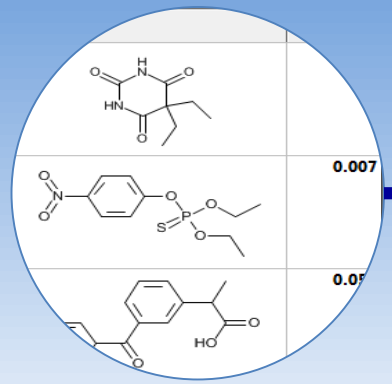
Predicting Five Rat Acute Toxicity Endpoints with ANNE Models

Michael Lawless, Pankaj R. Daga, Marvin Waldman, Robert Fraczkiewicz, Robert D. Clark, John DiBella, and Michael B. Bolger

Rat Acute Toxicity ANNE Models

- ANNE methodology
  - Regression, binary classification, and multiclass models
- Data set curation
- Results
- Summary

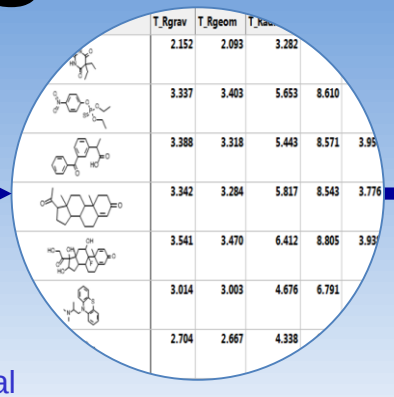
Building ANNE Models



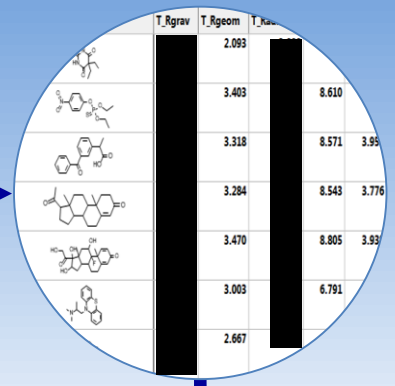
Structures and experimental data

Compute >300 Descriptors

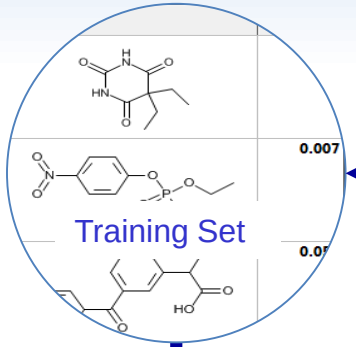
Constitutional
Topological
Hbonding
Ionization
Electrotopological
Charge/reactivity



Remove low variance and correlated descriptors



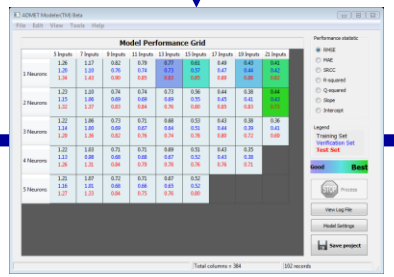
Test Set Selection
1) Kohonen map
2) Stratified sampling
3) Random
4) K-means
5) Manual



Training Set

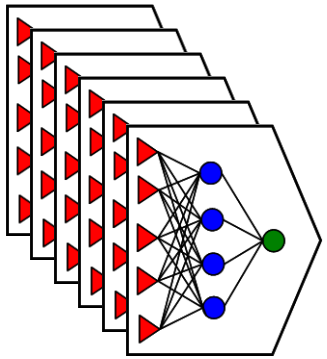
- No. of neurons and descriptors
  - Create models with different architectures
- Sensitivity analysis
  - Which descriptors create the best model?

ANNE Training



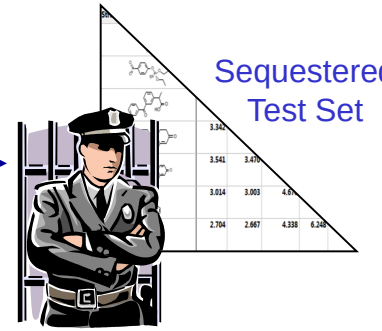
Grid of ANNE Models
(X descriptors by Y neurons)

Select best model



Ensemble of 33 ANNs

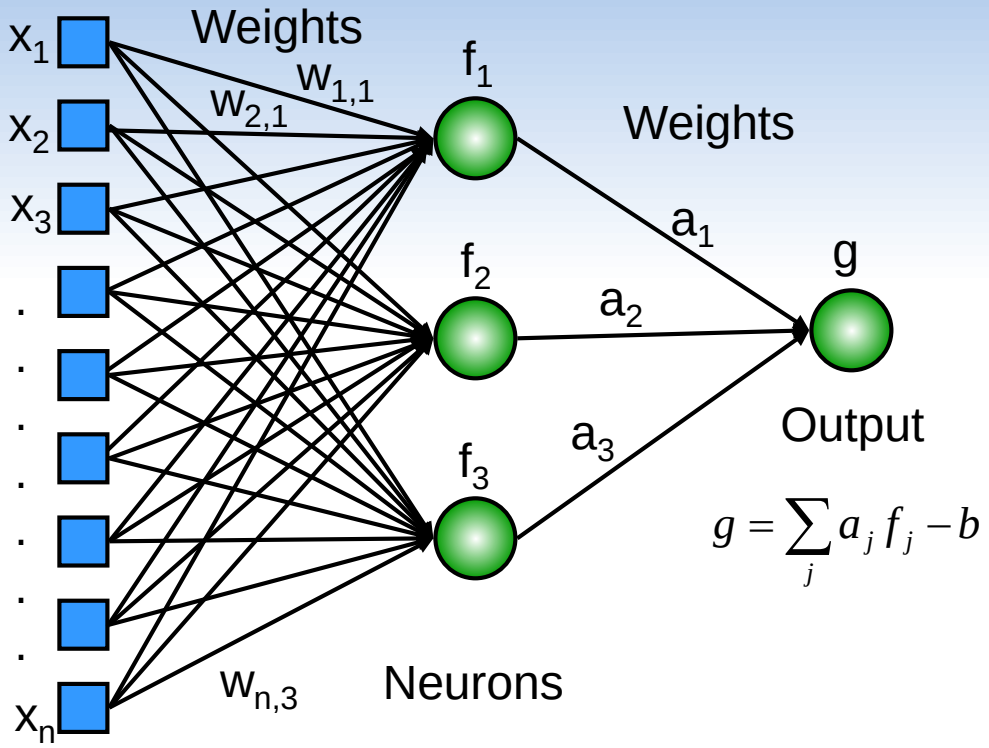
Apply model to test set



Sequestered Test Set

Regression Neural Network

Descriptors: x_i
Normalized to range 0.0-1.0



$$f_j = \tanh\left(\sum_i w_{ij} x_i - t_j\right)$$

$$g = \sum_j a_j f_j - b$$

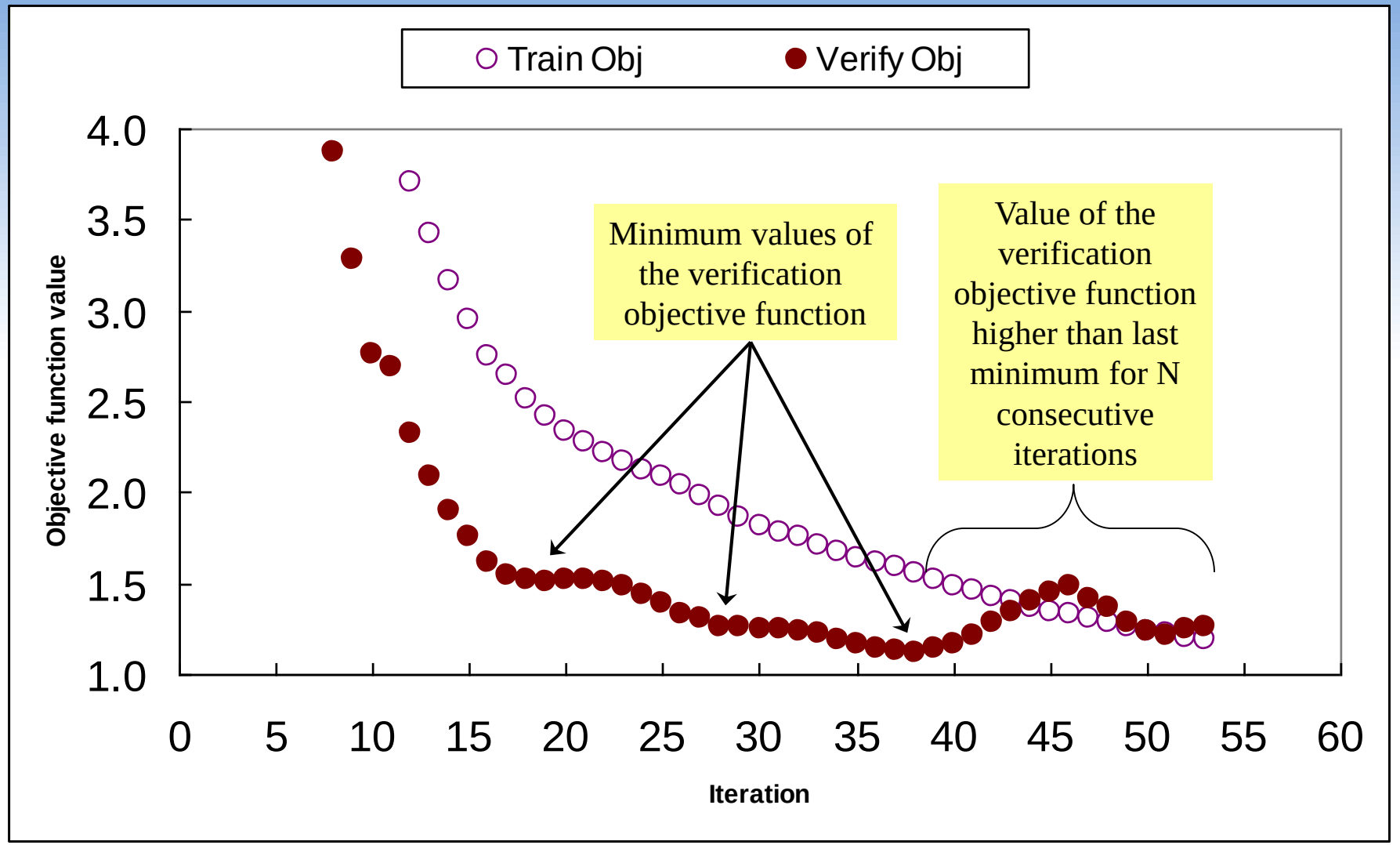
Weights are adjusted iteratively to optimize model performance on the training set

$$Obj = \sum_{k=1}^{n} (y_k - g(k))^2$$

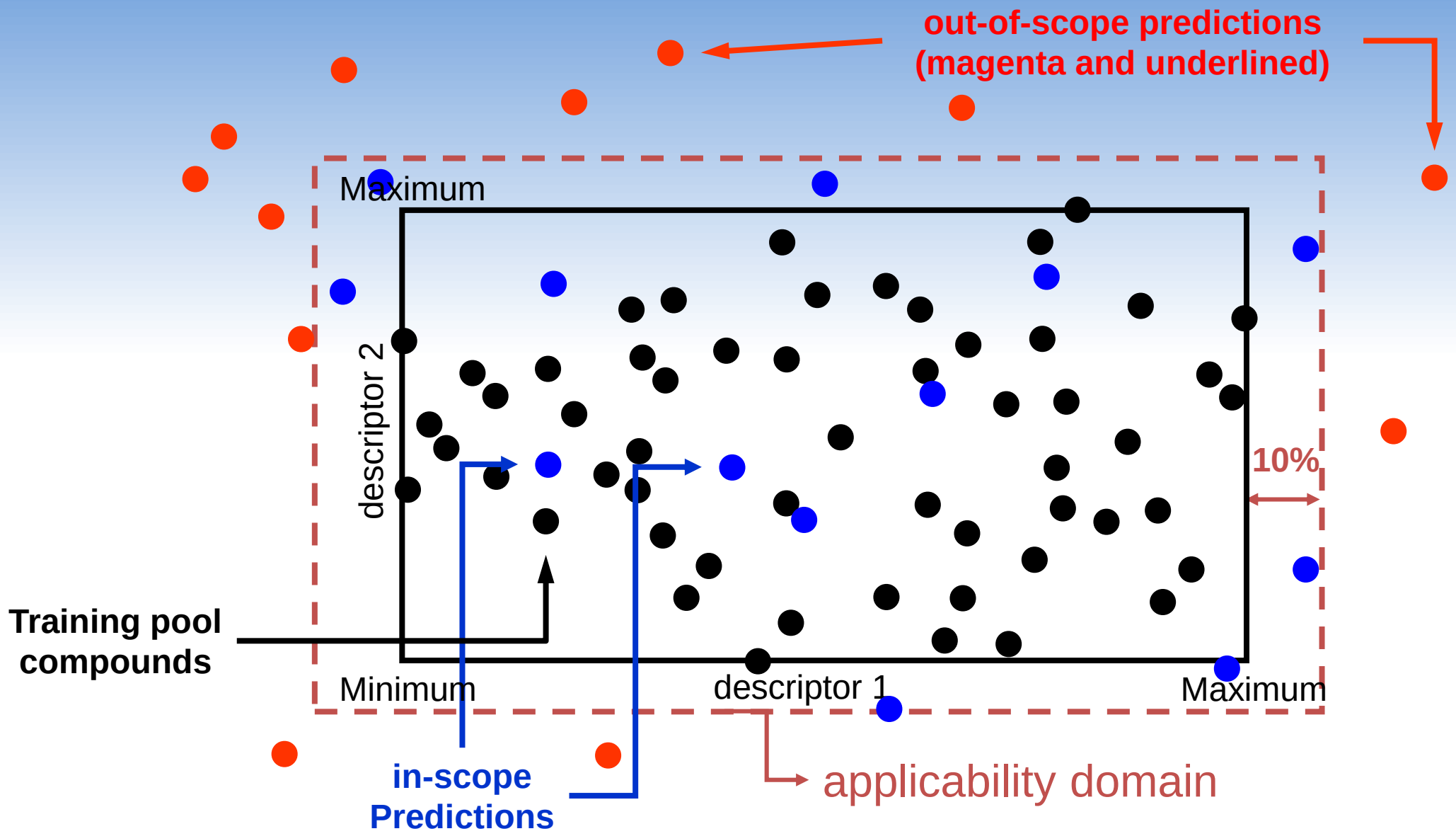
where y_k is the observed value for observation k

4

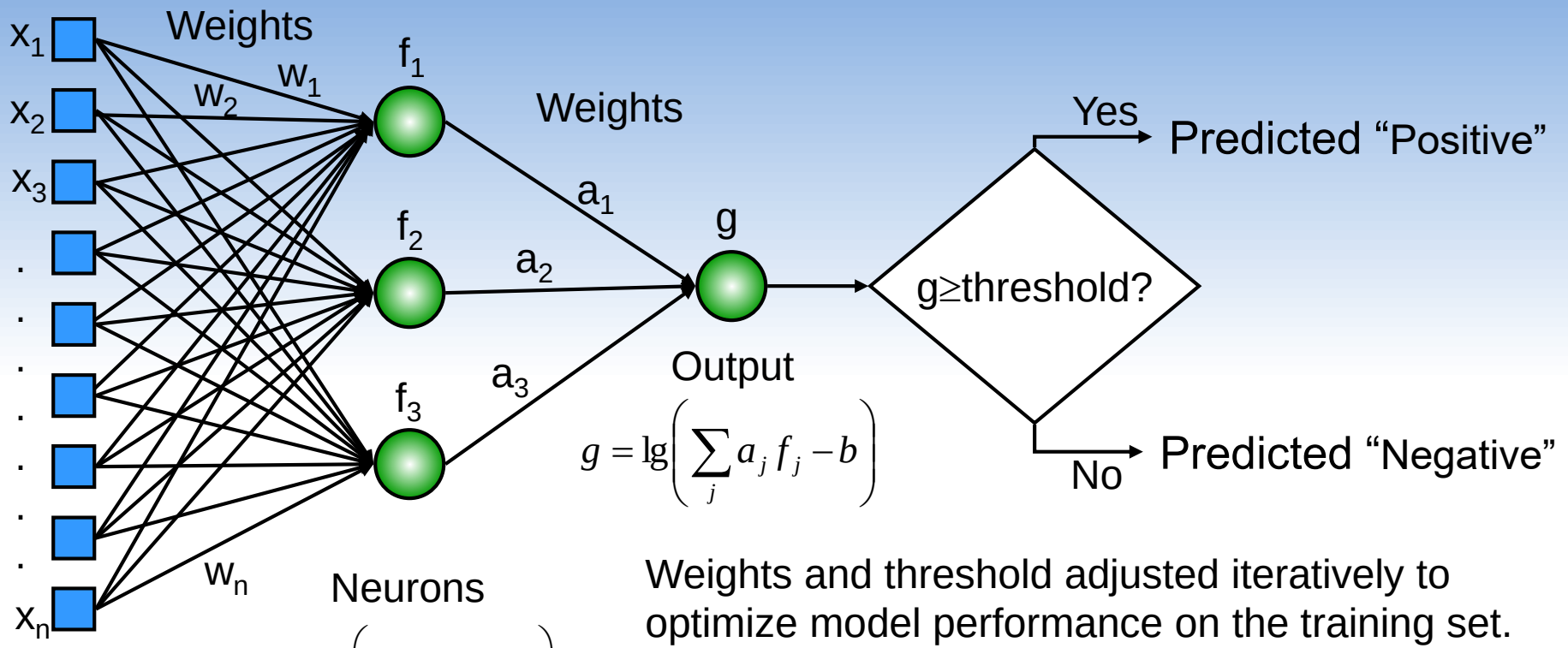
Early Stopping Prevents Overtraining



Applicability Domains



Binary Classification Neural Network



$$f_j = \tanh\left(\sum_i w_{ij} x_i - t_j\right)$$

$$g = \lg\left(\sum_j a_j f_j - b\right)$$

Descriptors: X_i
Normalized to range 0.0-1.0

Weights and threshold adjusted iteratively to optimize model performance on the training set.

$$Obj = \sum_{k=1}^{n} w_0 (1 - c(k))(g(k))^2 + w_1 c(k)(1 - g(k))^2$$

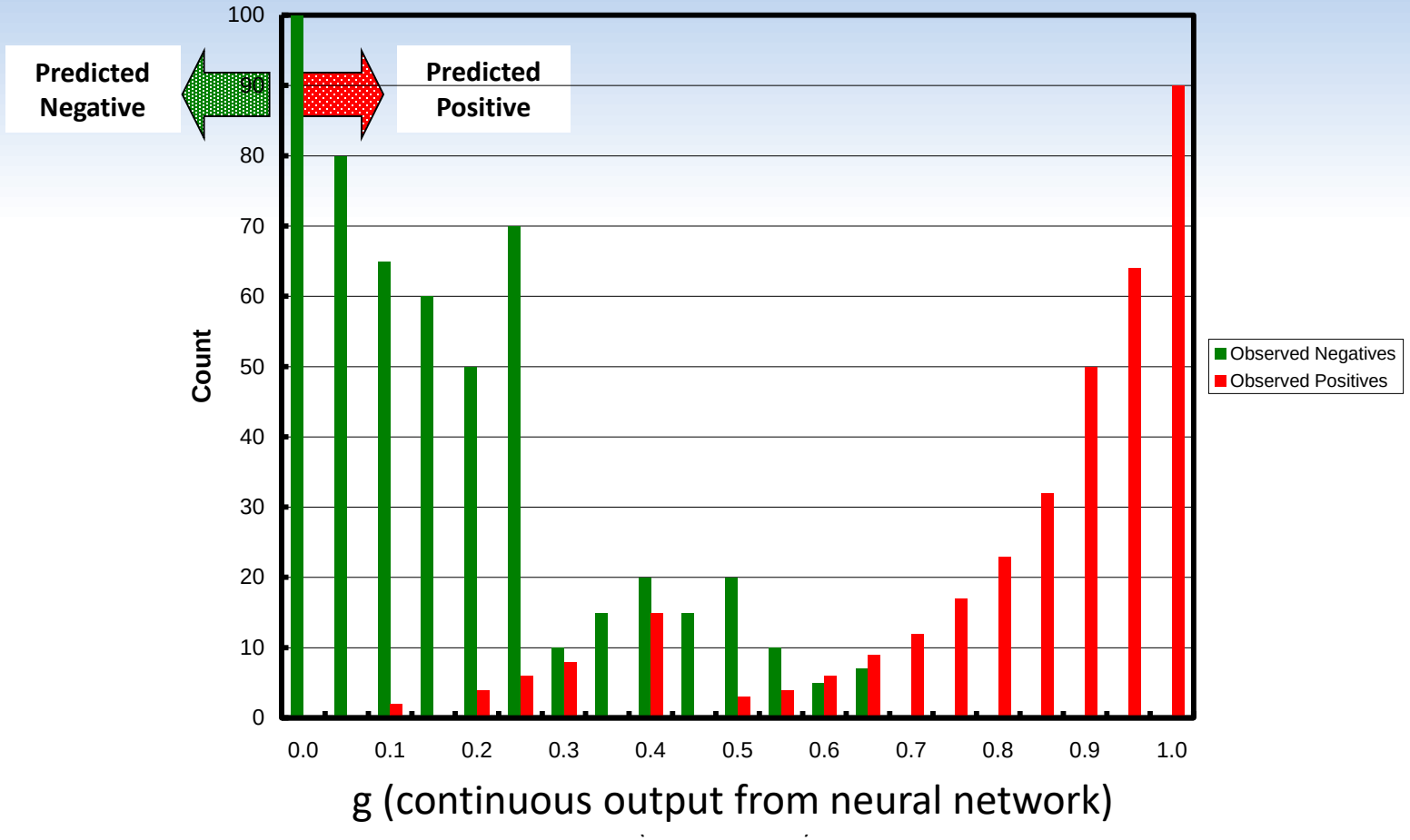
where *c*(*k*) is 0 if observation *k* is in the negative class and 1 if observation *k* is in the positive class.

7

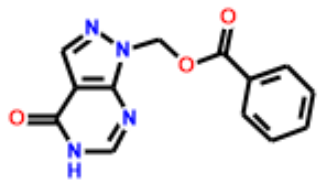
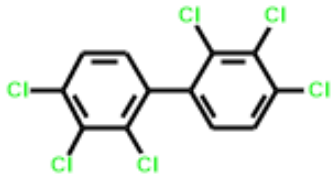
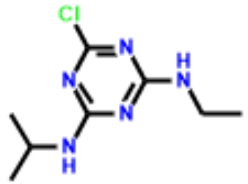
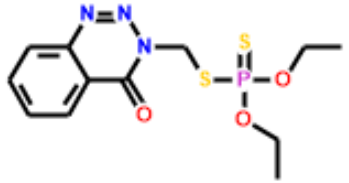
Threshold Adjustment

The neural network outputs a continuous value (g) between 0 and 1 for each compound. The graph below illustrates a possible distribution plot of the observed negatives (green bars) and positives (red bars) of these values.

A threshold value is adjusted to give the best classification statistics



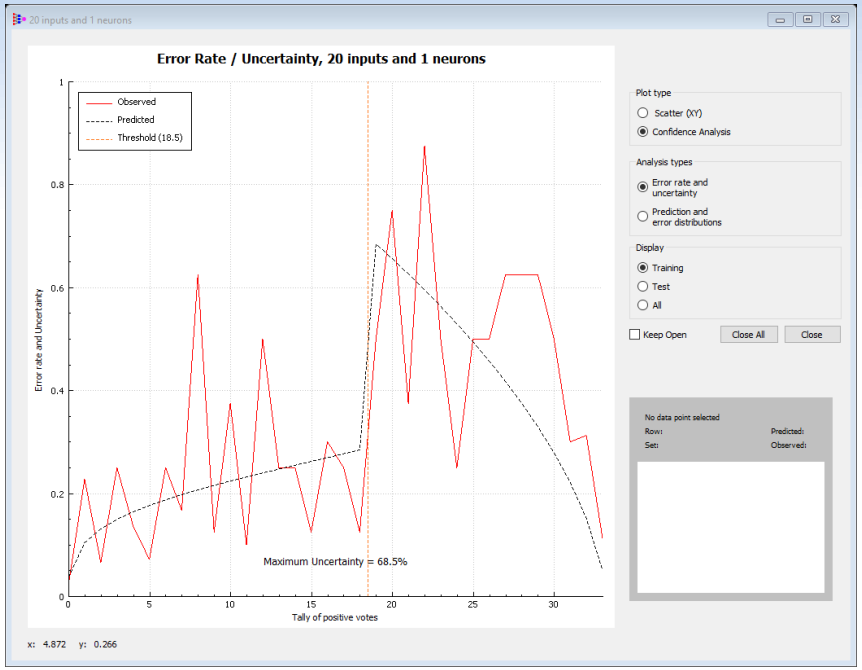
Confidence Estimates

	Structure	Identifier	VeryTox_Model2	VeryTox_Model6
21		Allopurinol, 1-[benzoyloxymethyl]-	FALSE (99%)	FALSE (99%)
22		Aroclor 1260	FALSE (99%)	FALSE (97%)
23		Atrazine	FALSE (99%)	FALSE (99%)
24		Azinphos ethyl	TRUE (57%)	TRUE (62%)
25		Bendroflumethiazide	FALSE (89%)	FALSE (90%)

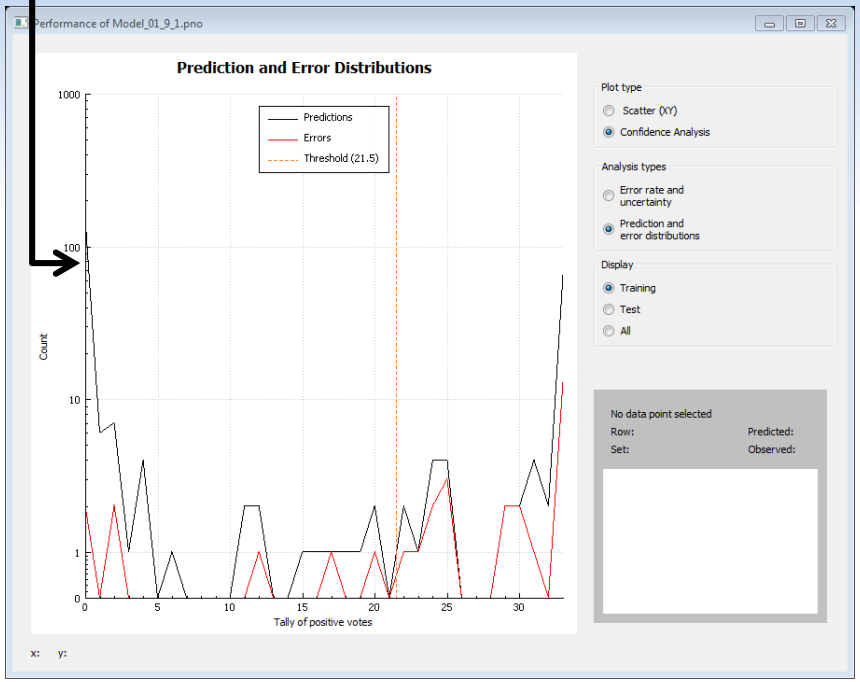
Confidence estimates are computed for binary classification predictions. These are displayed in parenthesis next to the binary prediction.

Confidence Analysis

Prediction uncertainty is based on the degree of concordance among individual neural networks in an ensemble.

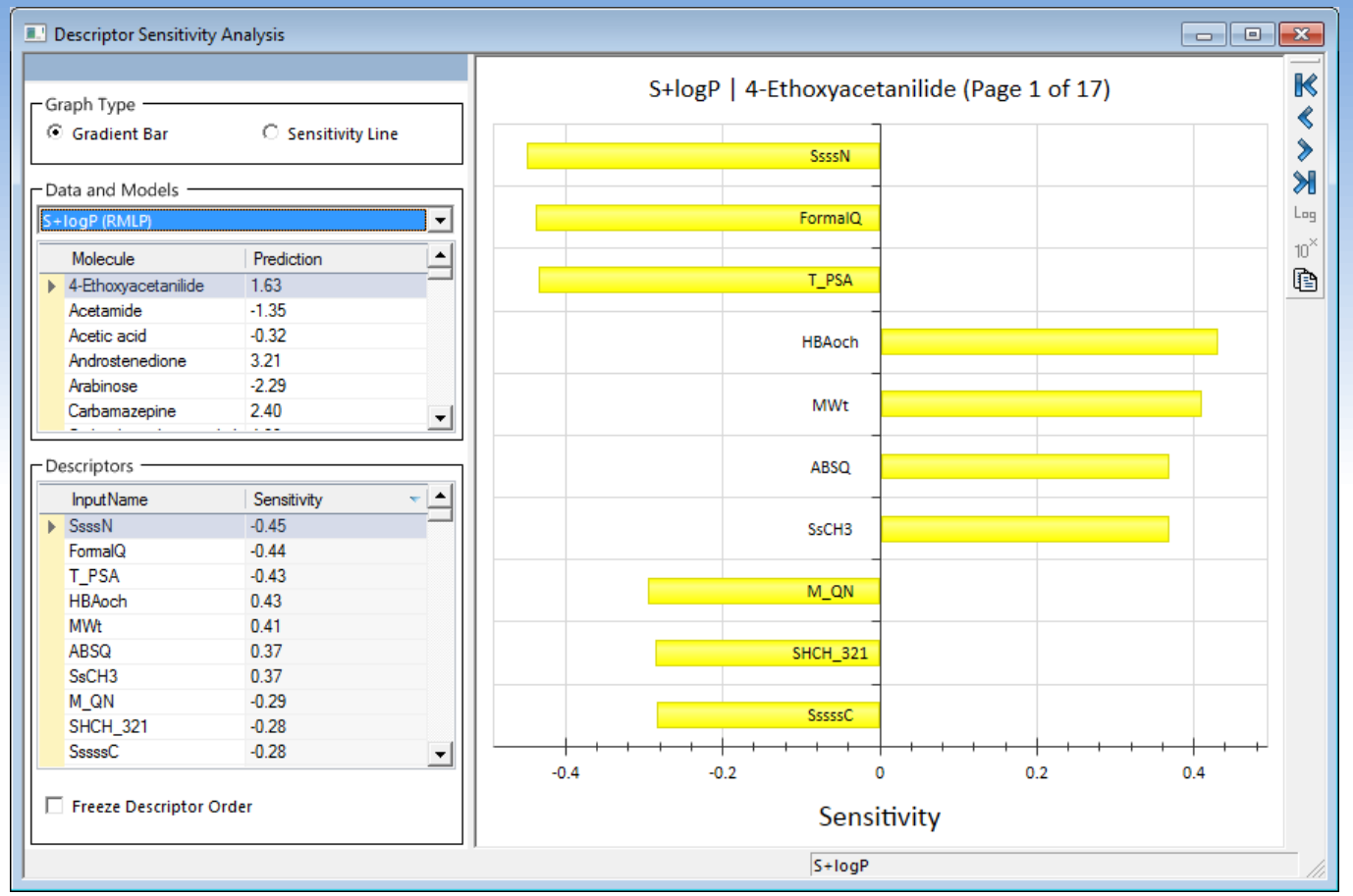


log scale



- Continuity correction for natural error rate
- Observed distributions of experimental positives and negatives are fitted to separate beta binomial distributions

Removing “Black-Box” Stigma with ANN

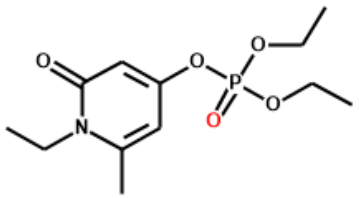
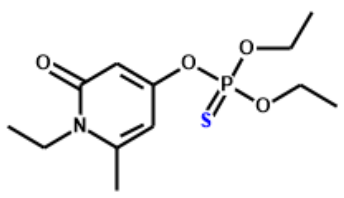
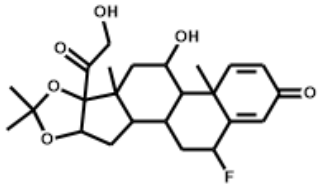
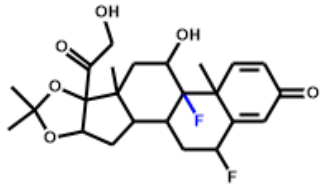
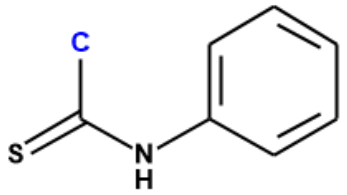
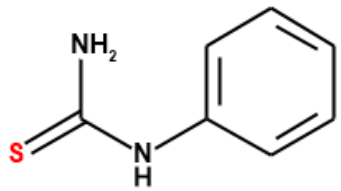
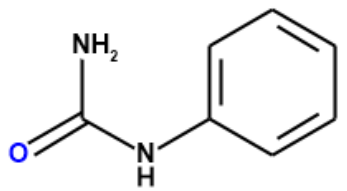
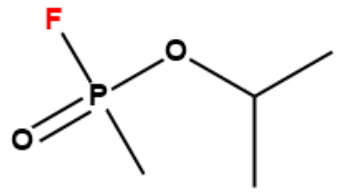
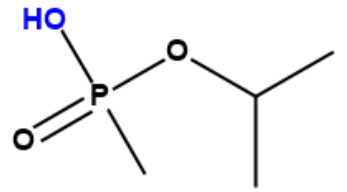


- Descriptor Sensitivity Analysis (DSA) shows the predicted effect of changing each model descriptor on the corresponding property **for the selected molecule**.
  - This provides guidance to the medicinal chemists as to how they may improve a property by changing some structural feature(s) of the molecule

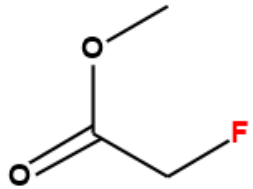
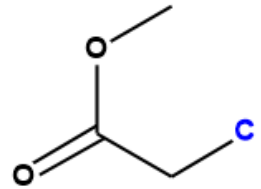
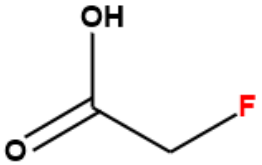
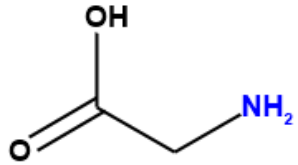
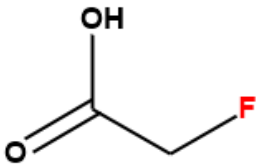
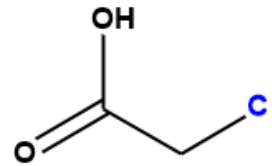
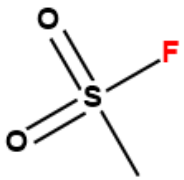
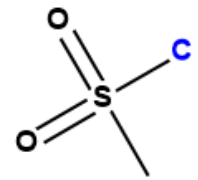
Data Set Curation

- Standardize functional groups and tautomers
- Exclude molecules that contain atoms other than H, C, N, O, P, S, F, Cl, Br, or I
- Find and analyze duplicate structures and neighbors
  - Remove compounds with widely varying endpoint values

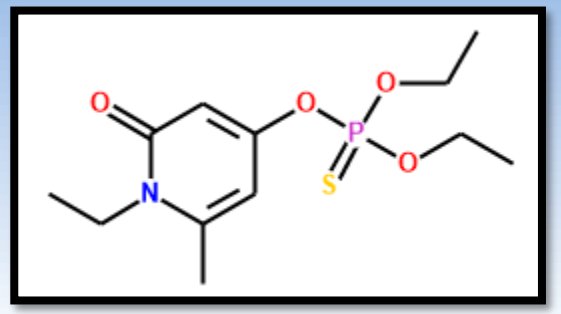
Matched Molecular Pair Analysis

	GHS=1	GHS=5
1		
2		
3		
4		
5		

9 pairs of compounds have a single heavy atom modification that converts the compound from GHS=1 to GHS=5. These pairs represent “activity cliffs”. A small structural change results in a large change in biological activity.

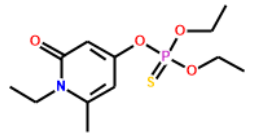
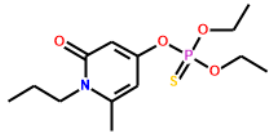
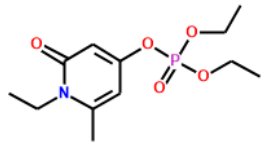
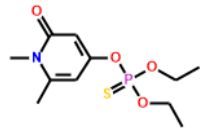
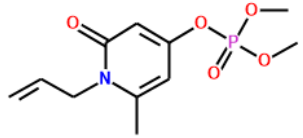
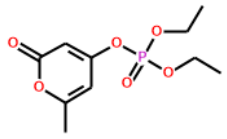
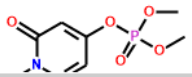
	GHS=1	GHS=5
6		
7		
8		
9		

Widely Varying LD_{50} of Similar Compounds

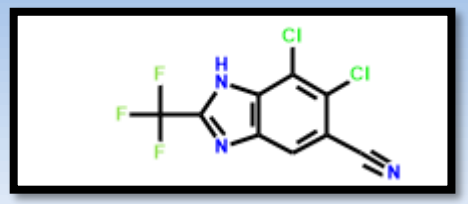


22787-58-2

All the close analogs have single digit LD50 values while this compound has a very high LD50.

	Structure	Identifier	LD50_mgkg
1		22787-58-2	7070
2		21327-31-1	9.75
3		22787-59-3	1.62
4		21409-78-9	5.97
5		22620-72-0	2.46
6		26662-09-9	3.34
7		22787-53-7	2.91

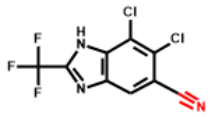
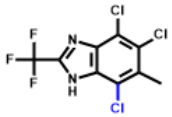
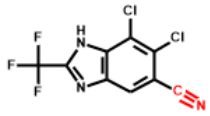
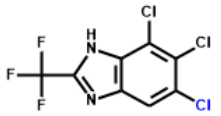
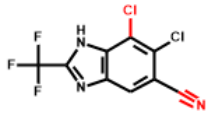
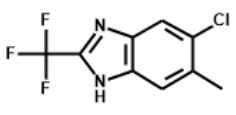
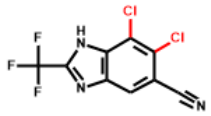
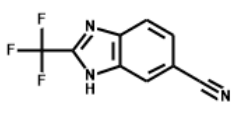
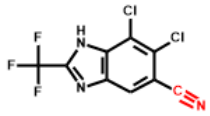
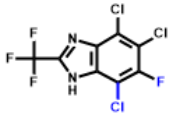
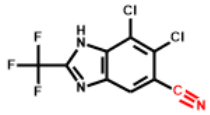
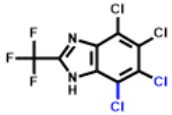
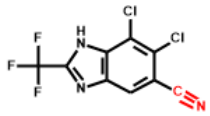
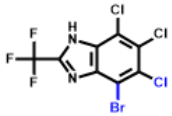
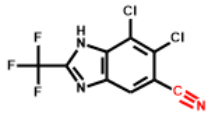
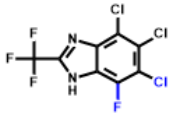
Widely Varying LD_{50} of Similar Compounds



89427-25-8

LD_{50} = 3,955 mg/kg

- 31 most similar structures
  - Tanimoto > 0.8
  - LD_{50} range is 0.245 to 77 mg/kg

Structure 1	Structure 2	Identifier 1	Identifier 2	Mismatches	Sim. Score	Change(LD50_mgkg)
		89427-25-8	4228-99-3	4	0.892	-3936.714
		89427-25-8	2338-27-4	4	0.889	-3953.481
		89427-25-8	14863-40-2	4	0.886	-3900.000
		89427-25-8	3671-61-2	4	0.886	-3950.589
		89427-25-8	18225-94-0	6	0.842	-3952.614
		89427-25-8	2338-29-6	6	0.842	-3954.755
		89427-25-8	4228-93-7	6	0.842	-3950.878
		89427-25-8	89427-26-9	6	0.842	-3954.153
		89427-25-8	89427-36-1	6	0.842	-3948.512

Submitted Models

Endpoint	Type	Training set size	Test set size
EPA class (1-4)	Multiclass	6,531	1,633
GHS class (1-5)	Multiclass	6,951	1,648
LD_{50}	Regression	5,037	1,209
LD50 > 2,000 mg/kg	Binary	7,059	1,246
LD50 ≤ 50 mg/kg	Binary	6,699	1,675

LD_{50} data set is smaller than the others because qualitative values, e.g., >2,000 mg/kg were not included.
Submitted 2 models from each endpoint

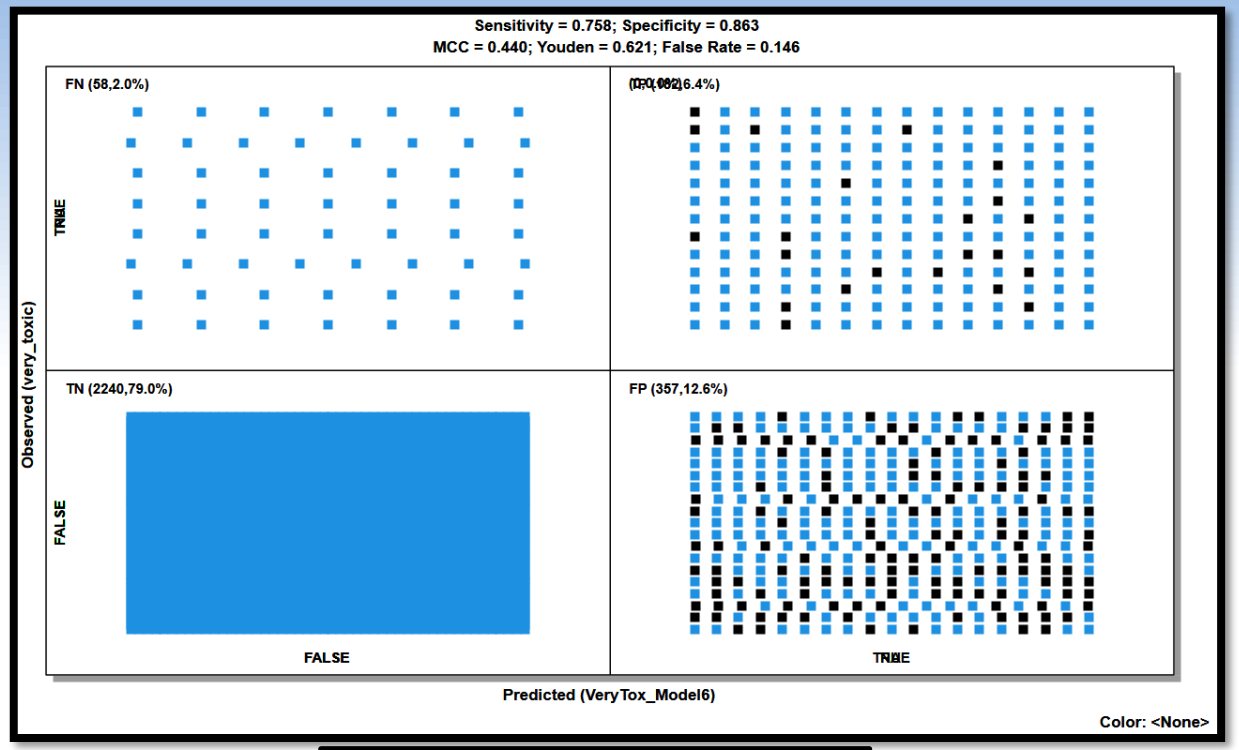
External Validation (Test) Set Predictions

Model	Endpoint	Data Size	Outside AD (%)	Performance stat[1]
EPACat_1	EPA class (1-4)	2812	50 (1.8%)	0.696
EPACat_2	EPA class (1-4)	2812	51 (1.8%)	0.691
GHSCat_1	GHS class (1-5)	2882	51 (1.8%)	0.666
GHSCat_2	GHS class (1-5)	2882	51 (1.8%)	0.671
LD50_1	LD_{50}	2172	41 (1.8%)	0.638
LD50_2	LD_{50}	2172	41 (1.8%)	0.605
NonTox_1	LD50 > 2,000 mg/kg	2887	54 (1.9%)	0.750
NonTox_2	LD50 > 2,000 mg/kg	2887	55 (1.9%)	0.748
VeryTox_1	LD50 ≤ 50 mg/kg	2891	52+166 with low confidence (7.5%)	0.873
VeryTox_2	LD50 ≤ 50 mg/kg	2891	53+185 with low confidence (8.2%)	0.825

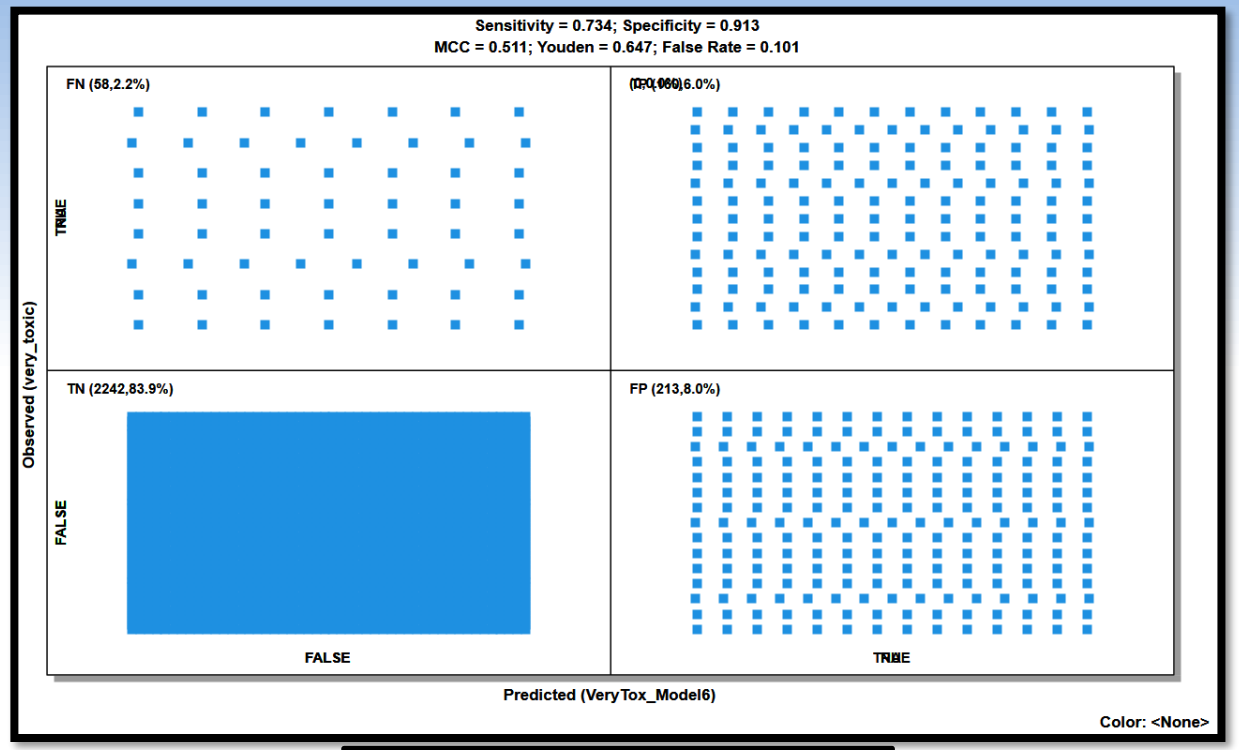
[1]TST_BA (test set balanced accuracy) for EPA, GHS, NT, and VT. TST_RMSE for LD_{50}

Using Confidence Estimates for VeryTox Model

External Prediction (Test) Set



Sn = 0.76; Sp = 0.86



Sn = 0.73; Sp = 0.91

Black points have low confidence (<28%)

Eliminating them decreases sensitivity (Sn) and increases specificity (Sp)

Summary

- Used ANNE technology to develop regression and classification models
- Curation identified activity cliffs and questionable LD_{50} values
- Model applicability domain is defined by the minimum and maximum descriptor values in the training set
- Confidence estimates are included in binary classification models
- Submitted two models for each of the five endpoints