IBMC
Institute of Biomedical Chemistry

Way2Drug
Understanding Chemical-Biological Interactions
PREDICTIVE SERVICES

XXXI Symposium on Bioinformatics and Computer-Aided Drug Discovery (BCADD-2025)

COMPUTATIONAL MODELING OF BIOACCUMULATION POTENTIAL OF PER- & POLY-FLUOROALKYL SUBSTANCES: MACHINE LEARNING BASED QUANTITATIVE READ-ACROSS STRUCTURE-PROPERTY RELATIONSHIP APPROACH

A. Chandra*, *A. Banerjee*, *K. Roy

Drug Theoretics and Cheminformatics Laboratory, Jadavpur University, 188 Raja S C Mullick Road, 700 032 Kolkata, India

যাদবপুর বিশ্ববিদ্যালয় Jadavpur University

DTC LAB

CDC

ATSDR

AGENCY FOR TOXIC SUBSTANCES AND DISEASE REGISTRY

PFAS Contamination has the potential to affect growth, learning, and behavior of infants and older children.

DEPARTMENT OF ENVIRONMENTAL CONSERVATION

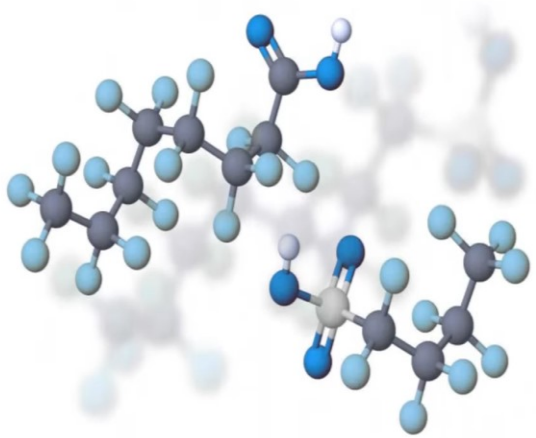
STATE OF ALASKA

European Environment Agency

Topics | Analysis and data | Countries

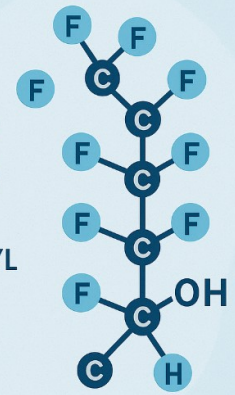
EPA United States Environmental Protection Agency | EPA 823R18004 | February 2019 | www.epa.gov/pfas

EPA's Per- and Polyfluoroalkyl Substances (PFAS) Action Plan



PFAS

PER- AND POLYFLUOROALKYL SUBSTANCES

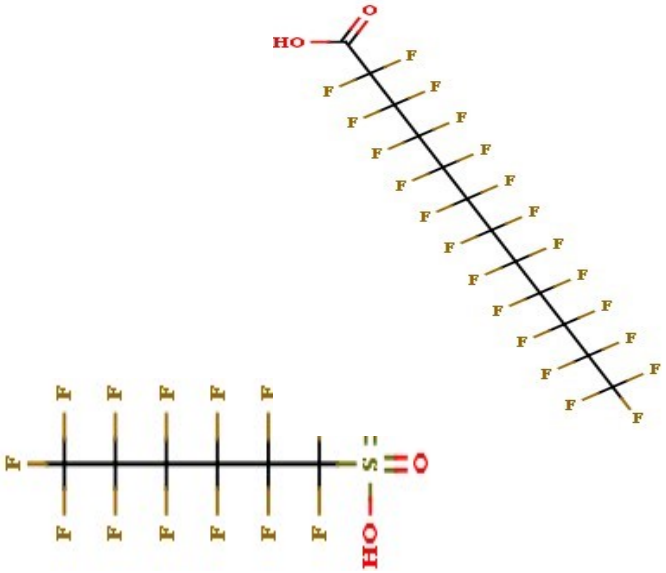


Registration, Evaluation, Authorisation and restriction of Chemicals

REACH

I) Introduction

- Per- and poly-fluoroalkyl substances (**PFASs**) are fluorine-substituted carbon chains that persist in the environment. They accumulate in aquatic life and pose risks to human health.



- The **q-RASPR** approach was implemented in this study.
- **Machine Learning (ML)**-based q-RASPR models were developed to predict the **bioconcentration factor (log BCF)** in fish.

- A Python-based tool **PFAS_(BCF)_Predictor v-1.0** was developed for data gap filling.
- The **bioaccumulation factor (log BAF)** of PFASs was predicted using the **Read-Across approach**.

Kowalska, D.; Sosnowska, A.; Zdybel, S.; Stepnik, M.; Puzyn, T. *Chemosphere 364*, **2024**, 143146.

II) Materials and methods

Data Collection and Data Curation

Descriptors Calculation and Data Pretreatment

Dataset Division and Feature Selection

QSPR Model Development

Optimization of Read Across Hyperparameters

Computation of RASPR Descriptors

Data Fusion and Feature Selection

MLR q-RASPR Model Development

Application of ML Algorithms

SVM, LSVM, RR, and RF

Prediction of Bioconcentration Factor (BCF)

BCF Prediction of True External Set

Development of Expert System for BCF Prediction

DTC LAB

PFAS (BCF) Predictor

PFASs → Important Features Selection → Read- Across → BAF Prediction

Collection of data

Freshwater fish (Teleostei taxonomic class)

Experimental bioconcentration factor (log BCF) data for 33 representatives

Experimental bioaccumulation factor (log BAF) data for 15 representatives

N training= 25
N test= 8

ML q-RASPR Model Development

BCF Prediction

PFAS

N training= 10
N test= 5

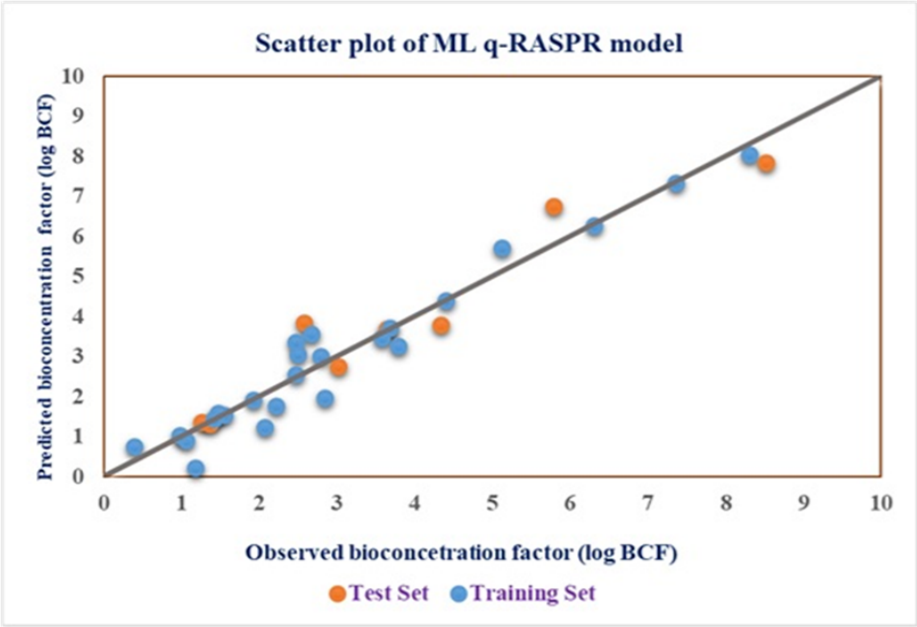
Read- Across

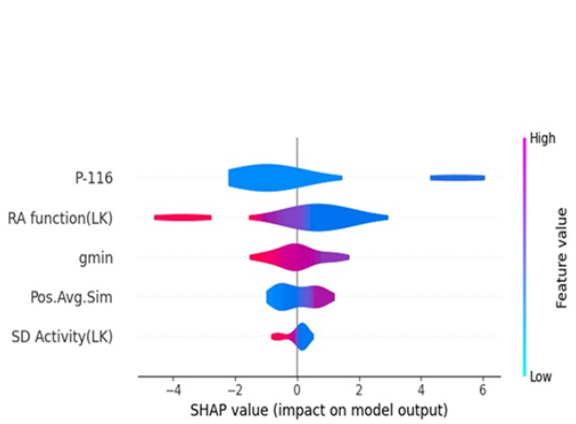
BAF Prediction

III) Results and Discussion

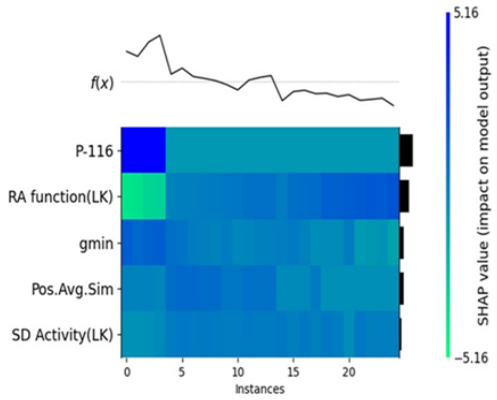
BCF prediction models for freshwater fish from the Teleostei taxonomic class.

Developed Models	N_{desc}	R^2_{train}	Q^2_{LOO}	Q^2_{F1}	Q^2_{F2}	MAE_{test}	$RMSE_{test}$
MLR QSPR	5	0.937	0.881	0.777	0.741	0.924	1.153
MLR q-RASPR	5	0.943	0.886	0.927	0.916	0.532	0.656
ML (LSVM) q-RASPR	**5**	**0.940**	**0.902**	**0.930**	**0.917**	**0.491**	**0.653**



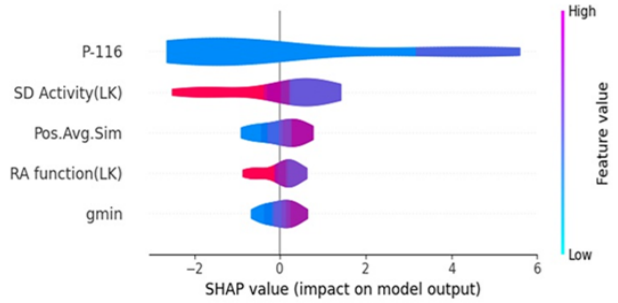


(i)

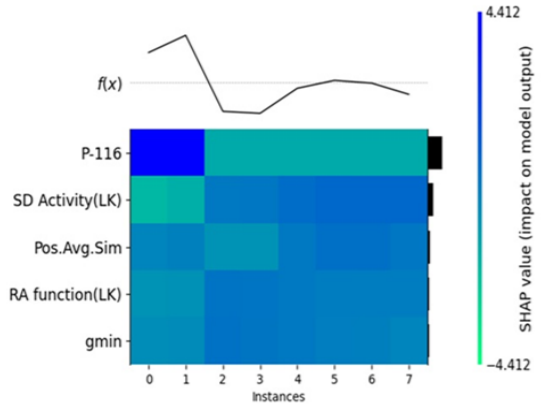


(ii)

Training Set



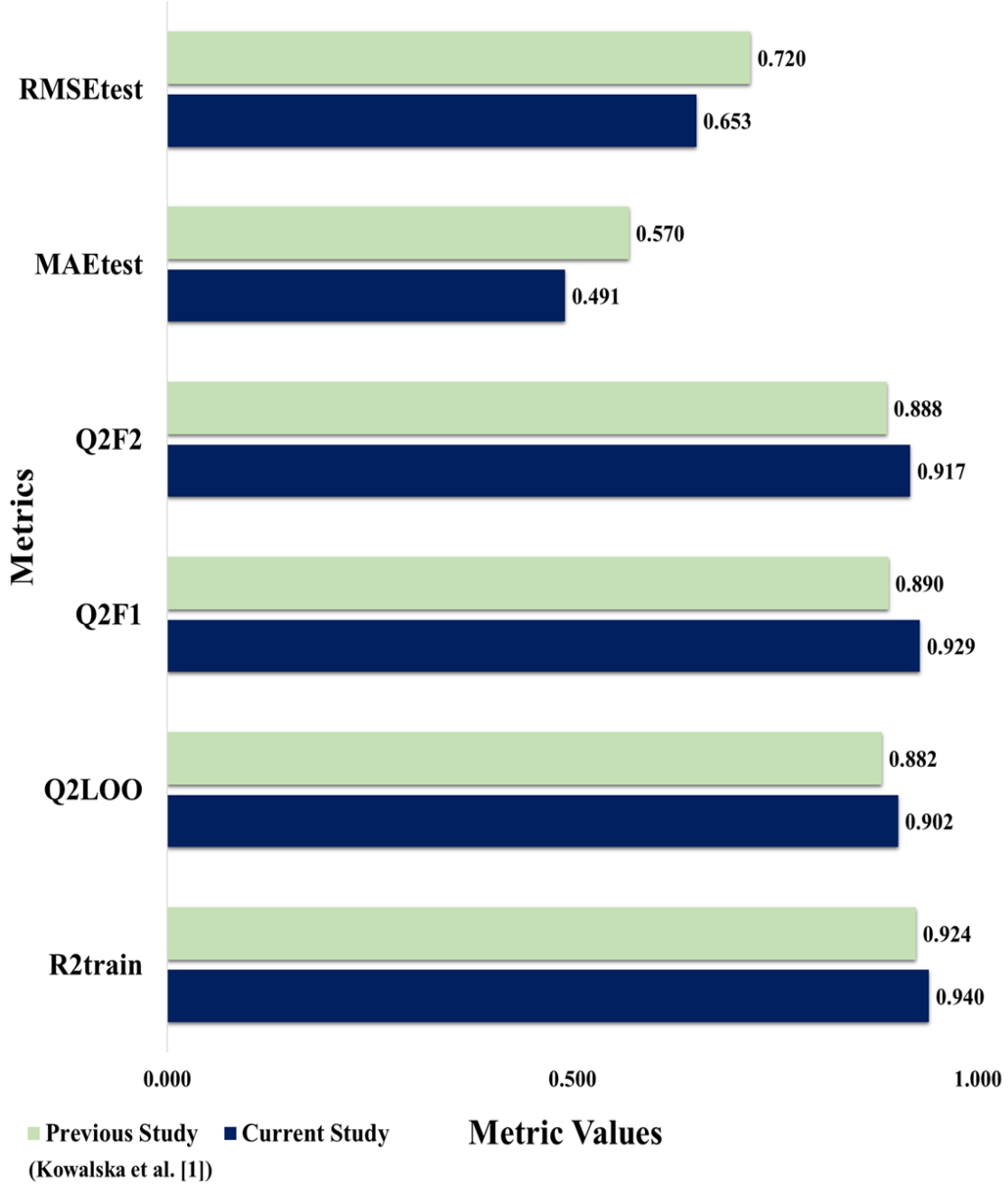
(iii)



(iv)

Test Set

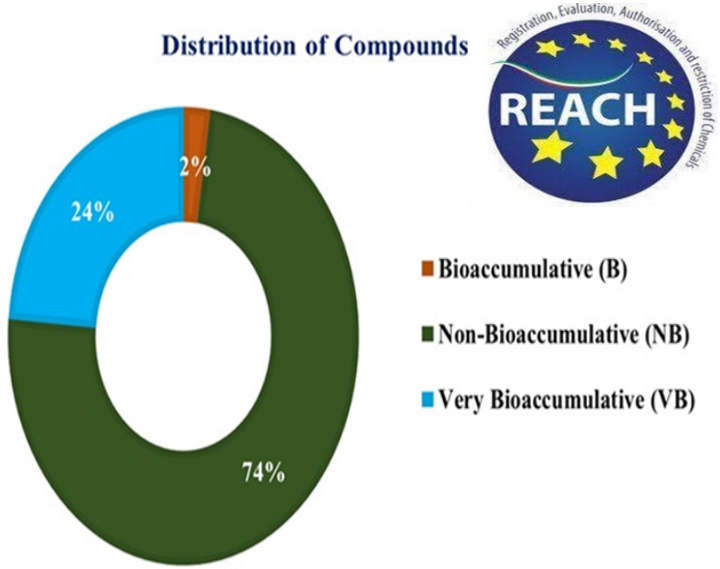
Comparison with the previously reported model for the bioconcentration factor (BCF) of PFASs

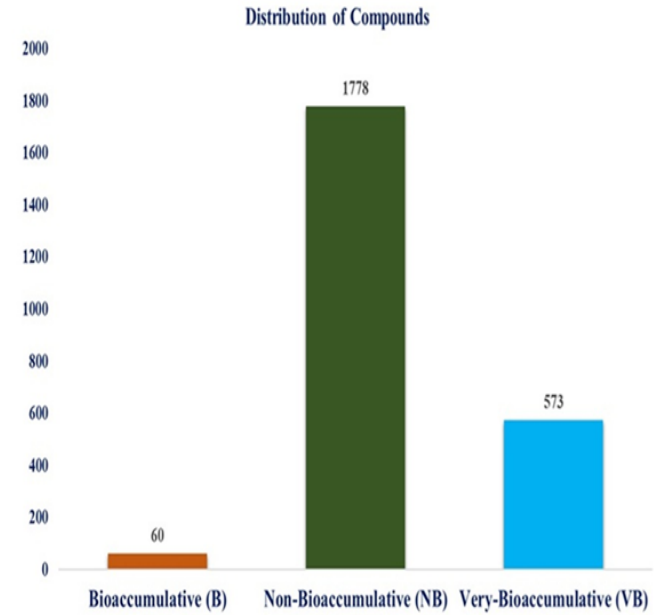


True External Set Predictions for log BCF

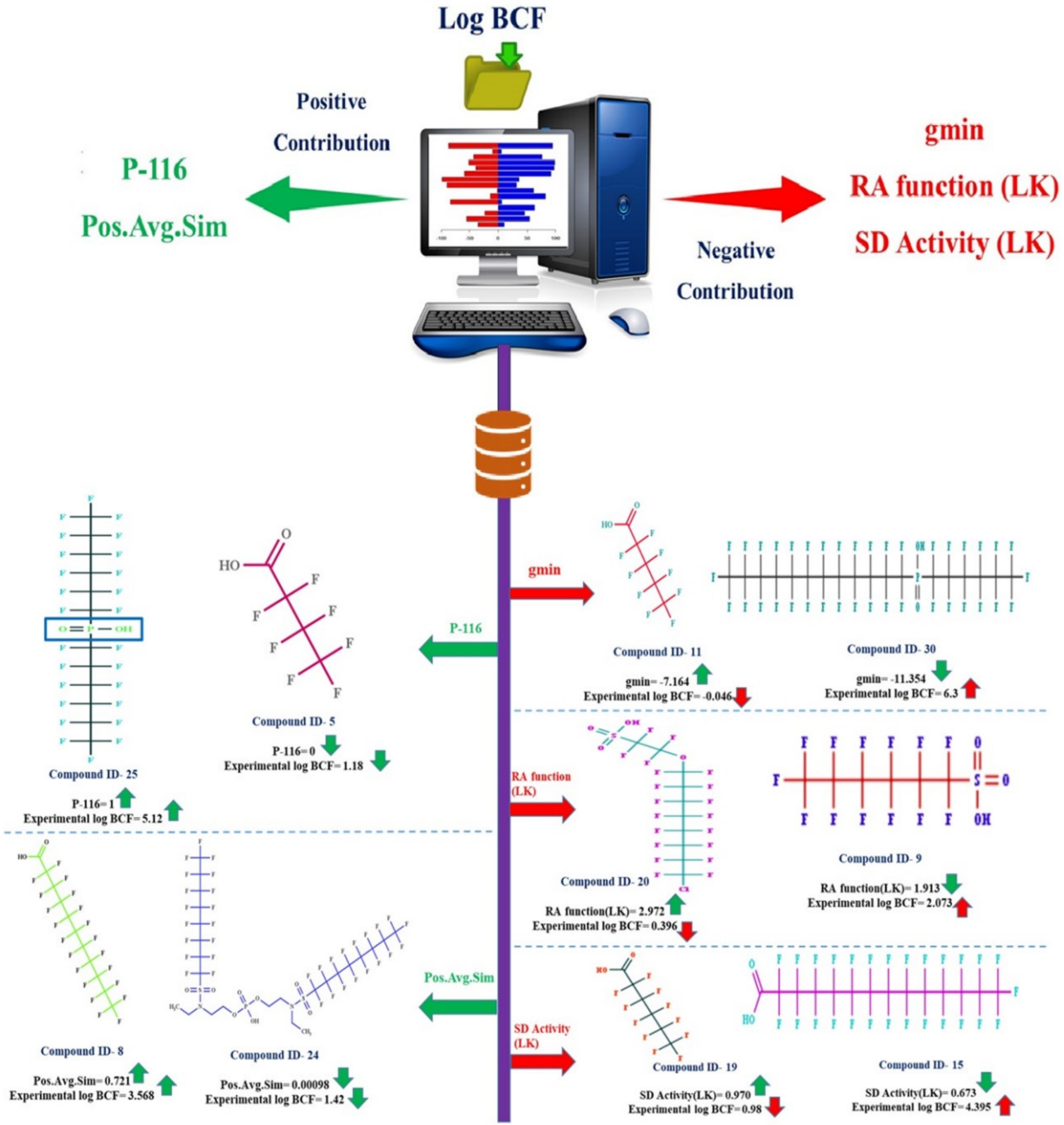
Category	Total Compounds	Reliability			Applicability Domain (AD) Status	
		Good	Moderate	Bad	IN	OUT
Current study	2411	2406	5	0	2409	2
Previous study (Kowalska et al. [1])	2522	-	-	-	2209	313

Distribution of PFASs based on the REACH guideline (Annex XIII).



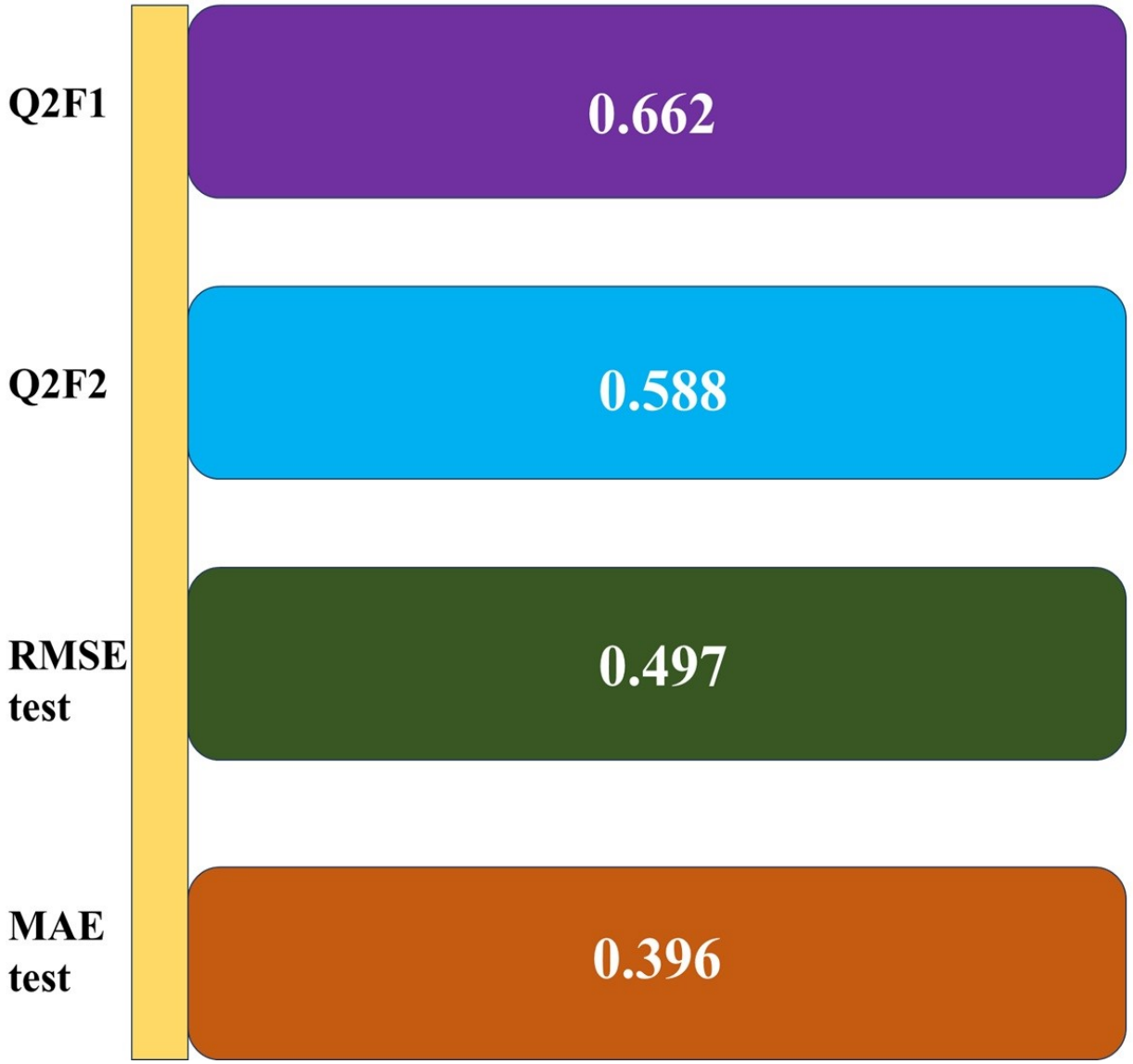


Mechanistic Interpretation (BCF Model)



PFAS_(BCF)_Predictor v-1.0

Prediction of Bioaccumulation Factor (BAF) using Read-Across



IV) Conclusions

PFAS are persistent and bioaccumulative, posing a continuous threat to the environment and human health due to their widespread distribution and endurance in the biosphere.

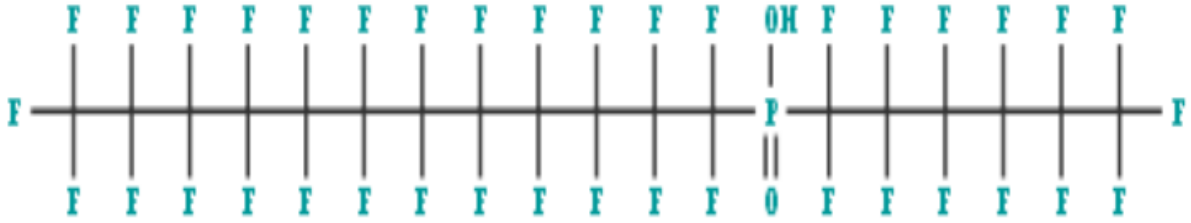
The study introduces the **ML q-RASPR** approach, which enhances prediction accuracy and interpretability for the bioconcentration factor (log BCF) estimation.

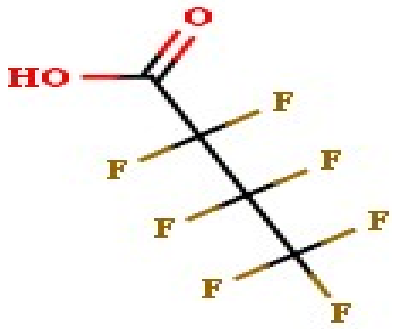
The developed **PFAS_(BCF)_Predictor-v1.0** tool predicts the BCF value of a true external set and classifies chemicals into the bioaccumulation categories as per **REACH guidelines (Annex XIII).**

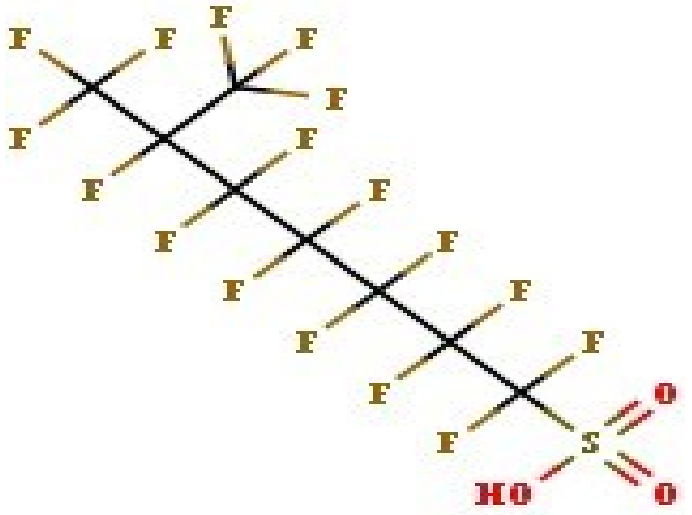
This approach substantially contributes to regulatory risk assessment by filling data gaps caused by missing experimental data for untested PFASs.

REFERENCES

1. Kowalska, D.; Sosnowska, A.; Zdybel, S.; Stepnik, M.; Puzyn, T. *Chemosphere 364*, **2024**, 143146.
2. Banerjee, A.; Roy, K. *Mol Divers. 26*, **2022**, 2847-2862.
3. **Chandra, A.; Banerjee, A.; Roy, K. *Sci. Tot. Environ*. 993, 2025, 180007**







Science of the Total Environment 993 (2025) 180007

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Prediction of bioconcentration factors (BCFs) and bioaccumulation factors (BAFs) for per- and polyfluoroalkyl substances (PFASs) using Read-Across and q-RASPR

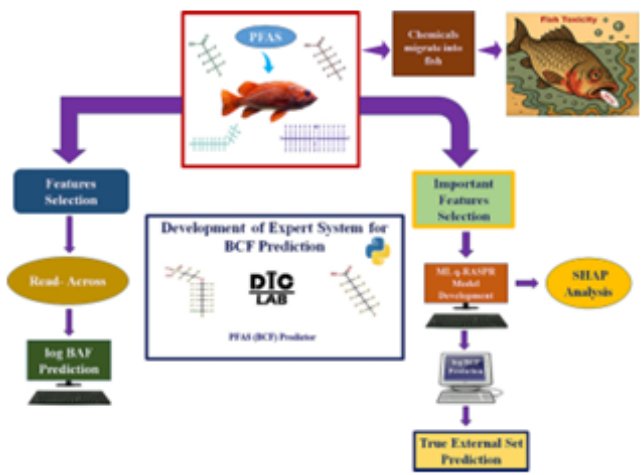
Akash Chandra [a], Arkaprava Banerjee [a], Kunal Roy [a,*]

[a] Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India

HIGHLIGHTS

- We developed a Machine learning (ML) based q-RASPR model for the BCF prediction of PFASs.
- Read-Across approach was used for the BAF prediction.
- A Python-based prediction tool was developed for accurate data gap filling.
- True external set predictions of BCF were made for 2411 untested PFASs.
- Untested PFASs were categorized into bioaccumulation classes defined by REACH.

GRAPHICAL ABSTRACT



IF- 8.0

IBMC
Institute of Biomedical Chemistry

Way2Drug
Understanding Chemical-Biological Interactions
PREDICTIVE SERVICES

XXXI Symposium on Bioinformatics and Computer-Aided Drug Discovery (BCADD-2025)

Thank you !