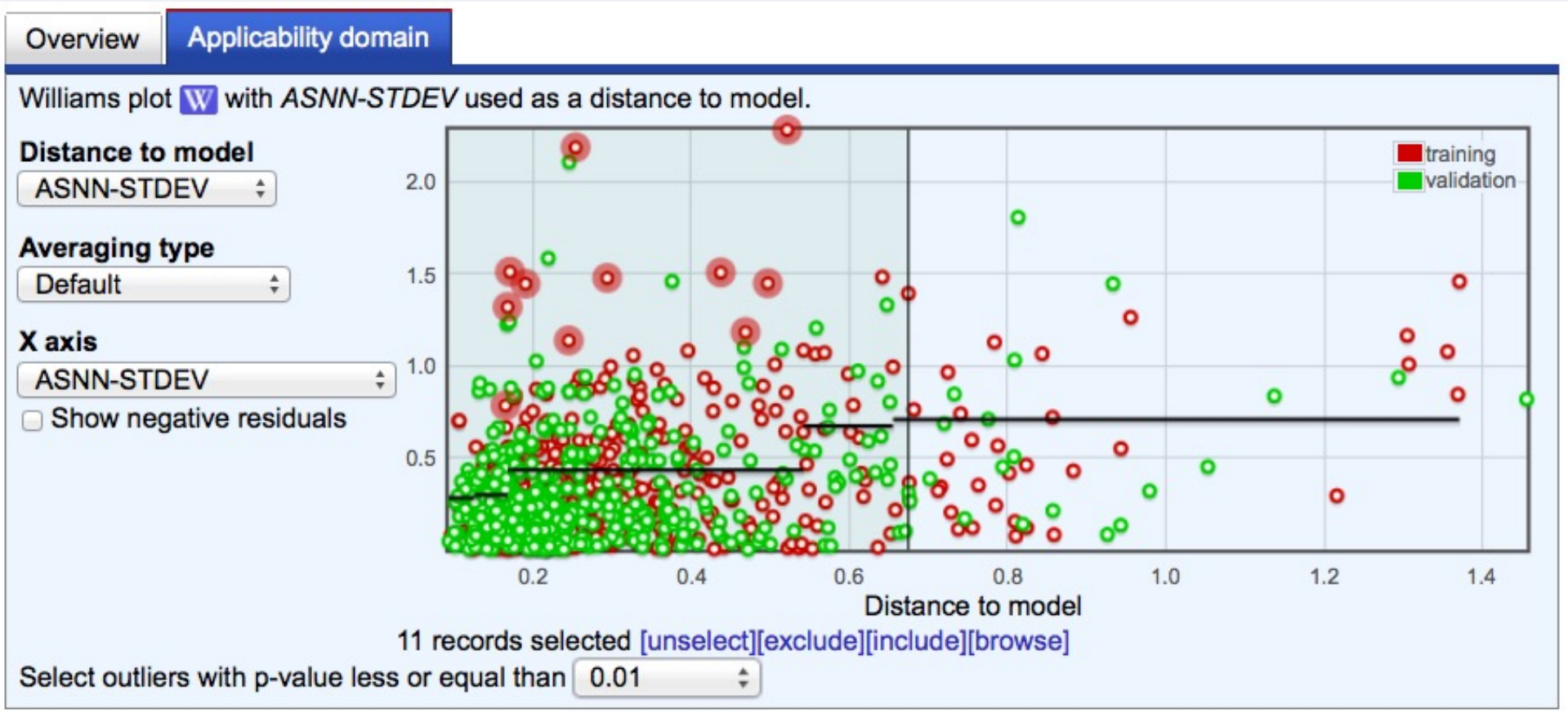


Property-based, ASNN model: DM does work!



Tetko et al, *J. Chem. Inf. Model*, **2008,** *48*, 1733-46**.**

Applicability domain assessment (regression)



- Several applicability domain measures (bagging-based for all methods; standard deviation, correlation in the property space, leverage, etc.)
- Automatic exclusion of outliers based on *p-value*

Accuracy of predictions for classification model

Overview | Applicability domain

Model name: Ames levenberg , published in Applicability domains for classification problems: Benchmarking of distance to models for Ames mutagenicity set. public identifier is 1
Predicted property: AMES
Training method: ANN

[OEstate]
Correl. limit: 0.95 Variance threshold: 0.0, Maximum value: 999999,
Levenberg, 1000 iterations, 3 neurons
ensemble=100 additional param PARALLEL=10
5-fold cross-validation
-

Calculated in 2402 seconds
Size: 450 Kb

Data Set	#	Accuracy	Balanced accuracy
Training set: Ames challenge training	4357 records (4359 selected)	78.1 ± 1.2	77.9 ± 1.3
Test set: Ames challenge test [x]	2181 records	79.9 ± 1.7	79.8 ± 1.7

Real↓/Predicted→	inactive	active
inactive	1521	495
active	460	1883
Training (Original)		

Real↓/Predicted→	inactive	active
inactive	802	207
active	232	940
Test (Original)		

Overview | Applicability domain

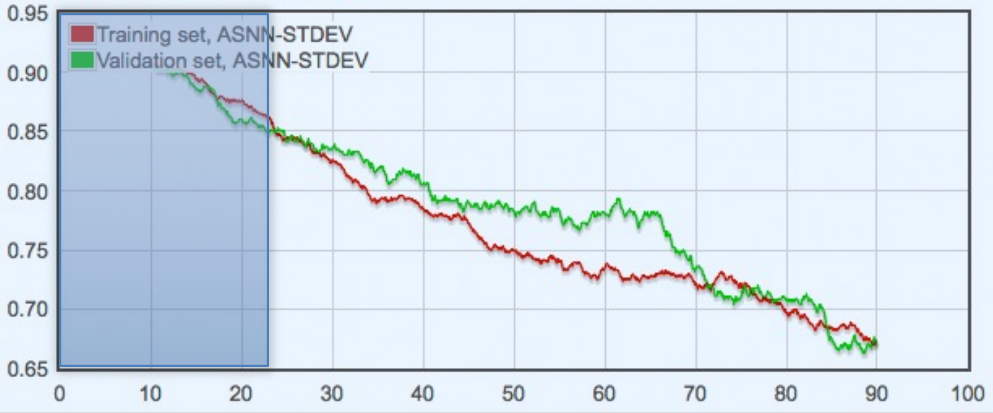
Wilkinson plot with *ASNN-STDEV* used as a distance to model.

Distance to model
ASNN-STDEV

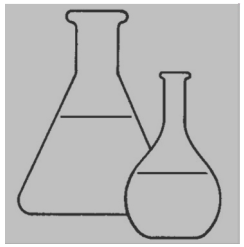
Averaging type
Default

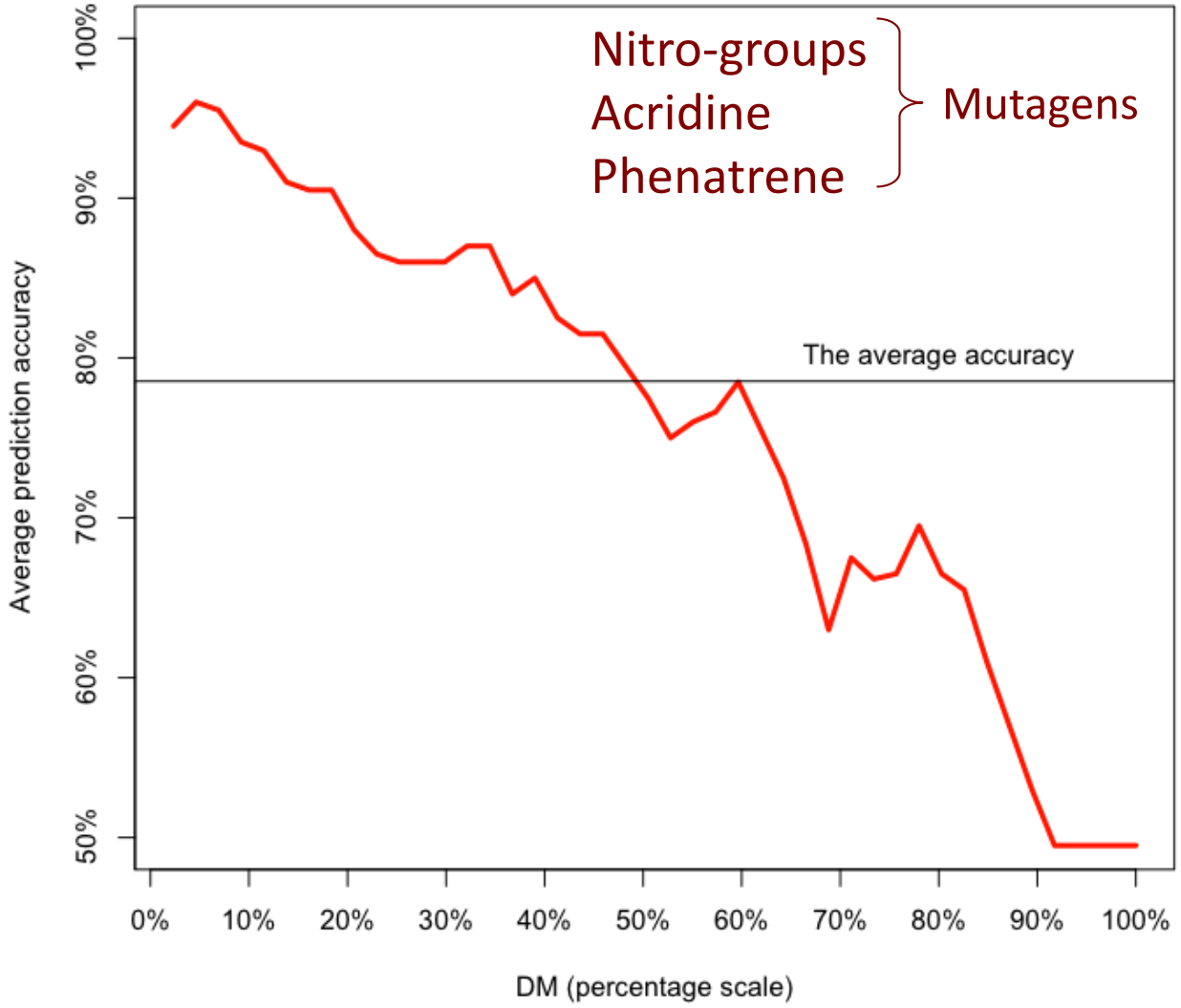
Window size:
20% of the set

X axis
Percentage of compounds

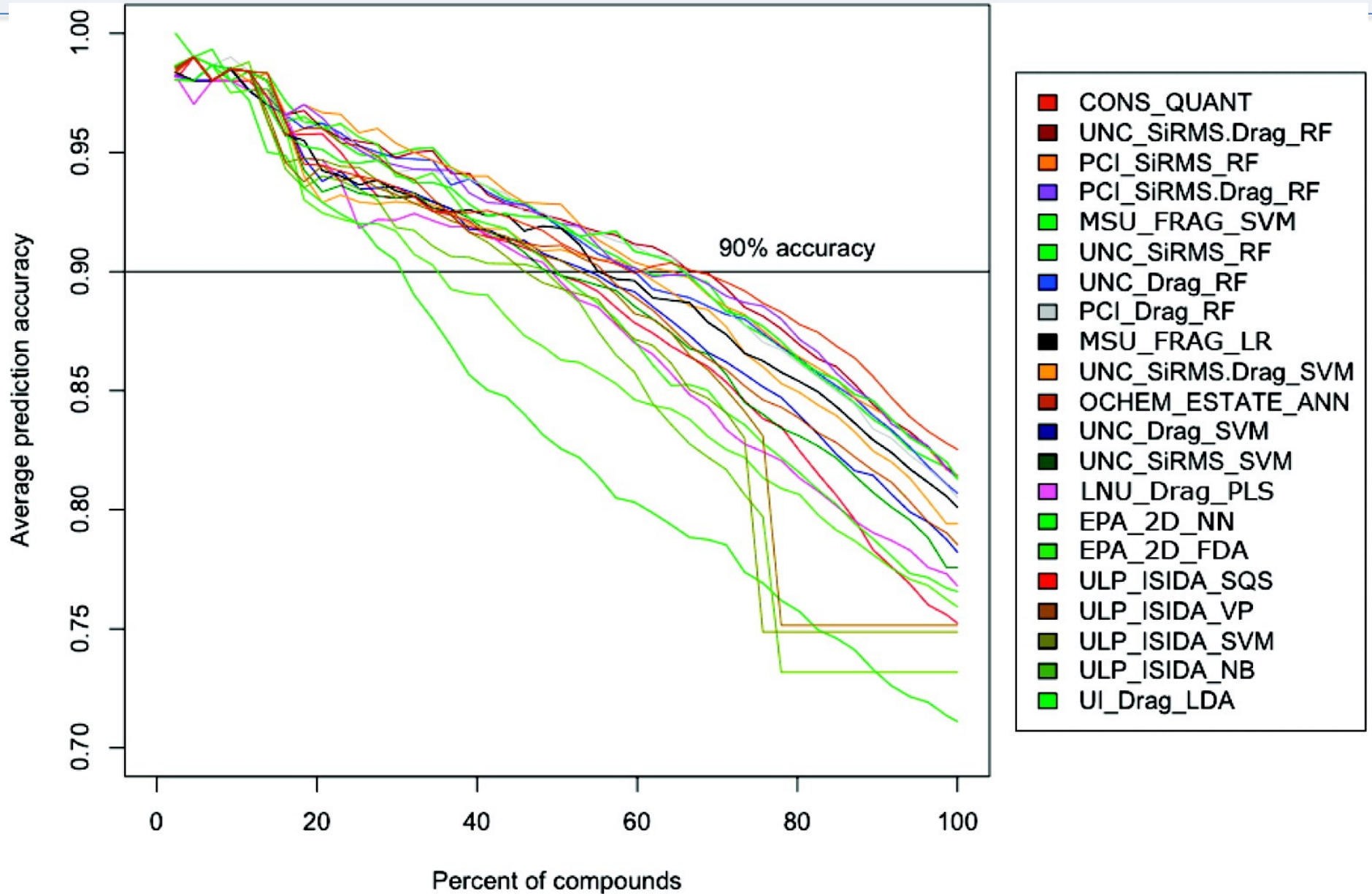


Discriminative power of model for AMES test





Accuracy of all models for AMES test set



Sushko et al, JCIM, 2010, 50, 2094 - 2111.

Model explanations

In November 2004, the 37th OECD's Joint Committee and the Working Party on Chemicals, Pesticides and Meeting of the Chemicals Biotechnology (Joint Meeting) agreed on the OECD Principles for the Validation, for Regulatory Purposes, of (Q)SAR Models.

The OECD Principles of (Q)SAR Validation

To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:

1. a defined endpoint;
2. an unambiguous algorithm;
3. a defined domain of applicability;
4. appropriate measures of goodness-of-fit, robustness and predictivity;
5. **a mechanistic interpretation, if possible.**

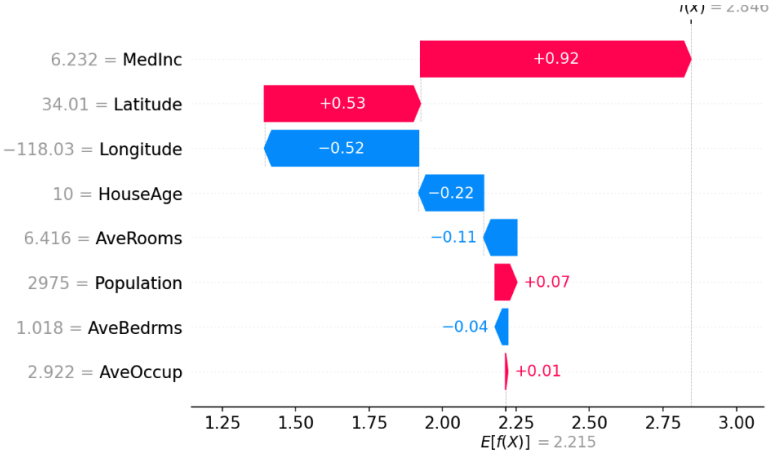
Model agnostic methods: SHAP values, LIME

EPFL

SHAP values

Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017).

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \right]$$

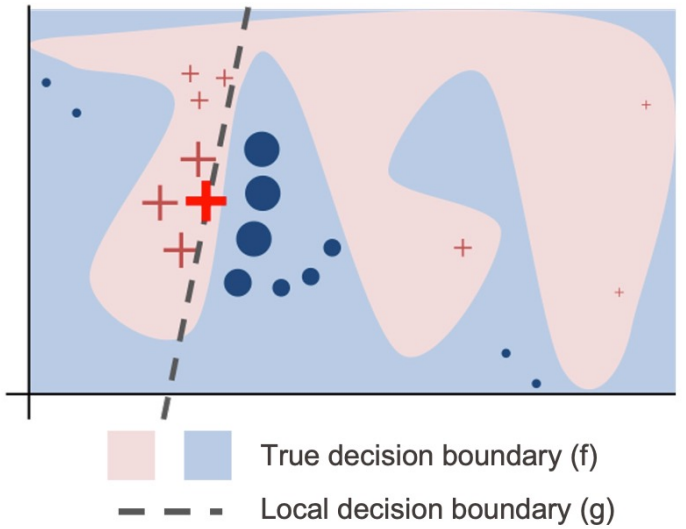


https://shap.readthedocs.io/

LIME

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "" Why should i trust you?" Explaining the predictions of any classifier." *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016.

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \ \mathcal{L}(f, g, \pi_x) + \Omega(g)$$



See full lecture of Dr. Wellawatte at https://ai-dd.eu/lectures

Linear classification model for AMES test

Mopac2016 descriptors

Y = 0.5378 – 0.3411*MullikenElectronegativity – 0.3277*LumoEnergy + 0.2389*IonisationPotential + 0.1178*FinalHeat + 0.05051*DipolPointCharge

GSFRAG

Y = 0.5372 + 0.1612*c10 + 0.1309*p1-1N – 0.1134*p2B + 0.05943*c3 + 0.05349*c9

E-state descriptors

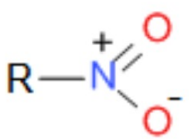
Y = 0.5375 + 0.09956*PSA + 0.08731*aCNOS – 0.08703*DONORS – 0.06814*SsCH3 – 0.04474*SssO

Dragon descriptors

Y = 0.5375 – 0.1173*GATS1m + 0.0954*MATS1e – 0.06558*SpMax_AEA(dm) + 0.05933*J_D/Dt + 0.05496*nR03

Structural Alerts

Y = 0.4733 + 0.191*Alert146 – 0.004113*Alert238– 0.1024*Alert213 + 0.1912*Alert214+ 0.01988*Alert196



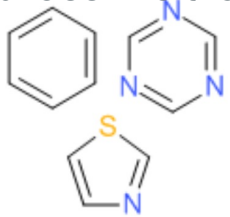
Aryl halides

R = aryl

SMARTS: a[F,Cl,Br,I]

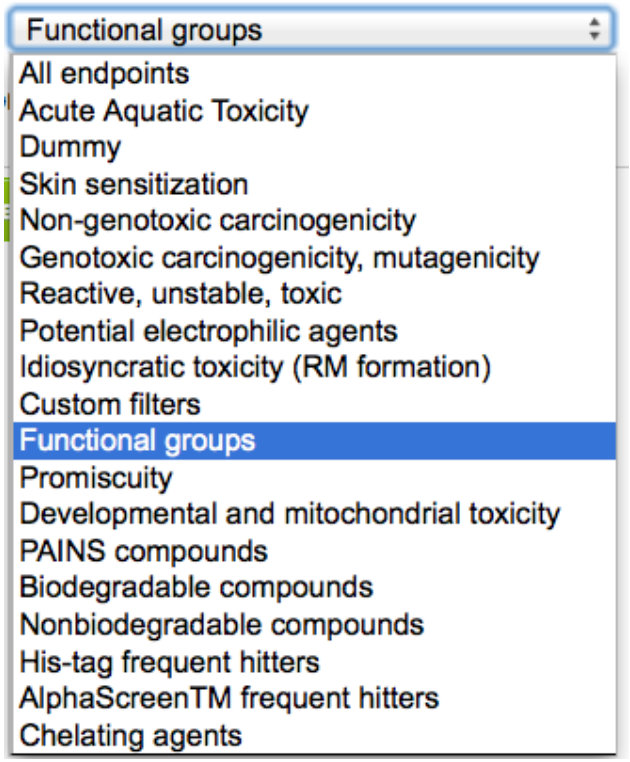
Alkyl halides

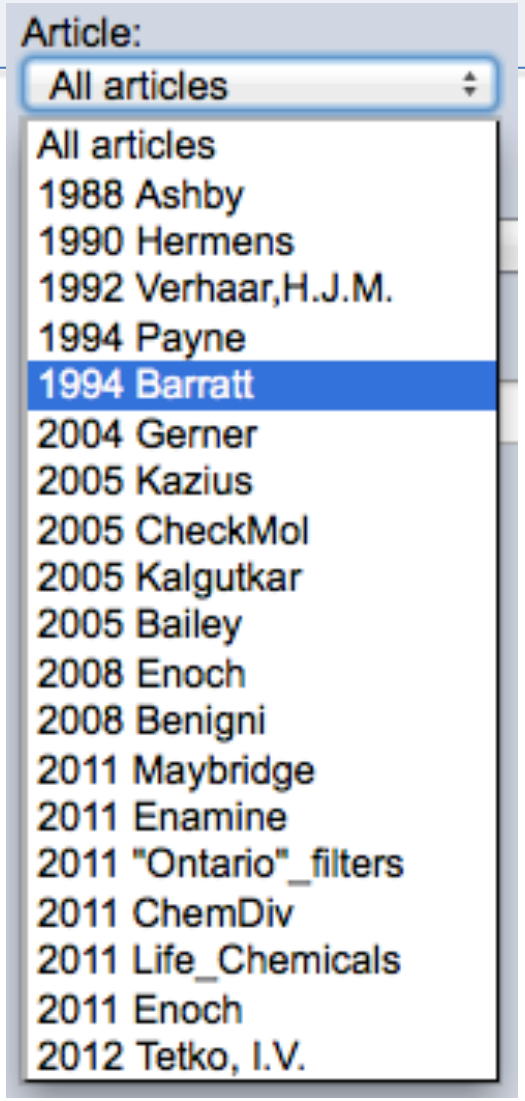
R = alkyl



ToxAlerts

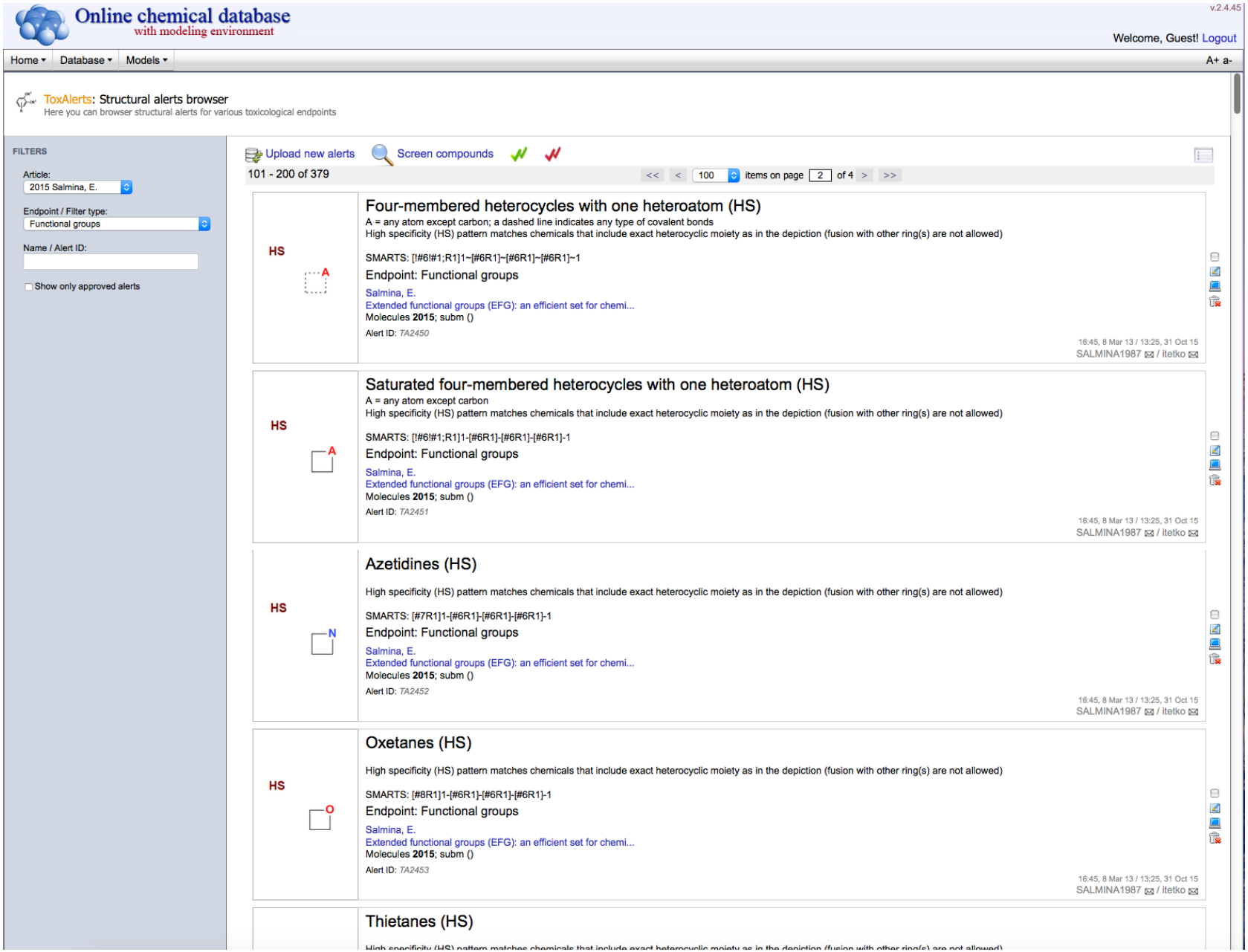
- Screening of compounds against published toxicity alerts, groups, frequent hitters
- Filter alerts by endpoints or publications
- Create or upload custom SMARTS rules





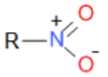
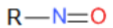
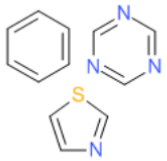
Sushko et al, JCIM, 2012, 52(8):2310-6.

Functional groups

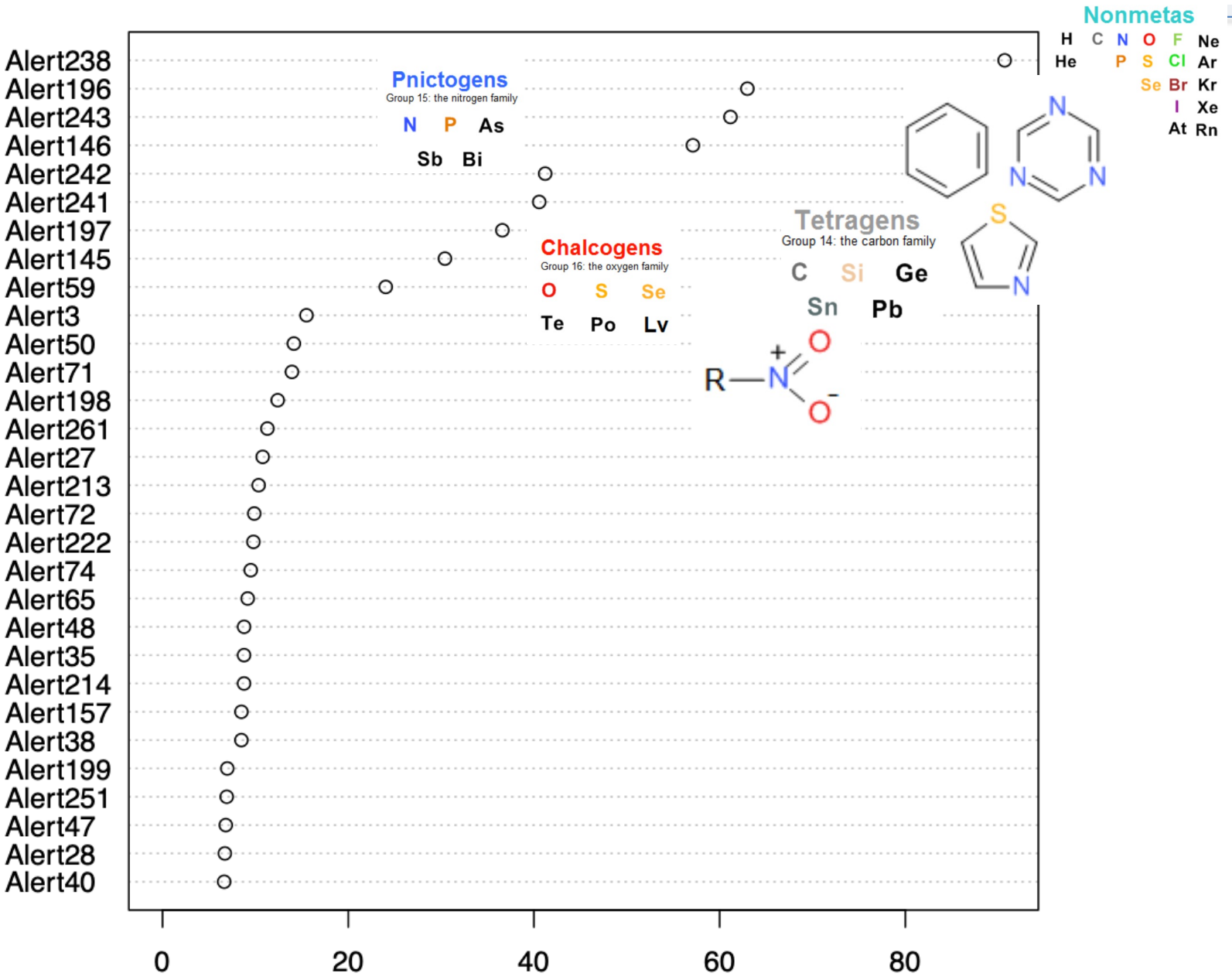


Overrepresented functional groups (AMES)

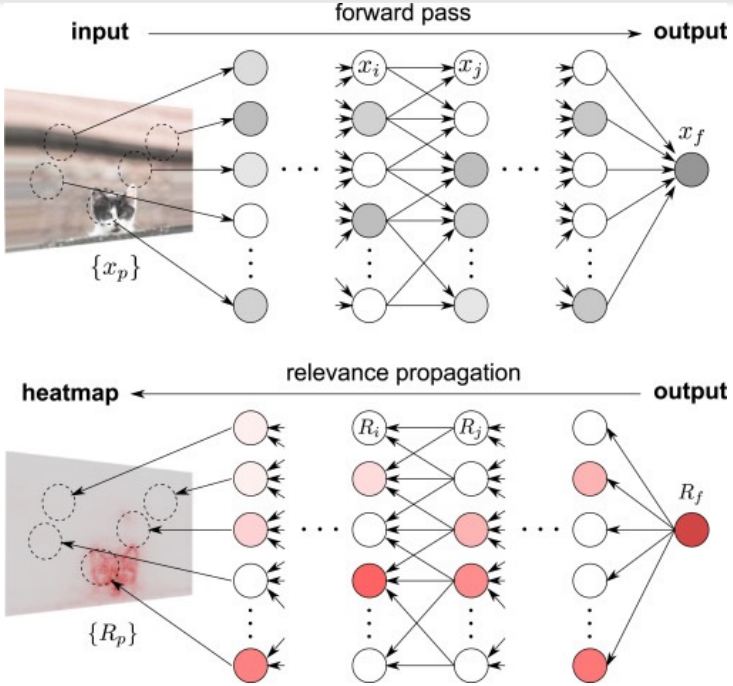
Active Inactive

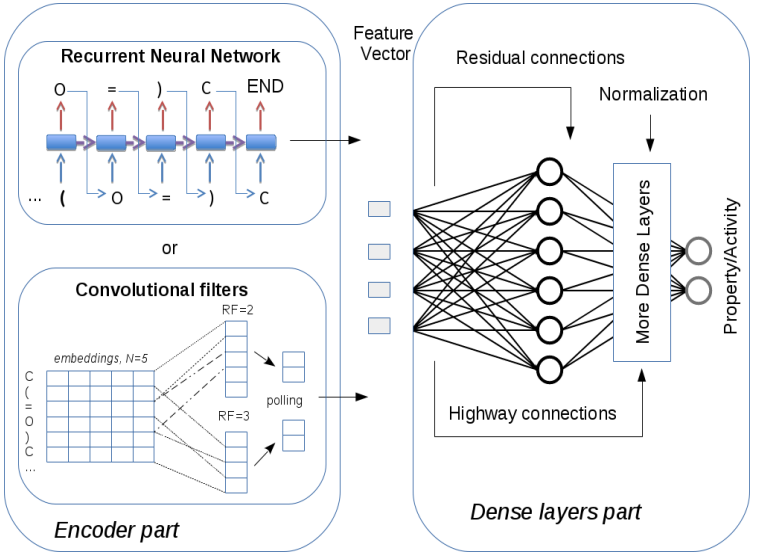
Descriptor	In set 1 (3514 unique molecules)	In set 2 (3023 unique molecules)	Enrichment factor	p-Value
	808 (23.0%)	167 (5.5%)	4.2	1.545E-94
Pnictogens Group 15: the nitrogen family **N P As Sb Bi**	2498 (71.1%)	1714 (56.7%)	1.3	5.687E-34
	219 (6.2%)	28 (0.9%)	6.7	2.796E-33
	2546 (72.5%)	1782 (58.9%)	1.2	7.944E-31
	2770 (78.8%)	2004 (66.3%)	1.2	3.461E-30

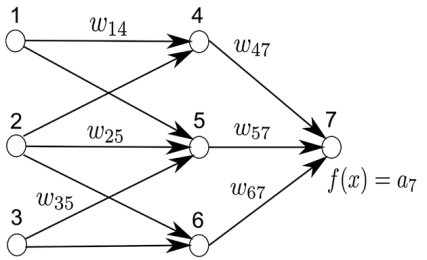
Importance of ToxAlerts in Random Forest model

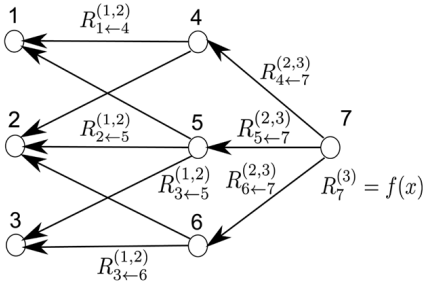


Layerwise Relevance Propagation (LRP)







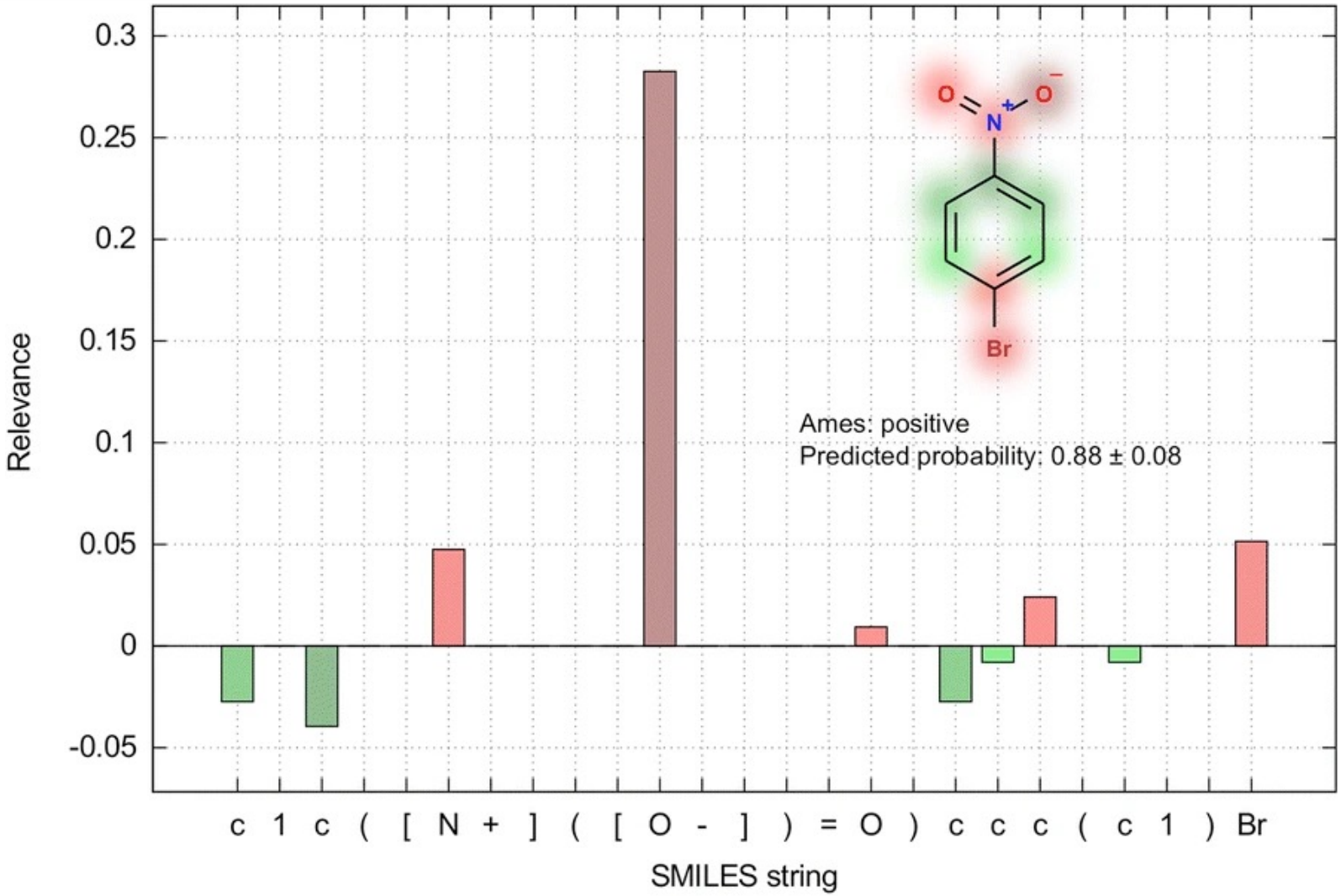


$$f(x) \approx f(x_0) + Df(x_0)[x - x_0]$$

$$= f(x_0) + \sum_{d=1}^{V} \frac{\partial f}{\partial x_{(d)}}(x_0)(x_{(d)} - x_{0(d)})$$

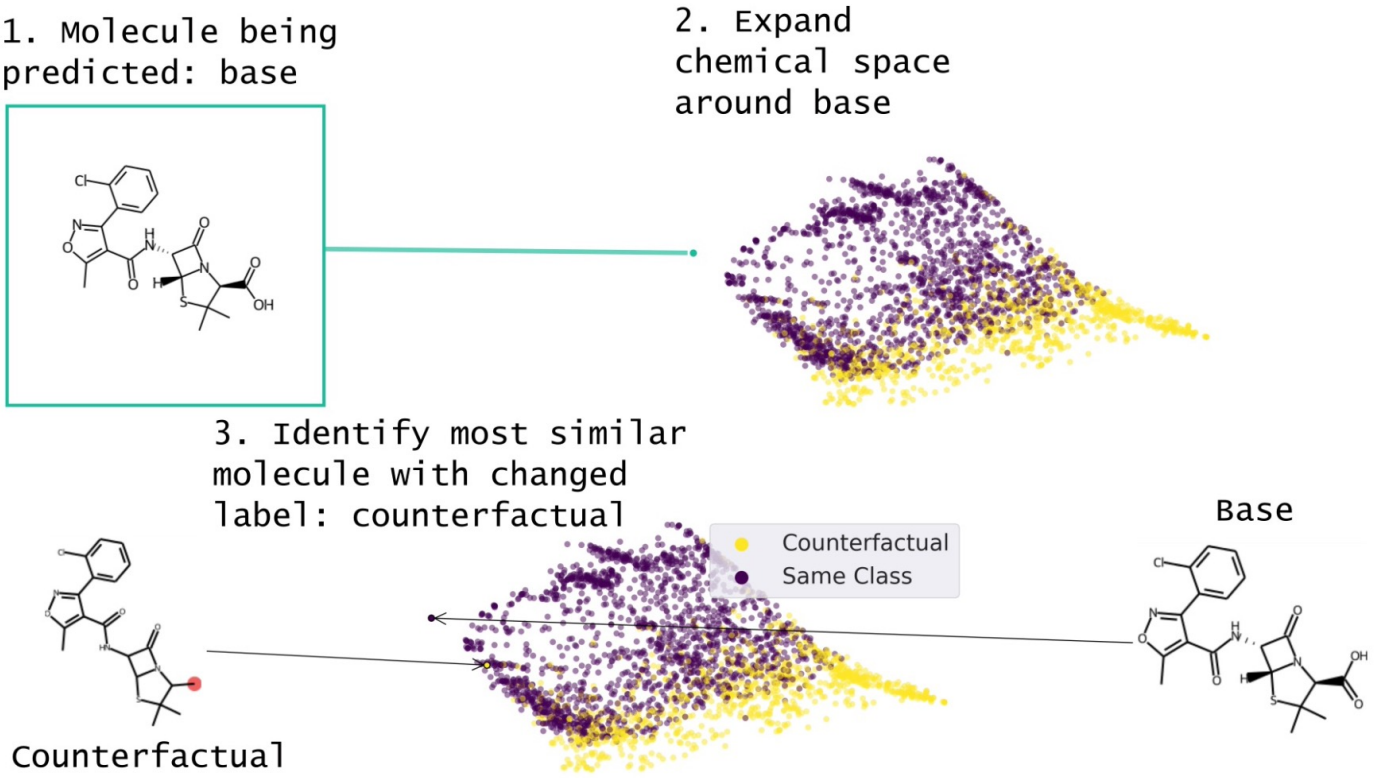
Bach, S. et al. *PloS One* **2015**, *10*, e0130140.

Interpretation of models



P. Karpov, G. Godin, I. V. Tetko, *J. Cheminform.* **2020**, *12*, 17.
https://github.com/bigchem/transformer-cnn

Contrafactual examples – Molecular Model Agnostic Counterfactual Explanations

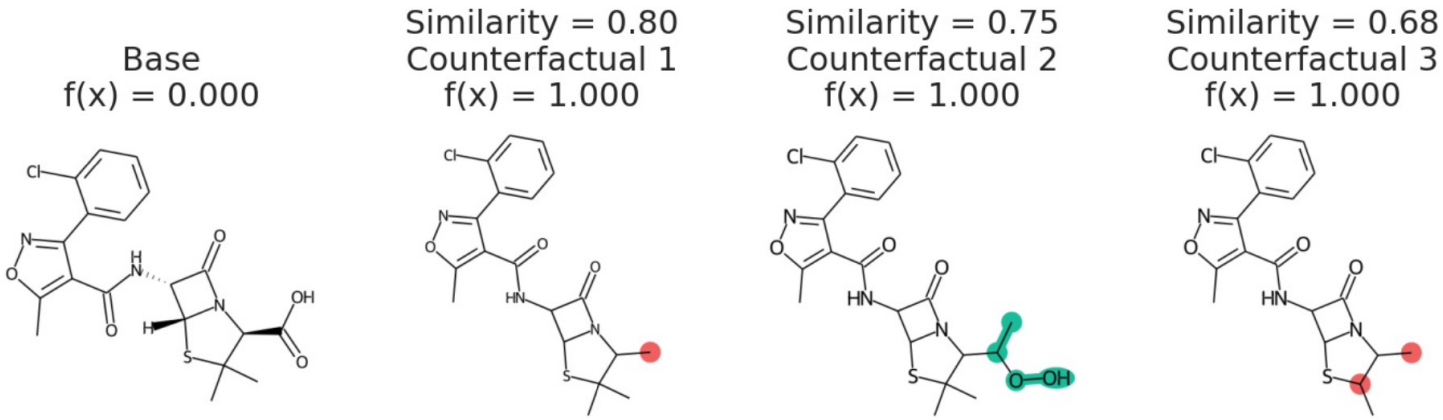


Wellawatte, G. P., Seshadri, A., & White, A. D. (2022). *Chemical science*, *13*(13), 3697-3705.

https://github.com/ur-whitelab/exmol

Example of interpretation of RF model for BBP

EPFL



Explanation: The negative example can be made to cross the blood brain barrier if the carboxylic group is altered.

Experimental observations: hydrophobic interactions and surface area govern BBB permeation (Boobier S, *et al., Nat Commun*. 2020)

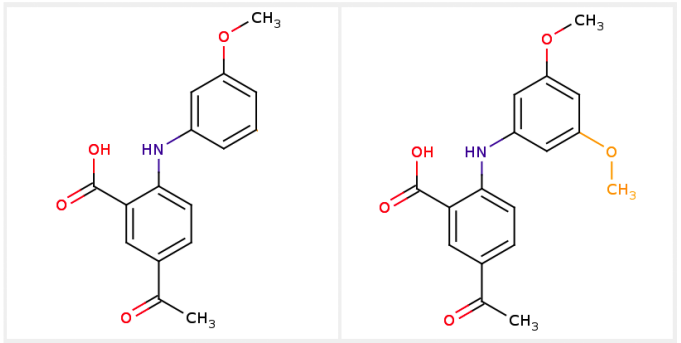
Wellawatte, G. P., Seshadri, A., & White, A. D. (2022). *Chemical science*, *13*(13), 3697-3705.

See full lecture of Dr. Wellawatte at https://ai-dd.eu/lectures

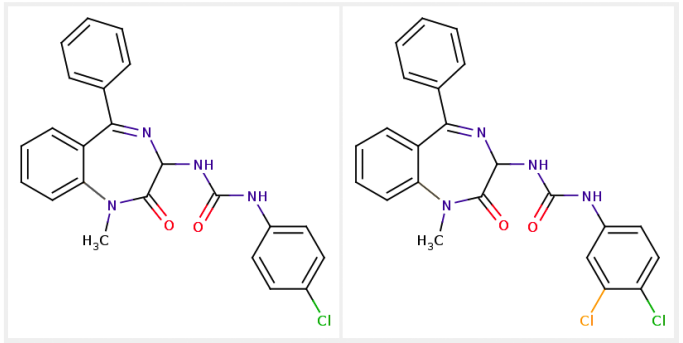
MMP definition

A molecular matched pair (MMP) is a pair of molecules that have only a (minor) single-point difference.
The typical way is to define a minor difference as a changed molecular fragment with less than 10 atoms.

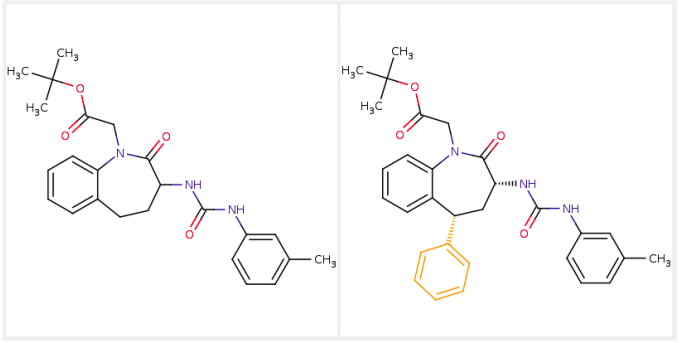
Pair 1



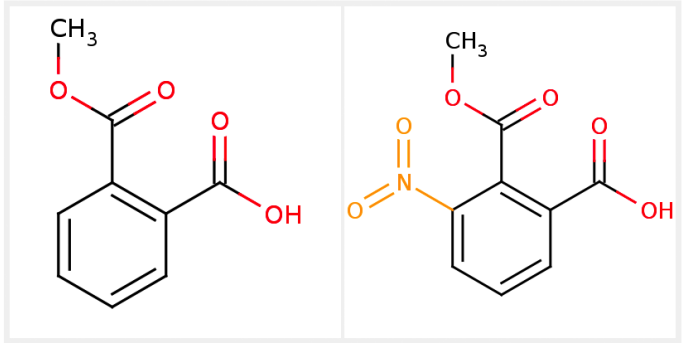
Pair 2



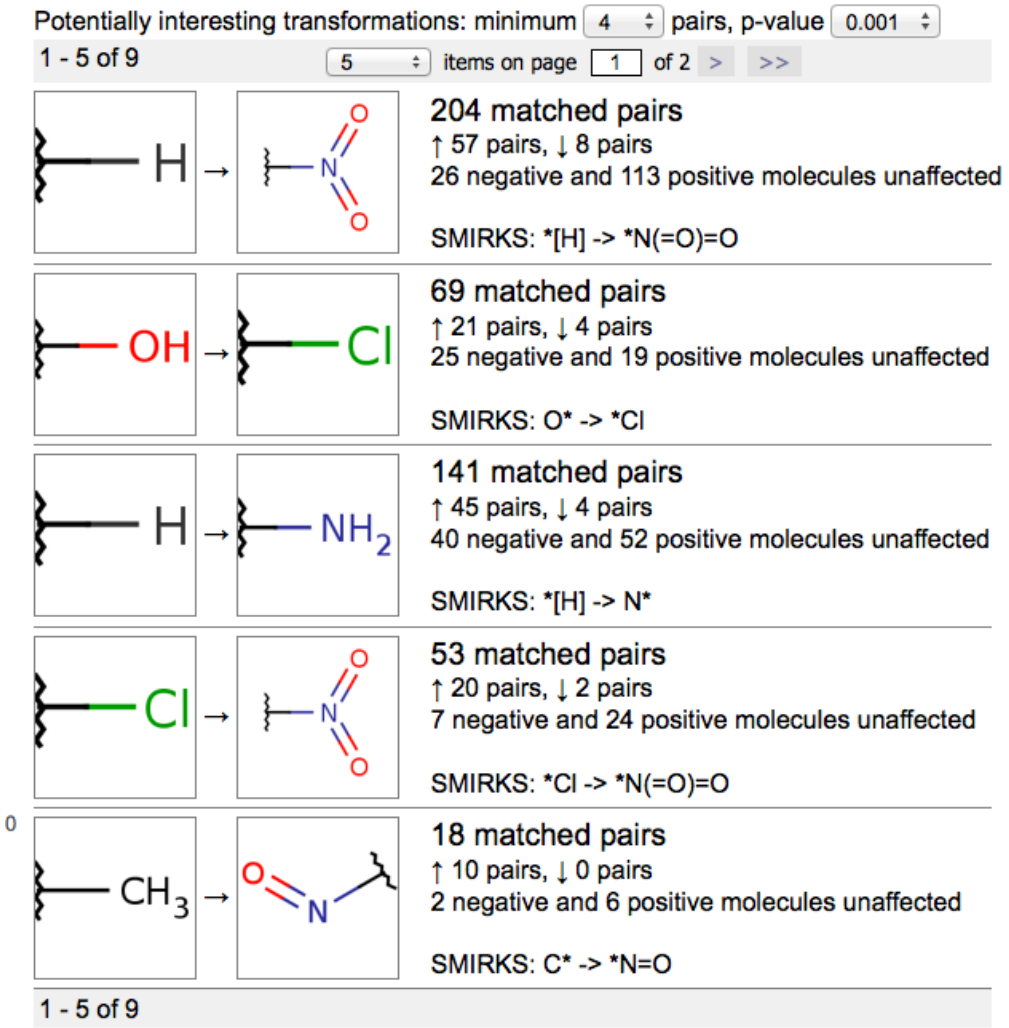
Pair 3



Pair 4

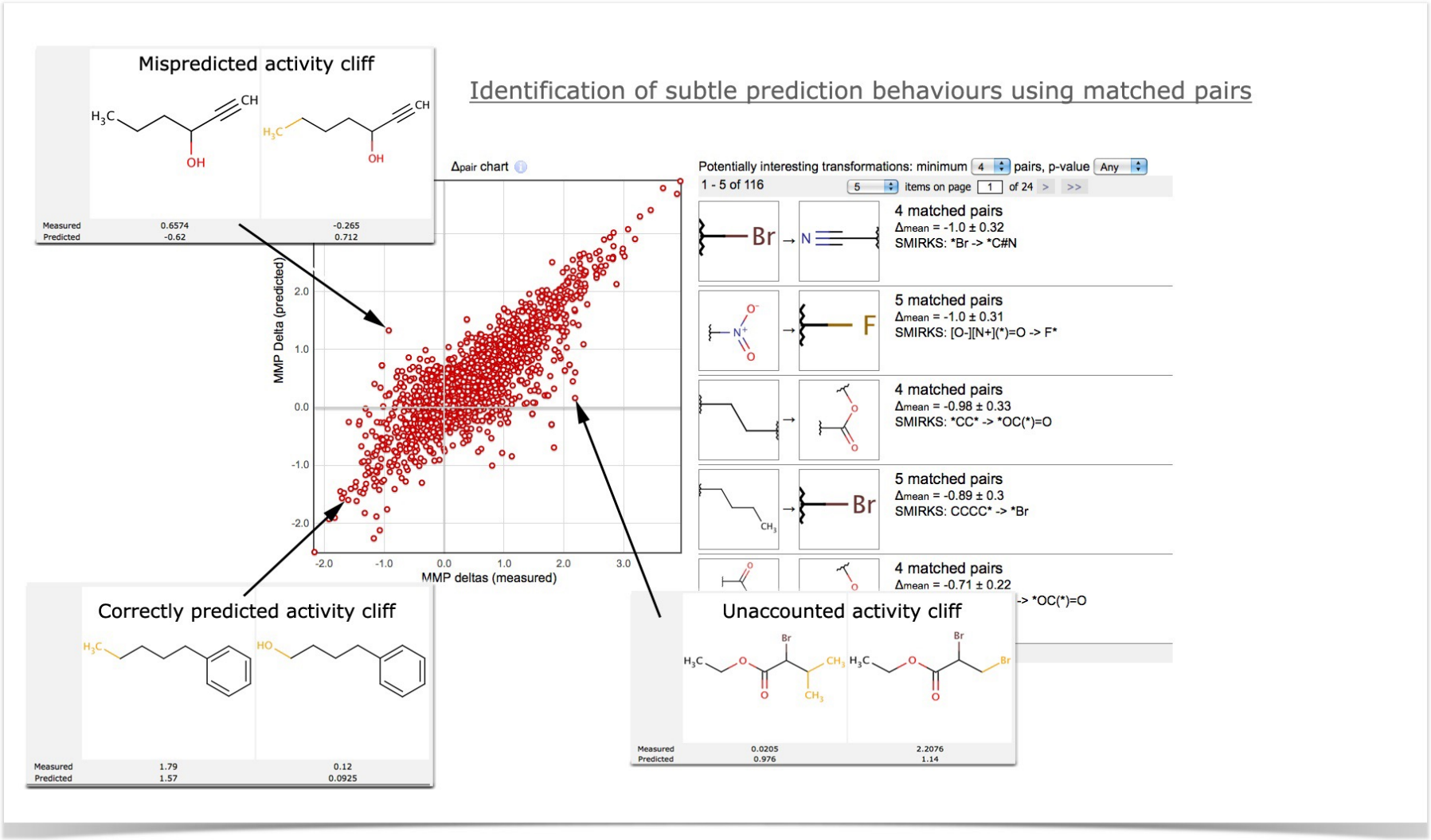


Data analysis using Matched Molecular Pairs



Identify molecular transformations that lead to significant change of activity (AMES test data are shown)

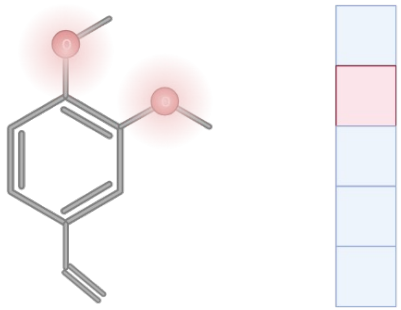
Identification of predicted activity cliffs



https://jcheminf.biomedcentral.com/articles/10.1186/s13321-014-0048-0

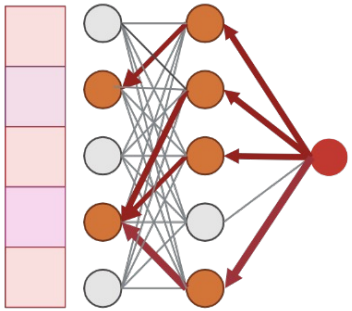
XAI methods

Pertubation-based XAI

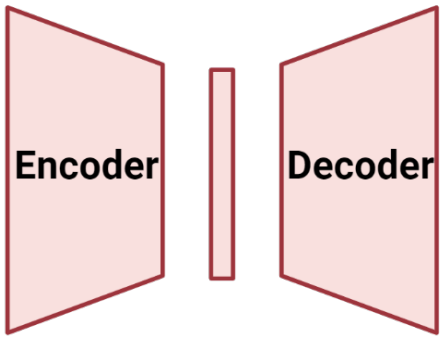


SHAP

Gradient-based XAI



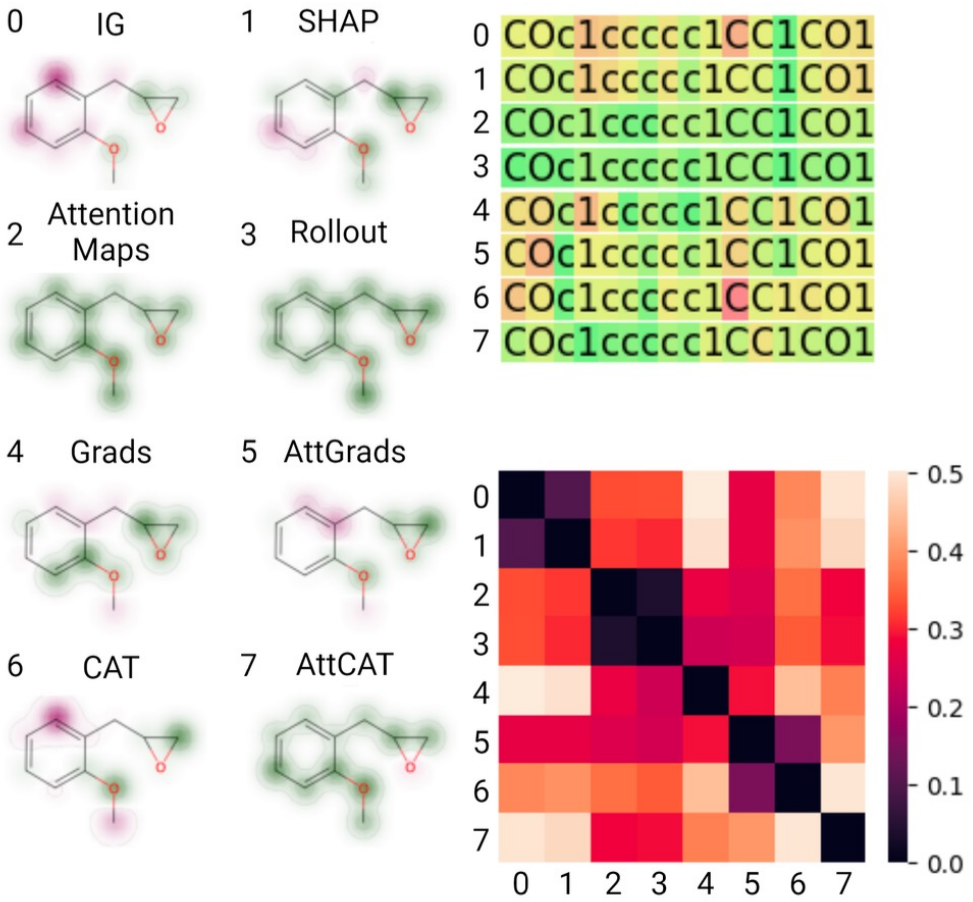
Integrated Gradients (IG)



Attention Maps, Rollout, Grads, AttGrads, CAT and AttCAT

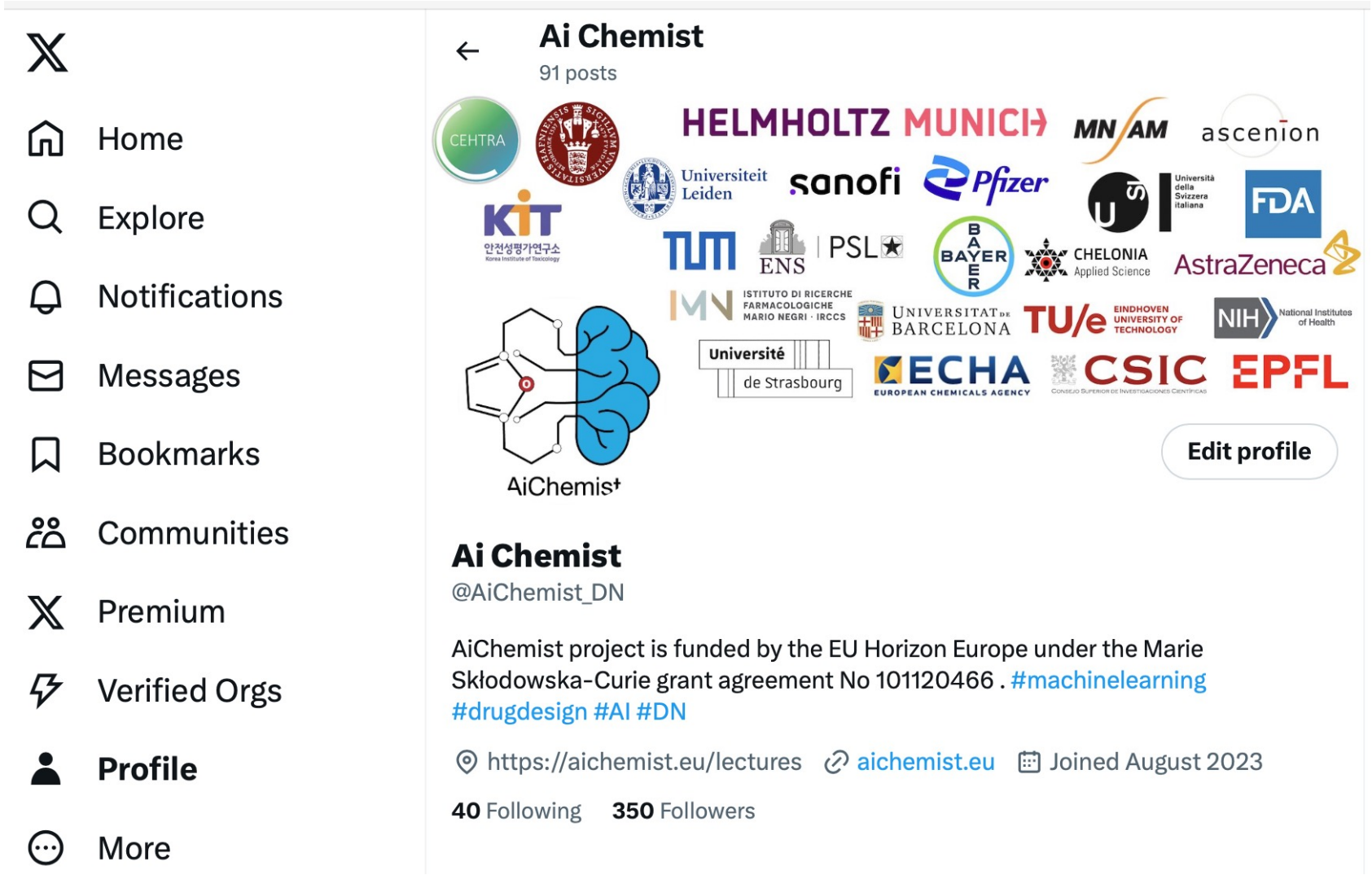
Hartog P et al, Using test-time augmentation to investigate explainable AI: inconsistencies between method, model and human intuition *J. Cheminformatics*, **2024**, 16 (1), 39.

Importance of features across XAI methods



Hartog P et al, Using test-time augmentation to investigate explainable AI: inconsistencies between method, model and human intuition *J. Cheminformatics*, **2024**, 16 (1), 39.

AiChemist – Explainable AI for molecules



Twitter: aichemist_dn or Bluesky: aichemist

Take home message

New methods based on representation learning successfully compete with traditional ones

Consensus modelling is a best approach to develop models with the highest prediction accuracy

Applicability domain of models and accuracy of predictions are crucial for their use and interpretation predictions

Model interpretation is essential for their acceptance by the end users (sometimes legally required)

Different XAI explanations do not always overlap; statistical evaluation is strongly required

Acknowledgements

Katya Ahmad
Thalita Chirino
Mark Embrechts

Andi Kopp
Peter Hartog
Fabian Krüger
Paula Torren-Peraire
Varvara Voinarovska
Nesma Mousa

Guillaume Godin (ex Firmenich)
Ruud van Deursen (DSM-Firmenich)

Michael Sattler (HMGU)

Alexander von Humboldt Stiftung/Foundation

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

AiChemist

eco
environmental chemoinformatics
marie curie initial training network

BigChem