



Machine Learning-based Solutions for PFAS Challenges in Water Systems

Nagababu Andraju

Yun Ji

Prakash Ranganathan

Introduction

Effects of PFAS

- Global Environmental Contamination.
- Long-term Persistence and Health Impact.
- Widespread Human Exposure.
- Contamination at Airforce bases in ND.
- No site-specific data available
- Government announced \$19M

Role of Machine Learning (ML)

- Classification, Toxicity Prediction and Water Treatment.



Figure 1. PFAS Sources.

ML for Classification

- Identify toxic and persistent compounds.
- Chemical properties and toxicity.
- Supervised and unsupervised approaches.
- Features: *Molecular Weight, Number of Fluorine atoms, Carbon Chain Length, etc.*

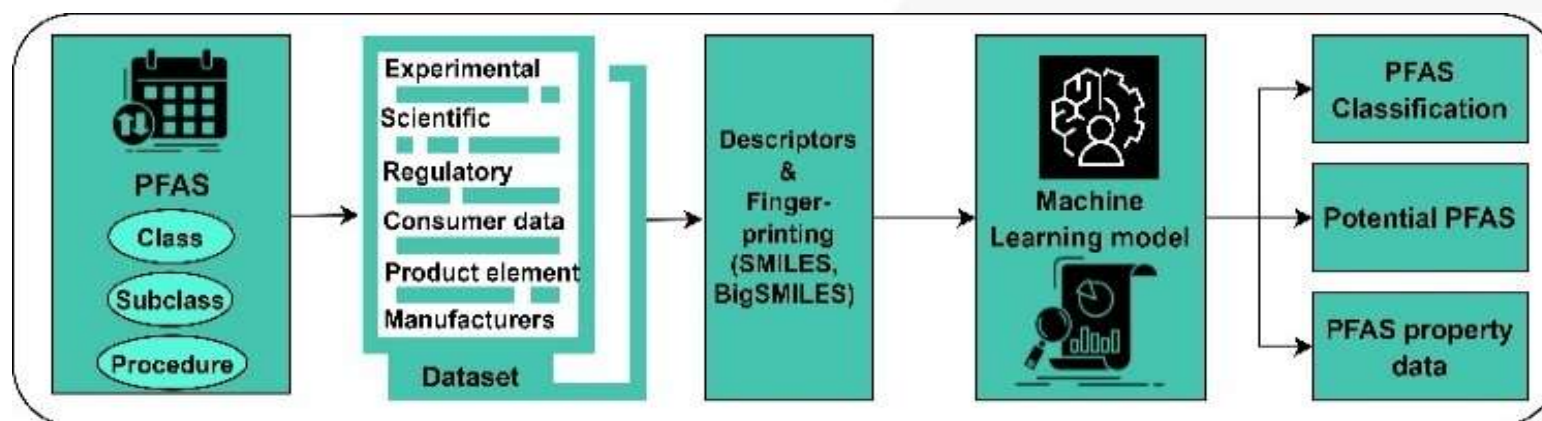


Figure 2. Common workflow for PFAS classification

ML for Toxicity Prediction

- Supervised approaches like SVM¹, RF², and NN³.
- Features: PFAS Structure and Toxicity values.
- Data Sources: ToxCast⁴, REACH⁵.

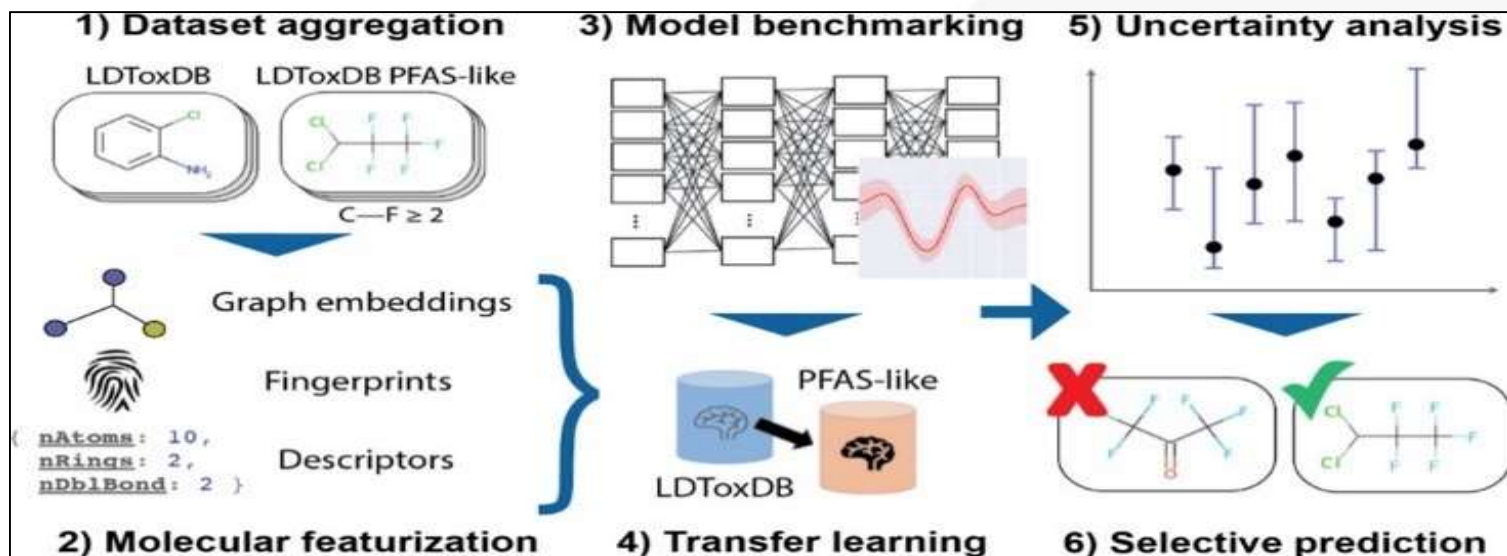


Figure 3. Common workflow for PFAS toxicity prediction

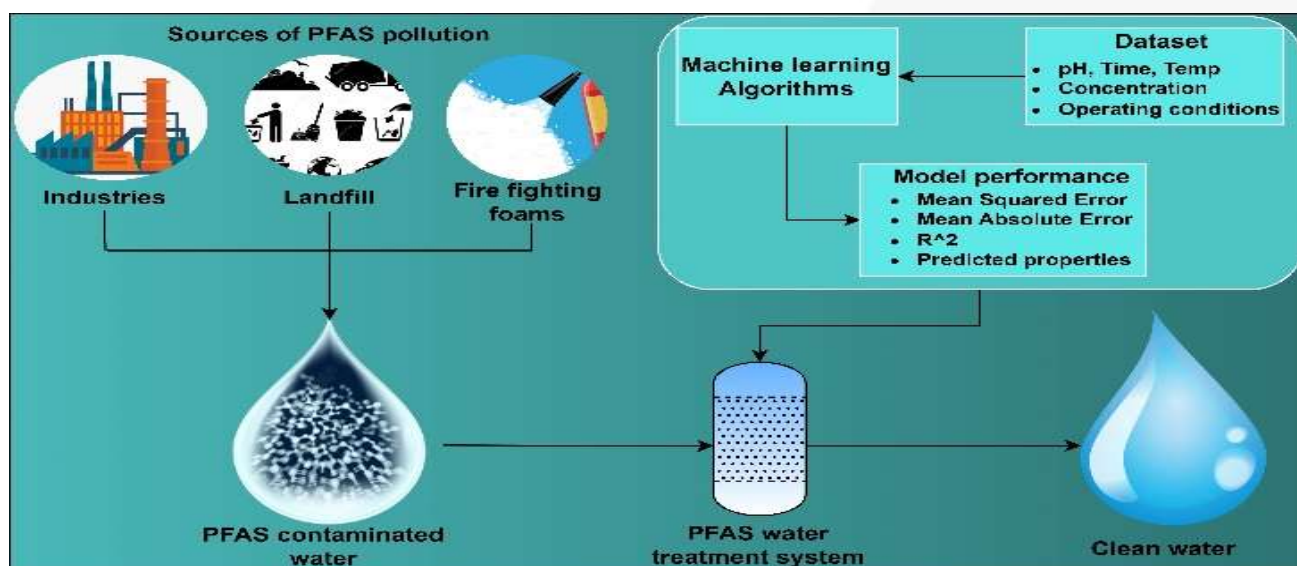
1 SVM: Support Vector Machines.
 2 RF: Random Forest.
 3 NN: Neural Networks.

4 ToxCast: U.S. Environmental Protection Agency's Toxicity Forecaster
 5 REACH: the European Chemicals Agency's Registration, Evaluation, Authorization, and Restriction of Chemicals



ML for PFAS Treatment

- Treatment processes Optimization and Efficiency Prediction.
- Experimental data have been used to train the model.
- Supervised approaches like SVM, XGBoost, RF, AdaBoost, NN, MLP¹, LR², and RR³.



- 1 MLP: Multi Layer Perceptron.
2 LR: Linear Regression.
3 RR: Ridge Regression.

Figure 4. ML workflow for PFAS removal from water



Conclusion

- ML is a powerful platform for exploiting PFAS data and providing insights for remediation solutions.
- Models XGBoost, RF, SVM, Bayesian network, etc. are proved.
- Improving ML models by data pruning and experimental validation.
- Integrating silo PFAS data sources into a single PFAS data platform with robust stakeholder participation.
- Addressing data quality challenges such as training data imbalances, interpretability issues, and secure data storage.



Thank you

Questions

nagababu.andraju@und.edu