P18-91

A fully QAF-regulatory compliant model for Acute Oral Toxicity Assessment (AOrTA)

Marie Darracq--Ghitalla-Ciock[1], Emel Ay-Albrecht[1], Carole Charmeau-Genevois[1]

[1]KREATiS, 38080 L'Isle d'Abeau – France

Contact us : www.kreatis.eu / contact@kreatis.eu

Introduction

Replacing animal tests for acute oral toxicity (AOT) remains a major challenge, as no regulatory acceptable new approach methodologies (NAMs) have successfully replaced *in vivo* studies for this test. In this context, *in silico* approaches present a promising methodology, contributing to the development of alternative strategies that align with the 3Rs principles. **New *in silico* tools are able** to provide **reliable predictions while adhering to the** QSAR Assessment Framework[1] (QAF) for **regulatory use**. The Acute Oral Toxicity Assessment (AOrTA) model was developed to **predict Acute Toxicity Estimates** (ATE) for AOT according to the **CLP classification methodology for regulatory purposes**. The previous pilot version of AOrTA, presented last year, was **enriched** with substances for which validated data were considered sufficiently reliable, and the model was **enhanced** to provide a **tool suitable for regulatory toxicology**.

Methods

The original dataset was expanded by the addition of 1,091 substances extracted from ECHA dossiers and scientific literature, resulting in a final dataset of **2,516 substances** (Figure 1). Only organic mono-constituents were retained for model development.

A **binary Support Vector Machine** (SVM) based on optimised MACCS keys, was chosen to predict AOT according to CLP classification (classified *vs* non-classified). Modelling was performed in accordance with QAF[1]. Model **hyperparameters and features** were defined using a 5-fold cross-validation genetic algorithm and validated with the validation set. Model **parameters** were subsequently optimised as presented in Figure 2. This workflow was designed to **ensure model stability, generalisation to unseen substances and an extended applicability domain**[2] (AD).

To **support regulatory decision-making**, a **reliability domain** combined with a kNN approach was implemented to **assign the predicted CLP category**, providing toxicologists with additional guidance for result interpretation.

AOrTA was subsequently compared with the 33 "Non-toxic" models (after weight of evidence) developed for CATMoS, by using its publicly available external test set[3]. Of the 2,895 substances in the test set, 353 substances already present in the AOrTA dataset were excluded, ensuring an independent and direct comparison between the two models.

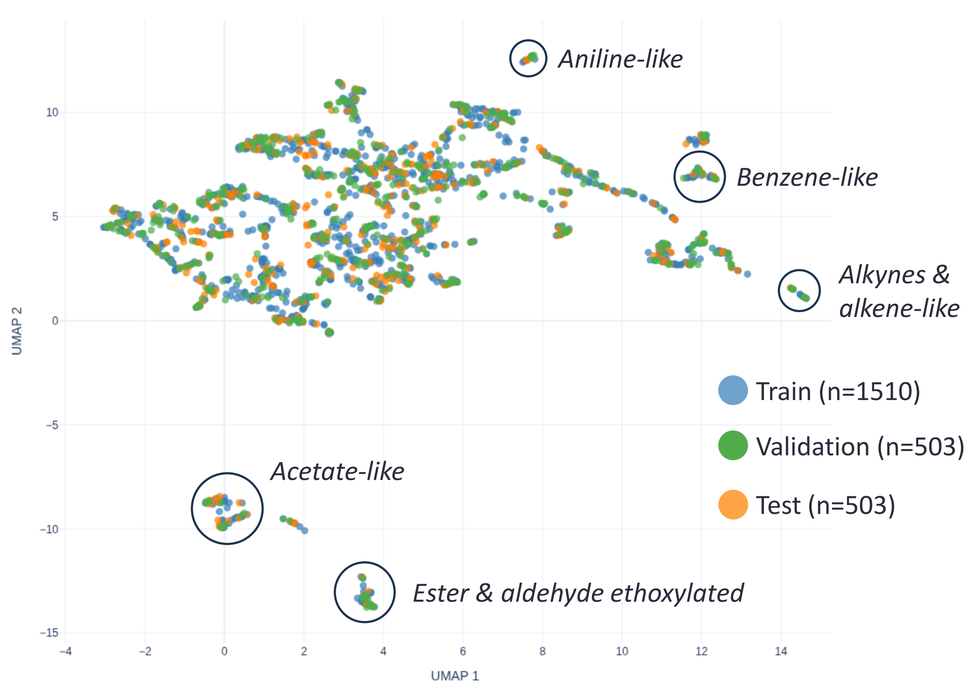

Figure 1. *UMAP enabling visualisation of the chemical space of AOrTA.*

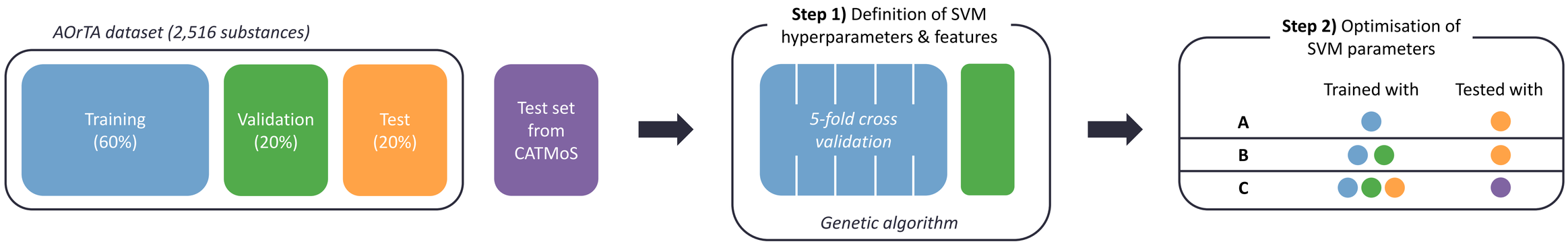

Figure 2. *Scheme of the modelling steps of AOrTA. The dataset was first split into training, validation and test sets. Step 1) The hyperparameters and the features of the SVM were defined using a 5-fold cross-validation genetic algorithm and validated with the validation set. Step 2) Finally, parameters of the SVM were optimised through different steps of training and testing (A, B and C).*

Results & discussion

The performance statistics across the training, validation and test sets indicate that AOrTA is a **robust and consistent model**, with comparable sensitivity and specificity across datasets (Table 1). The model showed a **strong ability to identify substances not classified for acute oral toxicity according to CLP**, whereas precision was lower for substances classified as toxic. This difference is influenced by the class imbalance in the dataset, comprising 1,600 non-classified substances versus 917 classified substances according to CLP (CLP 1 to 4).

In the direct comparison with the 33 models developed for CATMoS (Table 2), **AOrTA demonstrated higher performance for CLP classification**. This was achieved despite the substantially smaller development dataset available for AOrTA compared with CATMoS (2,516 *versus* 8,994 substances), suggesting that expert data curation, modelling strategy and AD definition contributed to its performance. Importantly, 524 of the 2,895 substances in CATMoS test set were outside AOrTA applicability domain. Thus, AOrTA's performance in classification of substances is due to its defined chemical space and comes at the cost of lower coverage. Overall, AOrTA provides performance comparable to the CATMoS models while **offering higher sensitivity for CLP-classified substances and an explicit reliability/applicability domain**.

Table 1. *Performance metrics for categorised substances and non-classified (NC) substances according to CLP. The left part of the table displays statistics prior application of the applicability domain, while the right part presents statistics after application of the applicability domain. *, Confidence interval (95%) are provided for step 1; (A) and (B) correspond to the different stages presented in Figure 2; CV, cross-validation.*

		Before Applicability Domain Defined					After Applicability Domain Defined				
		Sensitivity (classified)	Precision (classified)	Specificity (NC)	Precision (NC)	Balanced accuracy	Sensitivity (classified)	Precision (classified)	Specificity (NC)	Precision (NC)	Balanced accuracy
Step 1 (Cross validation)	Training set	88 ± 2%*	79 ± 2%*	87 ± 2%*	93 ± 1%*	-	-	-	-	-	-
	Validation set	71 ± 7%*	60 ± 4%*	74 ± 3%*	82 ± 4%*	-	-	-	-	-	-
Step 2 (A)	Training set	88%	78%	87%	93%	87.5%	-	-	-	-	-
	Test set	71%	67%	76%	79%	73.5%	72%	76%	88%	86%	80%
Step 2 (B)	Test set	72%	69%	77%	80%	74.5%	73%	79%	89%	85%	81%

Conclusion

AOrTA provides **high predictive capability for acute oral toxicity assessment**, supporting a more comprehensive evaluation while reducing the need for additional animal testing. The current version of AOrTA was intentionally developed using a simple descriptor-based approach, relying solely on structural MACCS keys, whereas many existing models also incorporate additional descriptors. Future development will therefore **investigate the integration of additional molecular descriptors** and **mechanistic information to further improve prediction reliability** and **support both regulatory assessment and R&D applications**.

Table 2. *Comparison of AOrTA v1.0 with 33 CATMoS models for the "Non-toxic" endpoint using the CATMoS external test set. Statistical results for CATMoS were retrieved from Mansouri et al. (2021). AD, applicability domain; (C) corresponds to the stage presented in Figure 2; NC, non-classified according to CLP.*

	Sensitivity (classified)	Precision (classified)	Specificity (NC)	Precision (NC)	Balanced accuracy
AOrTA v1.0 (C) (after AD)	75%	81%	82%	76%	78.5%
CATMoS 33 models for "Non-toxic" endpoint (after weight of evidence)	67%	*Not reported*	90%	*Not reported*	78%

[1] OECD (2024), (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions, Second Edition, OECD Series on Testing and Assessment, No. 405, OECD Publishing, Paris.
[2] Ay-Albrecht et al. (2026), A new high-accuracy QSAR model based on ultra-curated data for predicting thyroid receptor-related endocrine activity. Frontiers in Toxicology.
[3] Mansouri et al. (2021), CATMoS: collaborative acute toxicity modelling suite. Environ Health Perspect.

Scan QR code to download the poster
Available on https://kreatis.eu