Enabling Science via Analytical Informatics

Mixture-based Modeling of Chemical Ocular Toxicity Based on US EPA Hazard Categories

A Sedykh[1], N Choksi[2] , D Allen[2] , N Kleinstreuer[3] , W Casey[3], RR Shah[1] | [1]Sciome LLC, [2]ILS , [3]NICEATM, Research Triangle Park, NC

Advancing Science, Improving Lives

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

Abstract

For further information please contact
ruchir.shah@sciome.com

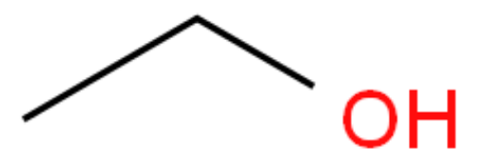
Computational prediction of eye irritation and corrosion potential of chemicals is one of the key strategies for animal-free evaluation of ocular toxicity. Over the years, the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) has compiled and curated a database of *in vivo* eye irritation studies from scientific literature and provided by stakeholders. The database contains around 800 annotated records of over 500 unique substances with their eye irritation categories according to Global Harmonized System (GHS) and US Environmental Protection Agency (EPA) hazard classifications. We developed a set of *in silico* models for EPA hazard classification categories at 100% and 10% potency thresholds (by mass or volume content) for the chemical substances in the eye irritation database, many of which are formulations and mixtures. Conventional models (based on chemical structure of the largest component of the test substance) achieve validated balanced accuracy in the range of 67-77% and 84-89% for the 100% and 10% potency thresholds, respectively. Comparatively, the mixture-based models, which account for all components in the substance by weighted feature averaging, showed higher accuracy of 69-78% and 85-91% for the respective potency thresholds. We also noted a strong trend between the pH feature metric calculated for each substance and its activity category. Namely, across all the models, calculated pH of inactive substances is on average 0.8 pH-units away from the neutral pH, while for active substances, it is >3 pH-units away. This pH dependency is especially important for complex substances that contain multiple components. In the future, these *in silico* models can benefit from additional high quality *in vivo* data sources (e.g., European Chemicals Agency dossiers) and by including additional variable inputs such as *in vitro* eye irritation test method results.

NICEATM Ocular Toxicity Data ("OcuTox DB")

- **810** curated data records with *in vivo* ocular toxicity (EPA and/or GHS categories) for **594** unique test substances (including cosmetics chemicals and formulations).
- Around **77%** of test substances are single compounds, while **~23%** are either salts or mixtures.
- Around **64**% of test substances occur once in the database, while **~36%** have multiple records (such as reports from different sources or results for different test doses from a single study).

Data record examples

Ethanol
CAS RN#: **64-17-5**



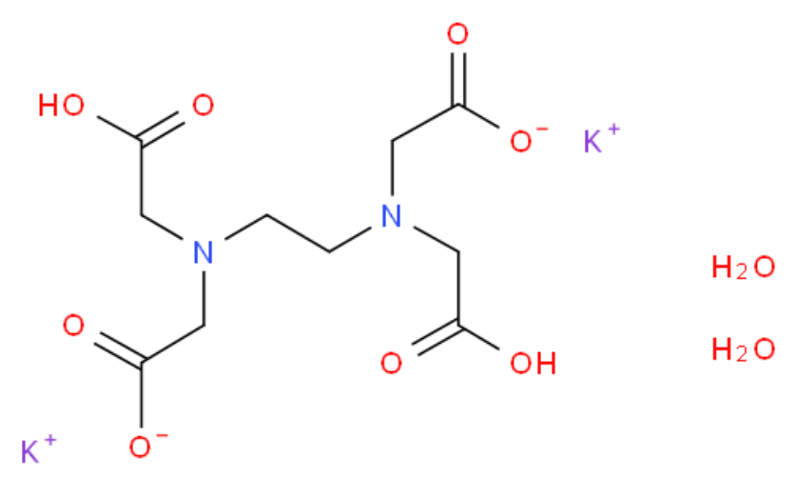
at 10% dose:
GHS: No Category
EPA: Category IV

at 79% dose:
GHS: Category 2B
EPA: Category III

at 100% dose:
GHS: Category 2A
EPA: Category I
EPA: Category II
EPA: Category III

EDTA, dipotassium
CAS RN#: **25102-12-9**

at 20% dose:
GHS: No category
EPA: Category III

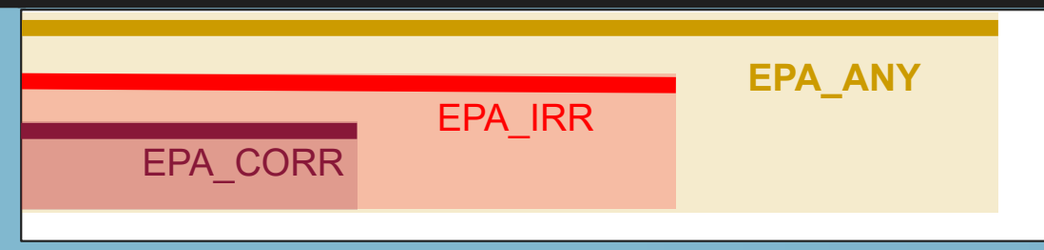


Eye Toxicity Hazard Classifications

In vivo effect on eye tissues	EPA	GHS
Corrosive or not reversible in 21 days	Category I	Category 1
Irritation, reversible in 8-21 days	Category II	Category 2A
Irritation, reversible in 1 – 7 days	Category III	Category 2B
Minimal effects, disappearing in 24h	Category IV	No category

Concordance of EPA vs GHS calls across data records

	EPA categories				
	I	II	III	IV	No data
GHS Cat.1	135	3			56
GHS Cat.2A	3	29	10		36
GHS Cat.2B		3	37		8
GHS No Cat.		6	114	201	72
GHS No data	2	2	10	1	62



Activity label binning schemes for binary classification models

We have assigned binary activity labels: EPA_CORR (Cat. I), EPA_IRR (Cat. I-II) and EPA_ANY (Cat. I-III) at two dose levels (10% and 100%) to **515** qualified substances.

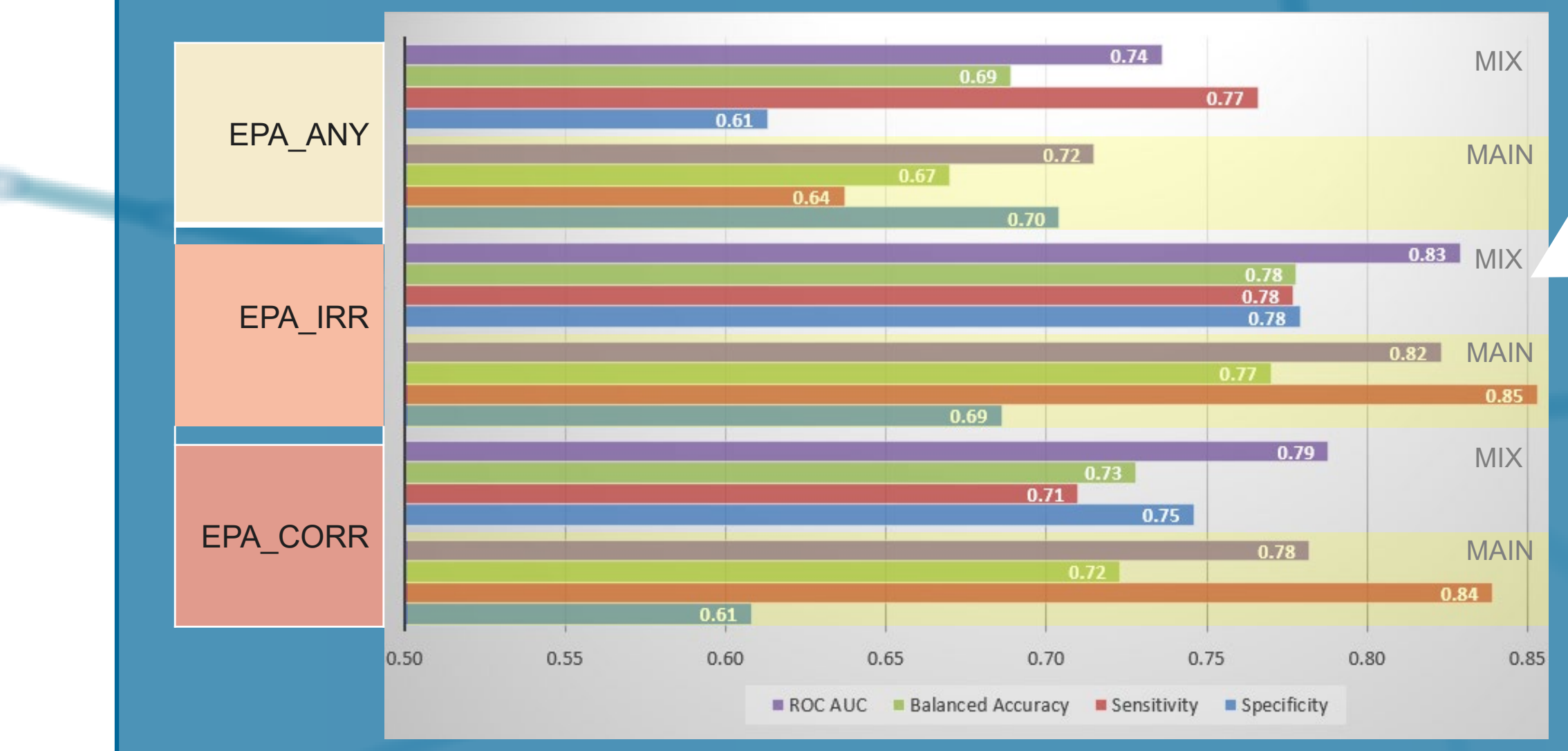
- ~20% of these labels were based on multiple data records.
- 6-12% of the above were discrepant (depending on label scheme and dose cut-off), leading to 1-2% of potential label-errors in the finalized datasets. For those cases we took most conservative call (highest EPA category reported).

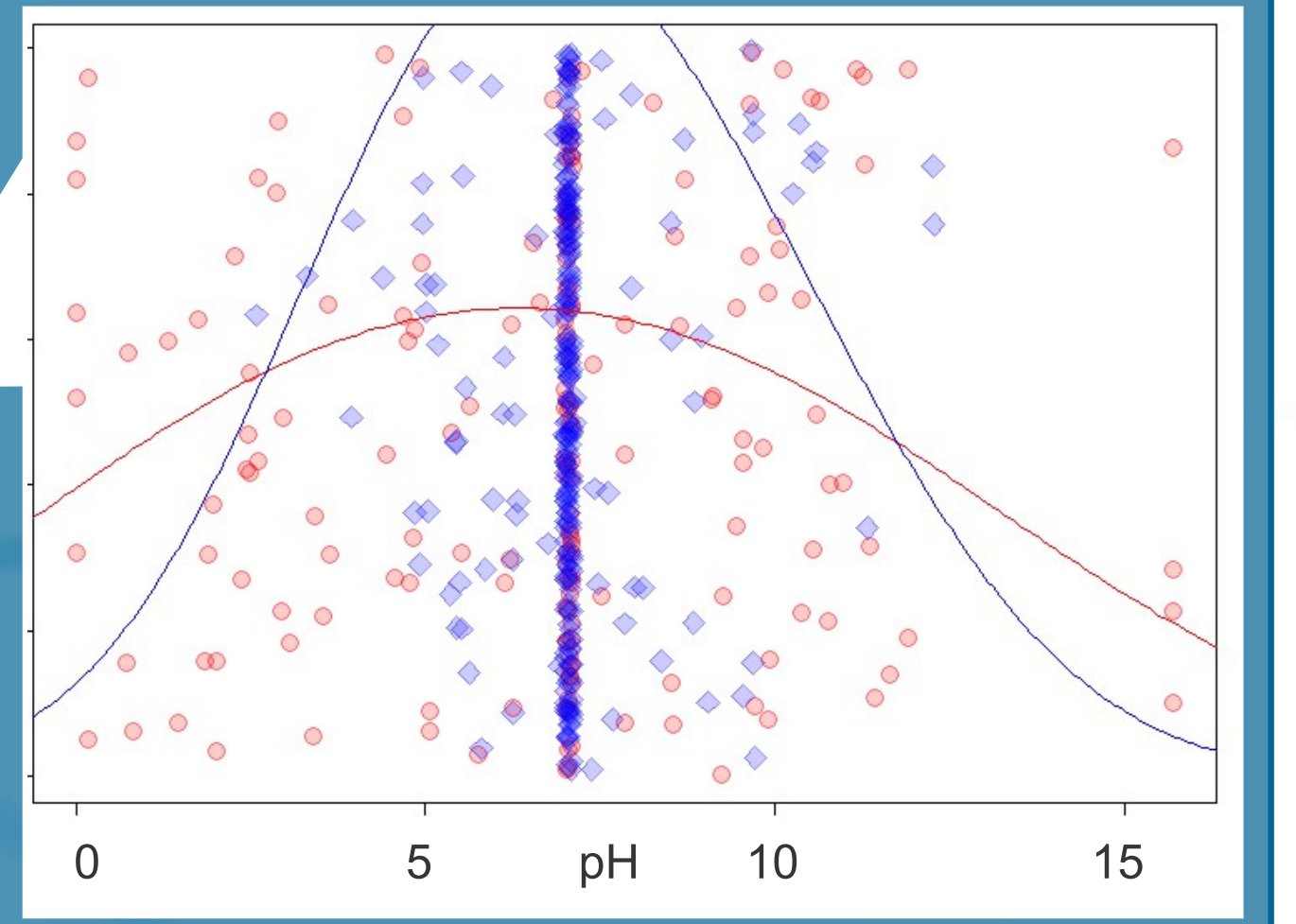
We note that EPA calls are, in general, more conservative, but when absent, GHS calls were used where appropriate.

Comparative Performance of OcuTox Models

"MAIN" (conventional) and "MIX" (mixture-based) QSAR models are compared by their out-of-bag performance for classification (sensitivity, specificity, and their average as balanced accuracy) and prioritization (by area under ROC curve) tasks.

- Models based on 100%-dose cut-off

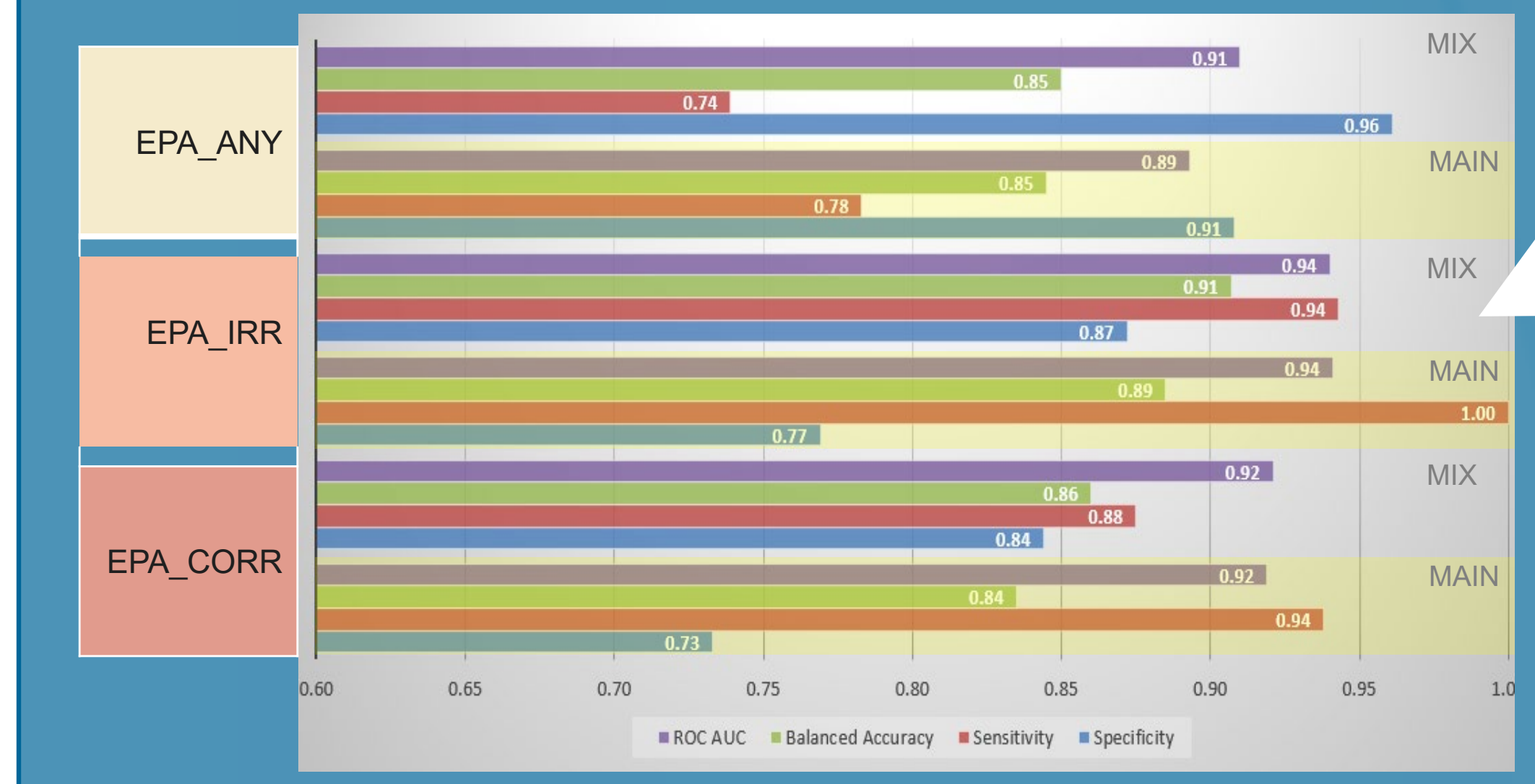


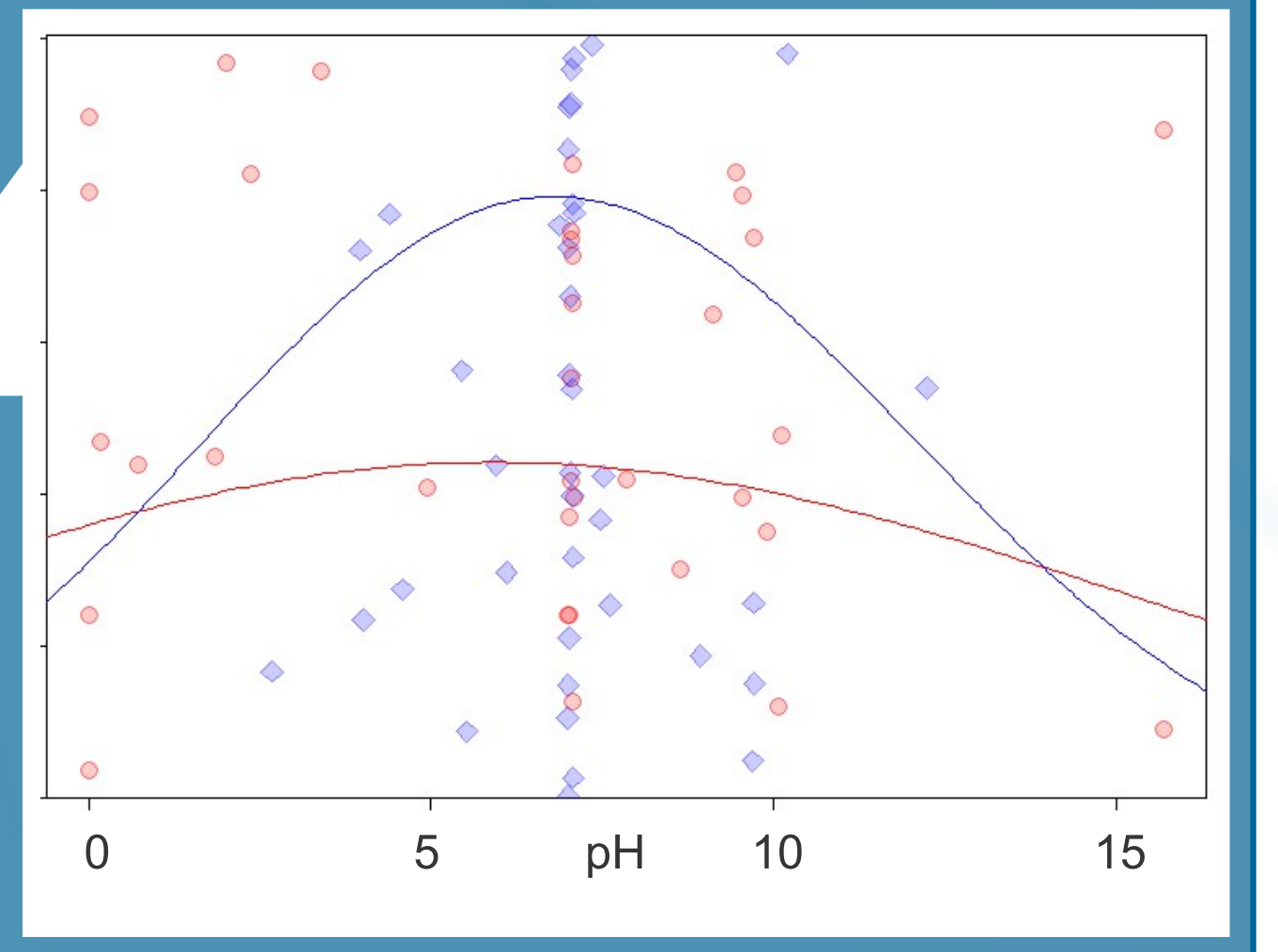


Density distribution of active and inactive substances in 100% and 10% training sets by their calculated pH value for EPA_IRR (MIX) models

○ active
◇ inactive

- Models based on 10% dose cut-off





- Above scatter plots show pH distributions for active and inactive substances. Most of the inactive substances cluster in the middle, neutral pH area, while many actives are well spread to the extremes (strong acid or alkali) of the pH values scale.

Constructing Binary OcuTox Datasets

Based on OcuToxDB, for each of the three endpoints (EPA_CORR, EPA_IRR, EPA_ANY), we formed two binary (*e.g.*, corrosives vs non-corrosives) datasets of unique, curated substances at two test doses ("potencies"): 10% and 100%.

Finalized ocular toxicity datasets and their composition

Dataset name	Dose cut-off	Endpoint	Inactive	Active
OCU_EPA_CORR_C	100% - 'C'	EPA_CORR	**311**	**155**
OCU_EPA_IRR_C		EPA_IRR	**258**	**184**
OCU_EPA_ANY_C		EPA_ANY	**142**	**333**
OCU_EPA_CORR_X	10% - 'X'	EPA_CORR	**45** *	**32**
OCU_EPA_IRR_X		EPA_IRR	**39** *	**35**
OCU_EPA_ANY_X		EPA_ANY	**152**	**46**

NB: Active calls at 10% were also used as active at 100%; inactive calls at 100% were also used as inactive at 10%
* retained based on the structural similarity of 100%-dose inactives to the corresponding 10%-dose actives

Modeling Details

Chemical features

- Mordred descriptors (github.com/mordred-descriptor/mordred)
- Structural alerts (Chemotyper, SMARTS for heavy metals and electrophiles)
- pH, acidity and basicity features (ADMET Predictor)

Substance representation approaches

- **MAIN** – largest chemical component (conventional approach)
- **MIX** – fraction-weighted average of features for all components

Machine learning method

Random Forest models with out-of-bag validation (33% of external data)

Conclusions

- Mixture-based models slightly outperform conventional QSAR versions, which is likely due to the higher accuracy of the mixture approach for tested formulations (~20% of data), especially when those act simply as acidic or basic agents on the ocular tissues.
- Models based on 10%-dose threshold show better performance. However, these are based on much smaller datasets, which limits their utility.
- For both dose thresholds (10% and 100%) and approaches (MAIN and MIX), the EPA_IRR scheme (EPA Categories I-II defined as active) achieves higher accuracy than other binning schemes.

Acknowledgements

This project was funded with federal funds from the National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN273201500010C.