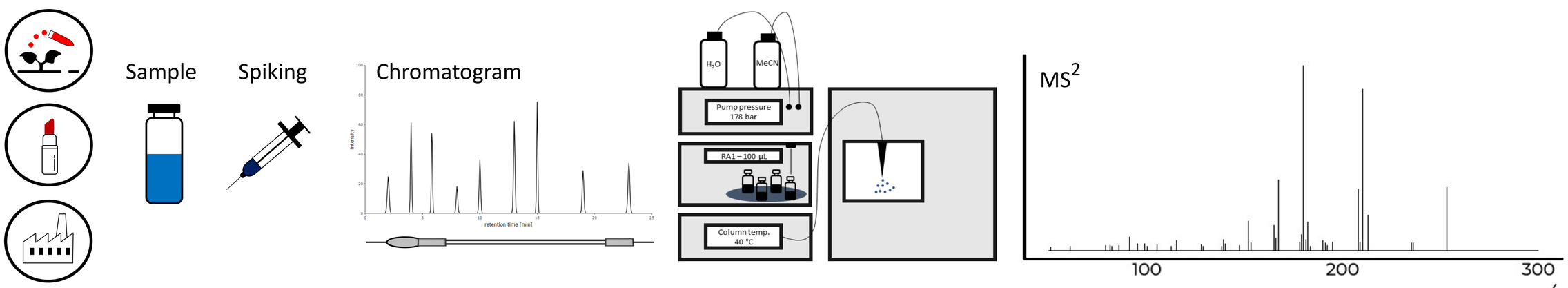
Reassessing Candidate Structures for NTS LC/HRMS^2 Features by Predicting Retention Times with Unsupervised Learning

Gordian Sandberg[a], Helen Sepman[a,b], Isak Samsten[c] and Anneli Kruve[a,b]

OBJECTIVE

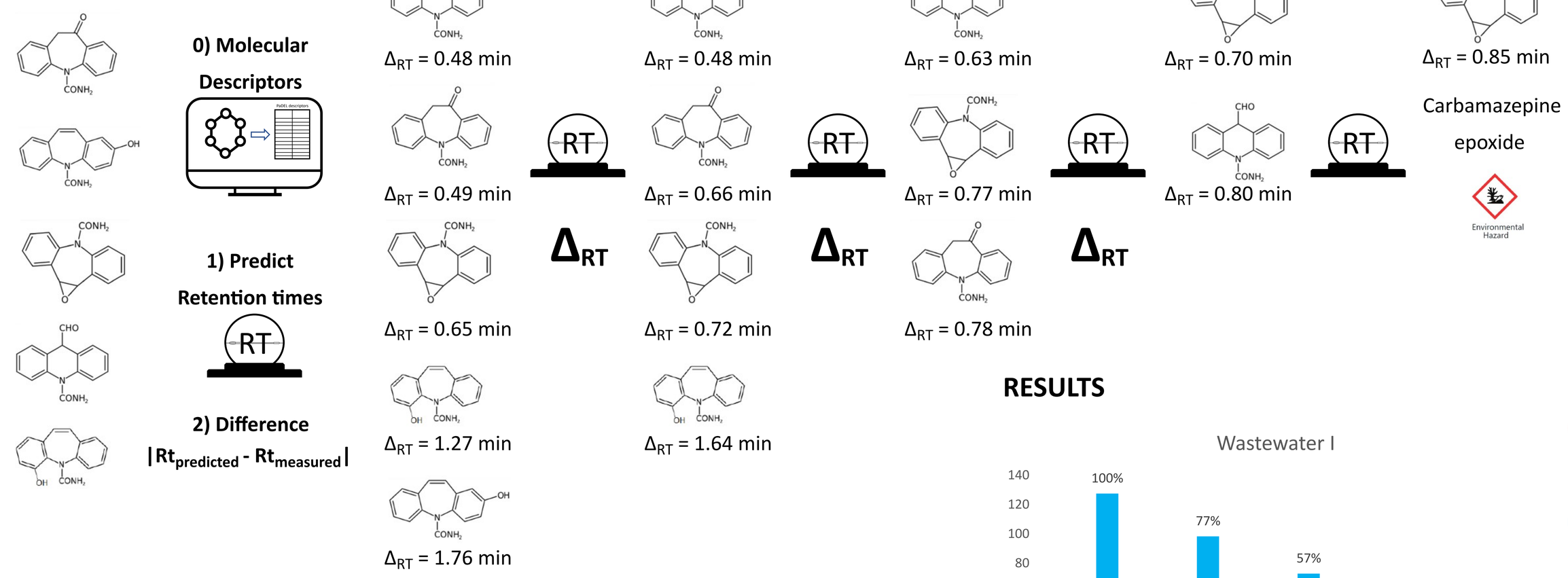
Detection of chemicals in complex mixtures with **non-target screening RPLC/ESI/HRMS**. Candidate structures are calculated and ranked by an iterative retention-time-prediction/candidate-structure-removal workflow **increasing the portion of correct candidate structures**.



WORKFLOW

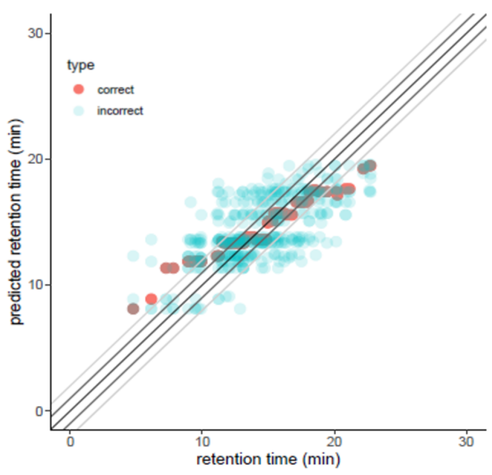
Molecular descriptors (PaDEL)[1], indicating the properties of each candidate structure (from MS^2 scorer **SIRIUS**[2]), are calculated. The candidates and respective PaDEL descriptors (input data) are used for training an extreme gradient boosting machine learning model (xgbTree), which **predicts the retention time (Rt)** of each candidate structure. The difference between predicted Rt of the candidate structure and measured Rt of the corresponding LC/HRMS feature is used to remove the candidate structure with the Δ_{RT} iteratively.

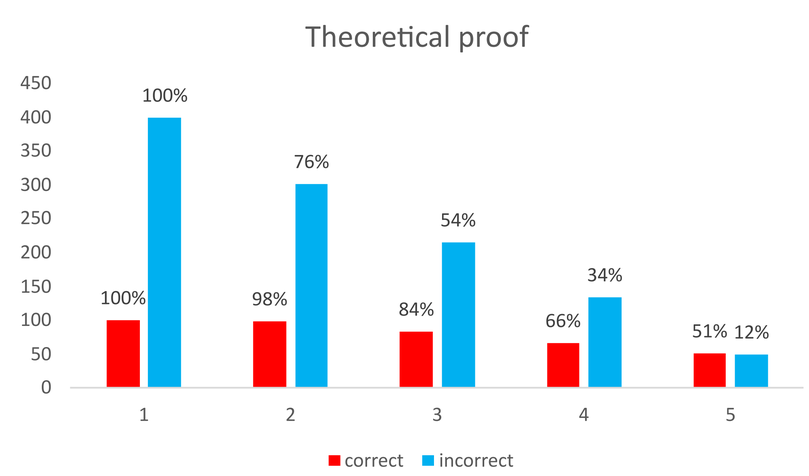
EXAMPLE



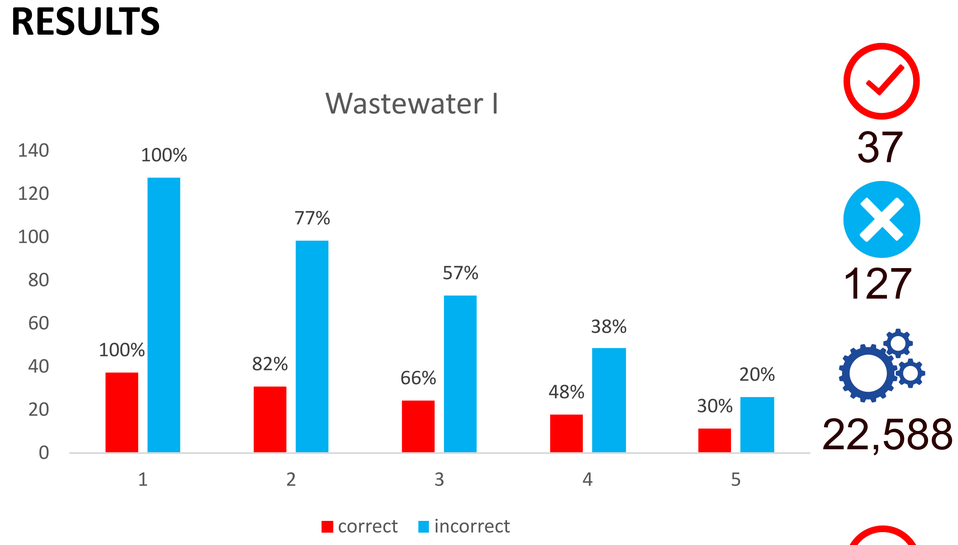
THEORETICAL PROOF

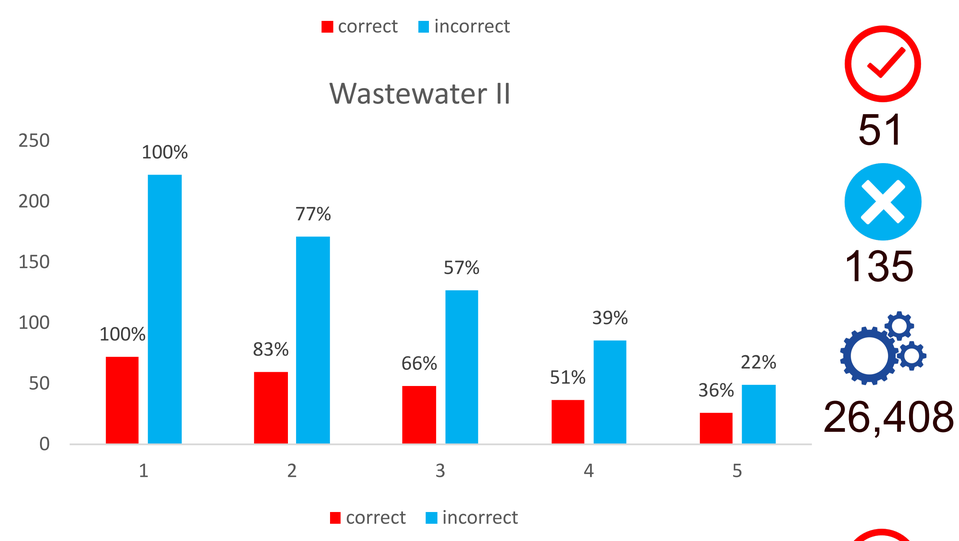
The correct candidate structures follow a relationship between input data (e.g. logP) and output data (retention time).

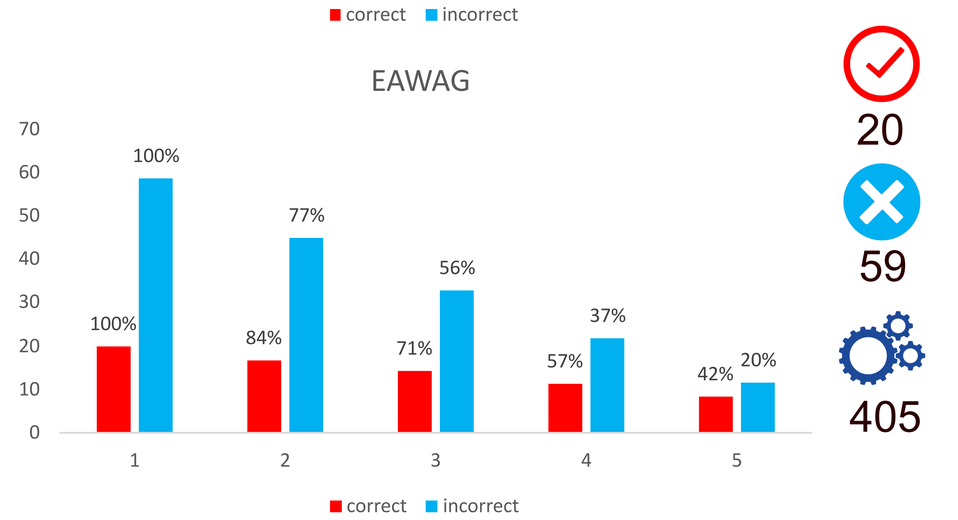




RESULTS







CONCLUSION

With this workflow it is possible to remove more incorrect than correct candidate structures. Promising theoretical results were generated and validated in two distinct studies of spiked chemicals in wastewater. MS^2 spectra from EAWAG on MassBank were used to test the workflow on chemicals, that were not contained in the SIRIUS training data. Although this training dataset was significantly smaller than the spiked wastewater datasets, removal efficiency was comparable or higher. This suggests, that the workflow is especially **suited for structures outside the SIRIUS training set**, increasing the percentage of correct candidates from 25% to 42% in the EAWAG dataset.

Funded by the Wallenberg Initiative Material Science for Sustainability

[1] Yap, C. W., *Comput. Chem.* **2011**, 32, 1466–1474.

[2] Dührkop, K. et al., *Nat. Methods* **2019**, 16 (4), 299–302.

Kruve Lab

[a] Department of Materials and Environmental Chemistry, [b] Department of Environmental Science, [c] Department of Computer and Systems Science, Stockholm Univeristy, Sweden