ACUTE ORAL SYSTEMIC TOXICITY

Consensus approach for modeling acute systemic oral toxicity and LD50 data using machine-learning and *in silico* approaches

Dr. Ahmed Abdelaziz Sayed | 11 Apr 2018

Acute Oral Systemic Toxicity

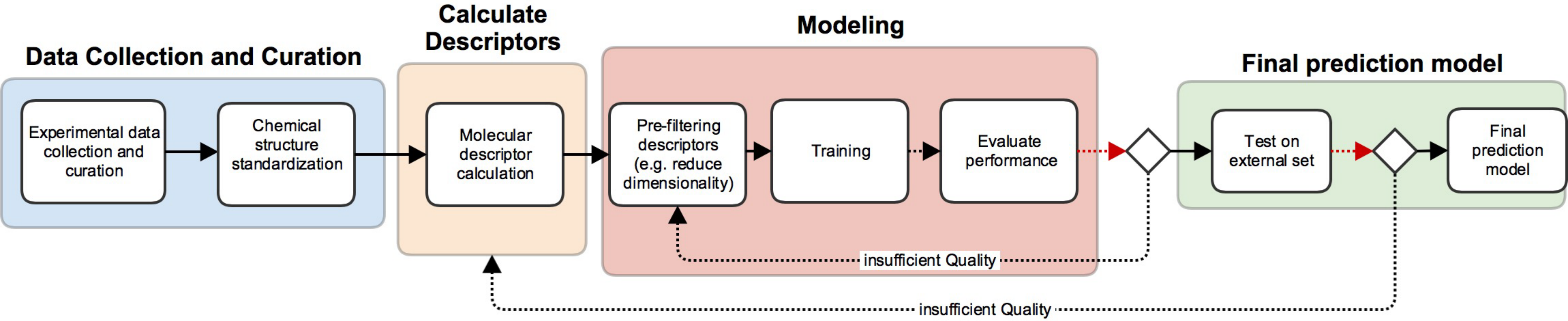
1. Acute toxicity studies occur over a short duration (≤ 24 h)
2. Intended to provide initial indication of toxicity in target organs and establish some dose-response relationships → **Not a single mechanism of toxicity**
3. The dose selection strategy is often to identify a low dose that causes no-effect and high dose that causes significantly adverse effects. → **high variability, usually multiple numbers are reported**
4. Rodent species are generally used for these studies.

Aim

1. Promotes the use of validated QSARs for regulatory purposes to replace animal use
2. Assists the selection of chemicals for prioritizing evaluation
3. Models can be used to investigate or confirm hazards or better understand mechanisms of toxicity

QSAR Model Building and Validation

To achieve OECD compliant model



Data Collection and Curation
Experimental data collection and curation
Chemical structure standardization
Calculate Descriptors
Molecular descriptor calculation
Modeling
Pre-filtering descriptors (e.g. reduce dimensionality)
Training
Evaluate performance
insufficient Quality
Final prediction model
Test on external set
Final prediction model
insufficient Quality

NICEATM /NCCT

Remove mixtures /inorganics, Chemotype normalization, Tautomers and duplicate, Manual inspection

0D, 1D, 2D, and 3D

R >0.95 Variance <0.01 Scaling

DNN, ASNN, FSMLR, k-NN, PLS, C4.5 (J48), RF

Balanced accuracy, AUROC Bootstrap aggregation

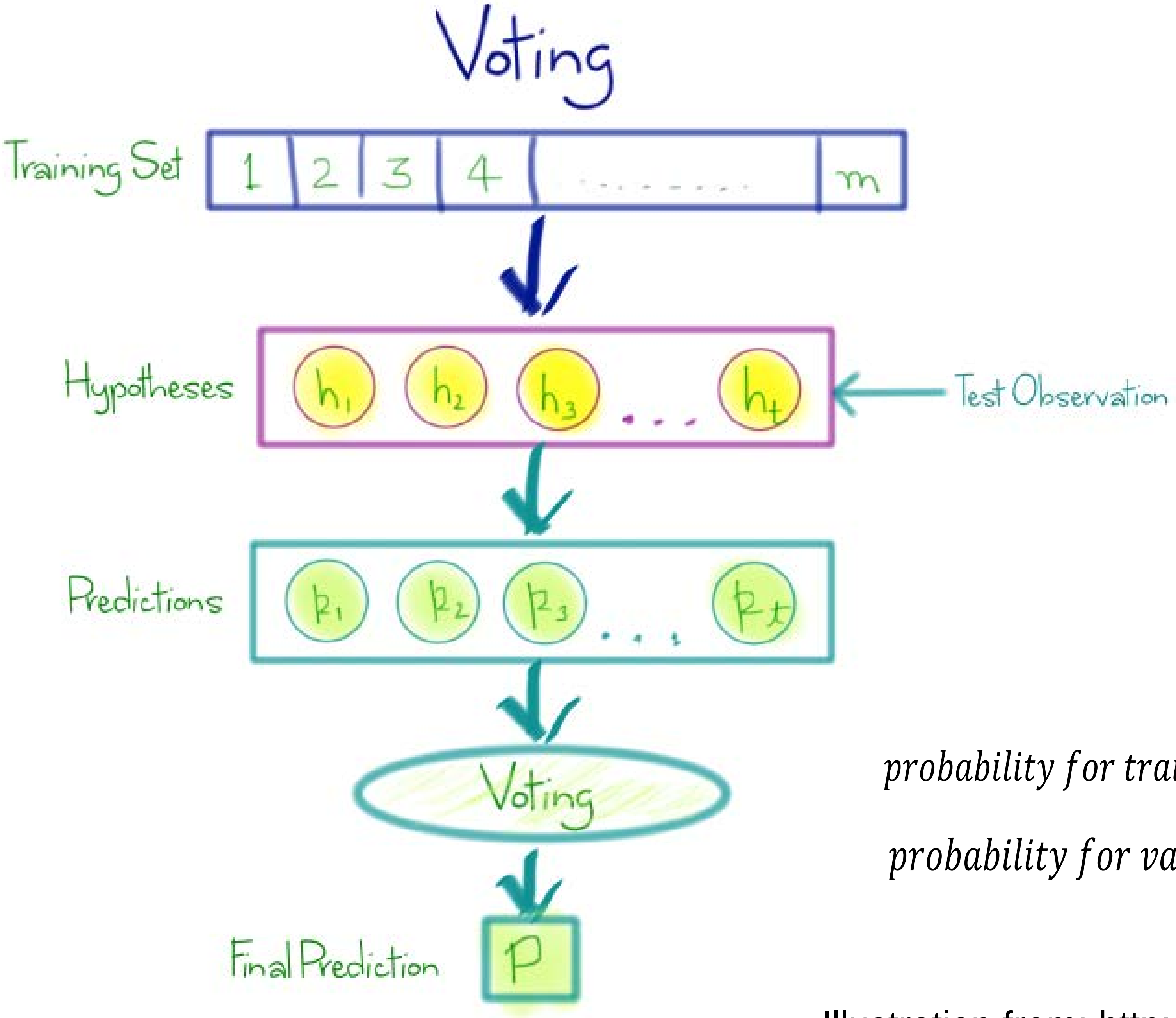
Descriptor packages

1. Five packages were selected from OCHEM: OEstate indices & AlogPS, Dragon, CDK2, as well as structural alerts.
2. Filtered out descriptors that are: constant or low variance (< 0.01)
3. Grouped descriptors with pair-wise Pearson's correlation coefficient (R) > 0.95

Learning algorithms

1. Seven algorithms were used: Deep neural networks (DNNC), Associative neural networks (ASNN), k-Nearest Neighbors (*k*NN), partial least squares (PLS), Fast stagewise multiple linear regression (FSMLR), Random forests (RF), and C4.5 decision tree (J48)
2. For each endpoint, more than 30 QSAR models were compared
3. In case of LD50 point-estimates, very-toxic and non-toxic classifications, the final model was a consensus among multiple underlying models.

Bootstrap aggregation



- Meta algorithm - aggregation of many models
- Validate the accuracy of the training set
- Utilizes random sampling with repetition

$$probability\ for\ training\ set\ selection = 1 - e^{-1} \approx 63.2\%$$

$$probability\ for\ validation\ set\ selection = e^{-1} \approx 36.8\%$$

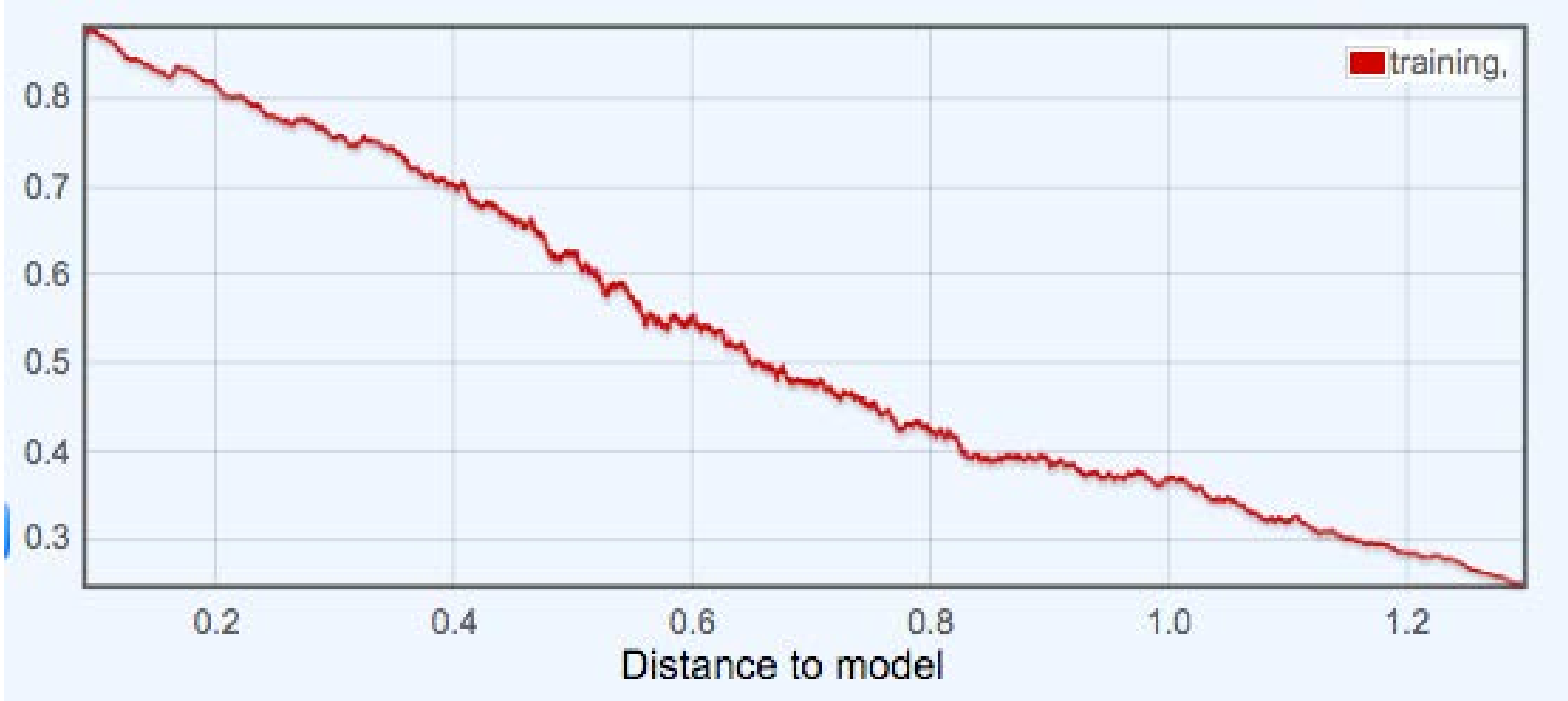
Illustration from: http://manish-m.com/?p=794

Bagging improves predictive ability

1. Overcome the unbalance between the classification classes (by under-sampling the majority class in the sub-models; stratified bagging)
2. Avoids bias in the learning process by calculating out-of-bag statistics (internal validation)
3. Balanced accuracy (for classification) and validated goodness-of-fit (Q2) was used for comparing models.
4. The bagging-standard deviation provides a distance-to-model measure that correlates with the model predictive power

Bagging-STD as a distance to model

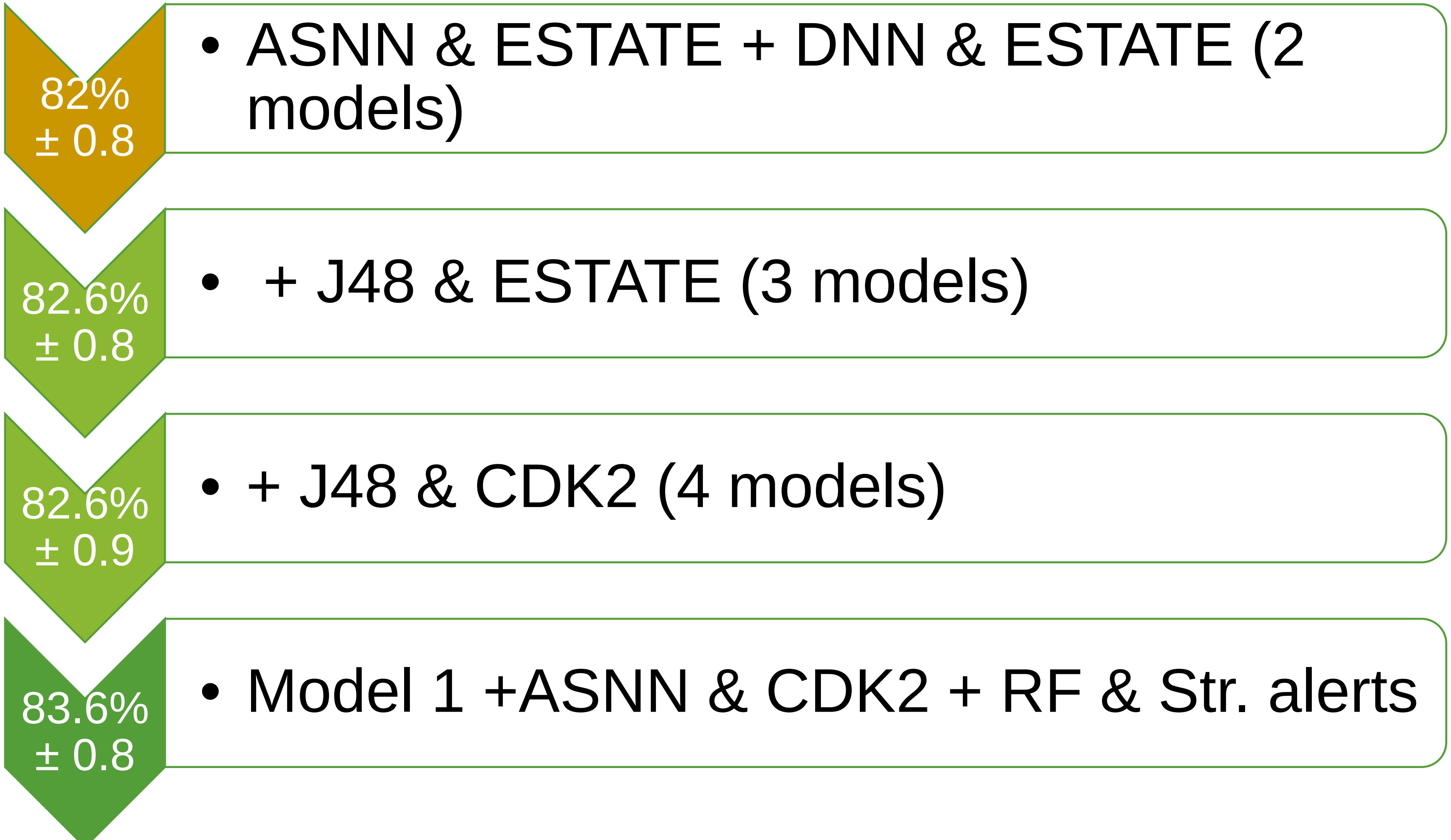
GHS classification model as an example



Williams plot showing distance-to-model applicability domain using **bagging-STD** as a distance measure for chemical categorization using GHS classifications

Consensus Improves predictive ability

Very-Toxic classification as an example



82% ± 0.8

- ASNN & ESTATE + DNN & ESTATE (2 models)

82.6% ± 0.8

- + J48 & ESTATE (3 models)

82.6% ± 0.9

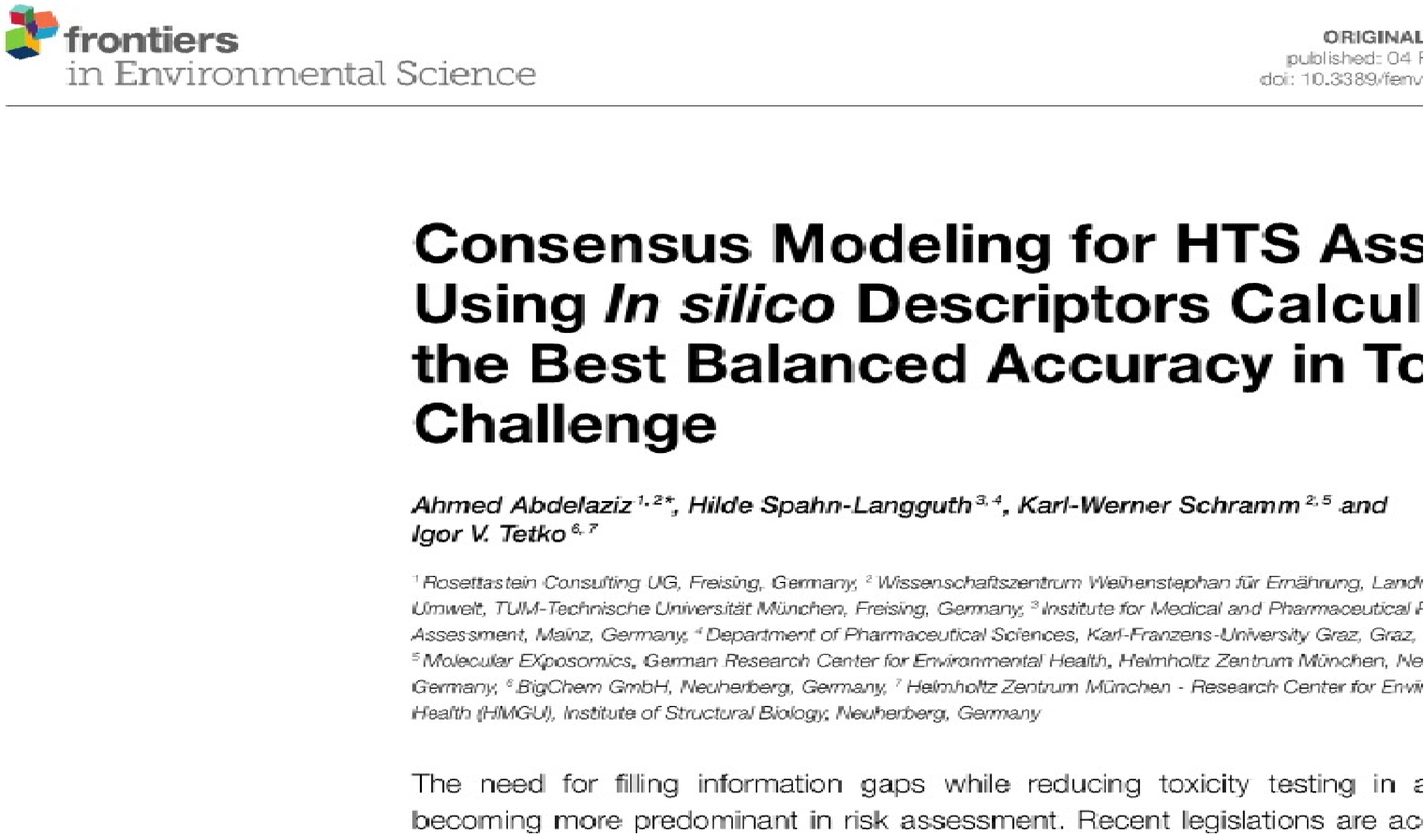
- + J48 & CDK2 (4 models)

83.6% ± 0.8

- Model 1 +ASNN & CDK2 + RF & Str. alerts

Tox21: Prediction of Molecular pathways' perturbation

Best Balanced Accuracy award in Tox21 Challenge



frontiers
in Environmental Science

ORIGINAL
published: 04 F
doi: 10.3389/fenv

Consensus Modeling for HTS Ass
Using *In silico* Descriptors Calcul
the Best Balanced Accuracy in To
Challenge

Ahmed Abdelaziz[1,2], Hilde Spahn-Langguth[3,4], Karl-Werner Schramm[2,5] and Igor V. Tetko[6,7]*

[1] Rosettastein Consulting UG, Freising, Germany, [2] Wissenschaftszentrum Weihenstephan für Ernährung, Land
Umwelt, TUM-Technische Universität München, Freising, Germany, [3] Institute for Medical and Pharmaceutical F
Assessment, Mainz, Germany, [4] Department of Pharmaceutical Sciences, Karl-Franzens-University Graz, Graz,
[5] Molecular EXposomics, German Research Center for Environmental Health, Helmholtz Zentrum München, Ne
Germany, [6] BigChem GmbH, Neuherberg, Germany, [7] Helmholtz Zentrum München - Research Center for Envi
Health (HMGU), Institute of Structural Biology, Neuherberg, Germany

The need for filling information gaps while reducing toxicity testing in a
becoming more predominant in risk assessment. Recent legislations are ac

Other criteria for consensus member selection

The criteria for consensus members selecction included (besides improvement in predictive power):

- Orthogonally of descriptor packages (value-of-information)
- Coverage of wider number of molecules (less calculation errors)
- Simplicity of calculation and interpretability (packages with fewer number of descriptors were preferred).
- Open descriptor packages were more favored to commercial ones to encourage adoption by the scientific community and regulators

Results

Datasets	Validated balanced accuracy	AUROC
Very-Toxic	83.6% ± 0.8 (Test: 0.86)	0.903 ± 0.01
Non-Toxic	79% ± 0.5 (Test: 0.79)	0.877 ± 0.01
EPA	58.5% ± 0.6 (Test: 0.74)	0.249 ± 0.01
GHS	49.1% ± 0.9 (Test: 0.69)	0.153 ± 0.01

Datasets	Q^2	RMSE	MAE
Training	0.33 ± 0.03	3.4 ± 0.2	1.67 ± 0.04
Test	0.17	0.77	

Tips for modelers

1. Only use validation statistics (not fitting) to guide your model building.
2. Models are as good as the underlying data. Access to additional data is always helpful (also can provide unfair advantage in competitions)
3. Modern machine-learning is about building consensus of models (usually consensus of ensemble models).
4. Make sure you understand the endpoint of concern that you try to build a model for.

Summary

- Different algorithms vary in their performance; but within a limit
- Bagging validation provided a good indication for the models' predictive power on external validation sets
- Bagging-STD provides a measure for the distance-to-models
- Consensus modeling improved the predictive ability of models
- The developed models are made publicly available at:

  https://amaziz.com/oraltox/

Summary (2)

- Predictive models for “very-toxic” and “non-toxic” classifications have high accuracy (both on bagging-validation and external test set) that can justify their use in regulatory context.
- The EPA and GHS classification models exhibit good predictive power but it is hard to pin-point the exact class (multi-class classification problem). They might still be useful for prioritization and risk assessment.
- The Point estimate prediction for LD50 was too poor to justify its regulatory use.

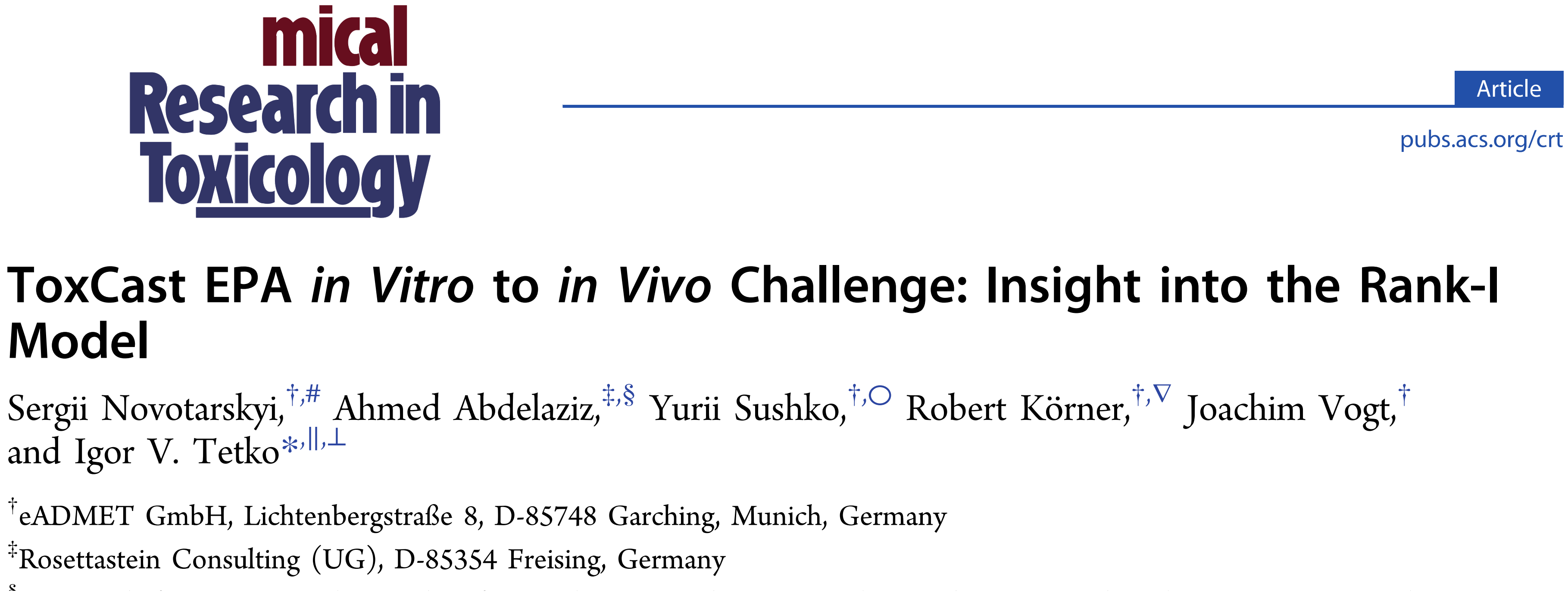
Future Research

Adverse Outcome Pathways, Integrated Testing Strategies and Toxicogenomics

- Investigating the value of inductive knowledge transfer for neural networks by using the same network to train related endpoints.
- Investigate models performance across different chemical classes.
- Use Bayesian statistics to build interpretable integrated strategy for supporting risk assessment for systemic oral toxicity.
- Extend the work to more end-points of regulatory interest.
- Incorporate data from cellular HTS and toxicogenomics assays.

Food for thought

- What is the ultimate goal to reach in terms of predictive power? Usually the answer is experimental variability.
- How well can we model endpoints such as LD50, LEL?



mical Research in Toxicology

Article

pubs.acs.org/crt

ToxCast EPA *in Vitro* to *in Vivo* Challenge: Insight into the Rank-I Model

Sergii Novotarskyi,[†,#] Ahmed Abdelaziz,[‡,§] Yurii Sushko,[†,○] Robert Körner,[†,∇] Joachim Vogt,[†] and Igor V. Tetko[*,∥,⊥]

[†] eADMET GmbH, Lichtenbergstraße 8, D-85748 Garching, Munich, Germany

[‡] Rosettastein Consulting (UG), D-85354 Freising, Germany

Thank you

Interested in the R scripts?

contact@amaziz.com