Hierarchical QSAR Modeling for Predicting Acute Oral Systemic Toxicity

Xinhao Li[1], Nicole C. Kleinstreuer[2], and Denis Fourches[1,*]

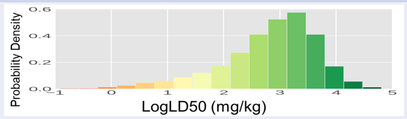
[1] *Department of Chemistry, Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695, United States*

[2] *National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, NIEHS, RTP, North Carolina 27709, United States.*

Background

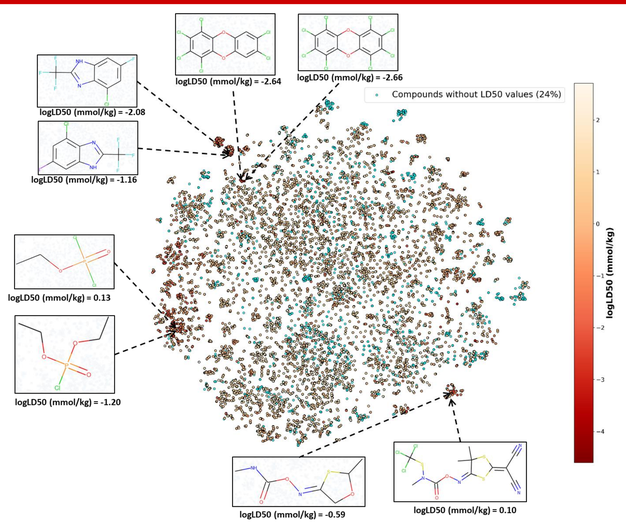
- Thorough toxicity evaluation is an important step to ensure the environmental safety of chemicals. *In vivo* procedures are not only costly and time-consuming, but the use of animals raises ethical issues and often has questionable relevance to human biology.
- Reliable *in silico* approaches to replace, reduce and refine animal testing for the evaluation of potential acute toxic effects are highly demanded by regulatory agencies.
- Quantitative Structure-Activity Relationship (**QSAR**) modeling is a major *in silico* approach relying on **machine learning techniques** and a set of **molecular descriptors** directly computed from chemical structures.

Rat Oral Toxicity Dataset*

Endpoints	Training Set (75%)				Test Set (25%)
Total	8,211				2,842
LD50 (mg/kg)	6,089				2,144
Toxic/Nontoxic	8,209	Toxic (LD50 < 2,000)		Nontoxic (LD50 > 2,000)	2,820
		4,661		3,548	
EPA Category	8,126	I	II	III	IV	2,842
		<50	50-500	500-5,000	>5,000	
		703	1,767	4,070	1,586	

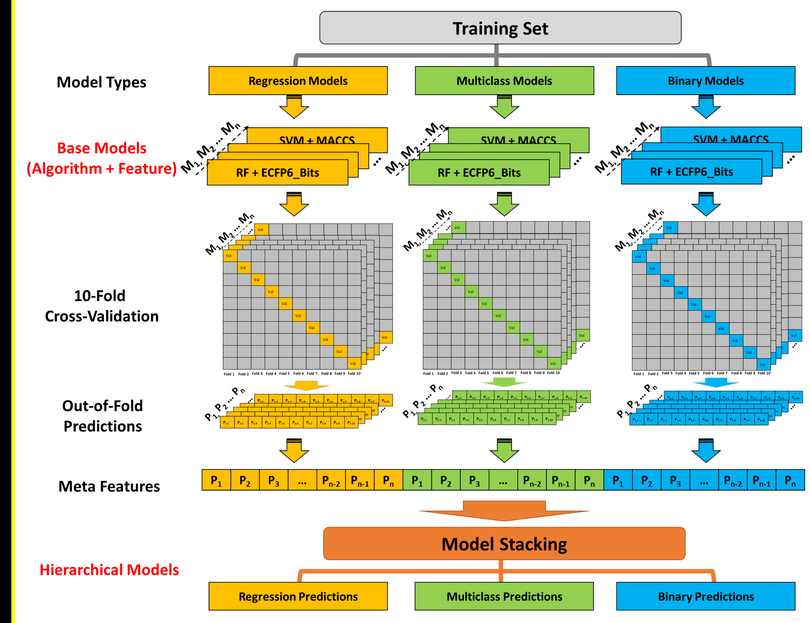
*The full description and the data is available at https://ntp.niehs.nih.gov/go/tox-models

Visualization of the Chemical Space



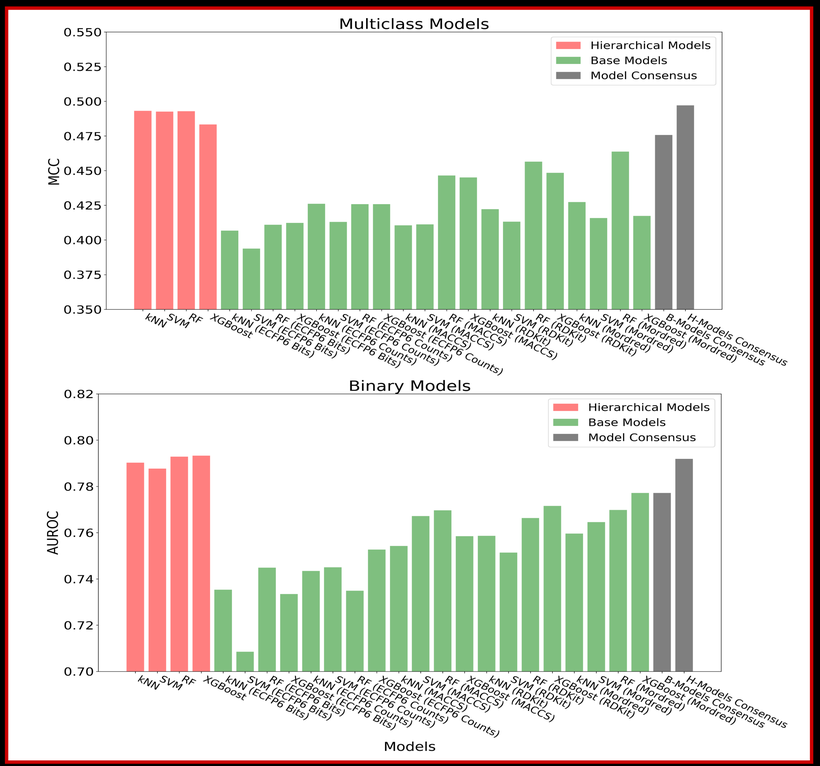
Clustering of the full dataset (11,056 molecules) by t-SNE with 2D bit-based ECFP6 (2,048 bits)

Hierarchical QSAR Modeling

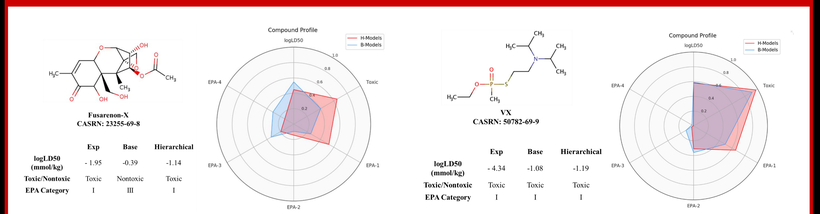


Overall workflow for building ***hierarchical* QSAR models**. ***Base*** regression, binary and multiclass models are built with diverse combinations of ***four*** machine learning algorithms (**RF, SVM, kNN and XGBoost**) and ***five*** chemical descriptors/fingerprints (**ECFP6_Bits, ECFP6_Counts, MACCS keys, RDKit descriptors and Mordred descriptors**). For each endpoint, a diverse set of 20 models were built at this stage (60 base models total). **Out-of-Fold Predictions** of base models are generated through 10-fold cross-validation. The out-of-fold predictions are concatenated together and used as input (**Meta Features**) for building hierarchical regression, binary and multiclass models.

Performances on External Test Set



Examples of Chemicals in External Test Set



Conclusion

- A dual-layer hierarchical QSAR modeling protocol was developed and applied to three acute oral systemic toxicity endpoints. The hierarchical models outperform the individual base models on all the three endpoints.
- Hierarchical H-QSAR modeling method relying on the full stacking of binary, multiclass, and regression models represents a promising approach for *in silico* chemical risk assessment and more generally, for blending individual QSAR models

Acknowledgement

NC State Chancellor's Faculty Excellence Program, NIEHS, CHHE.