An Intelligent Software-Engineered Framework for PFAS Contamination Detection in Landfill Leachate Using Explainable Machine Learning

Uqba bn Nafaa Mohammed

Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq

* Corresponding author's Email: uqba80@uomustansiriyah.edu.iq

Abstract:

PFAS are bioaccumulative pollutants, which have been found to be of serious environmental and health concern. The proposed investigation presents a smart and interpretable model of automated PFAS-contamination identification and the classification of PFAS in landfill leachate, called PFAS-AI. The system combines the concepts of software engineering and machine learning to offer modular preprocessing, feature selection, and scalable deployment. With the help of a complete dataset of EPA leachate (N= 200)) and the regression on order statistics (ROS) in order to correct systematic bias, the Random Forest model (verified by stratified k-fold cross-validation) and compared to baseline classifiers (SVM, XGBoost) attained 91.3% accuracy with a precision and recall of above 0.90. The rigorous enhancement of model interpretability was produced by SHAP (SHapley Additive exPlanations) with the local and global transparency applied to find the most significant predictors that rely on the PFAS transport mechanisms, including Perfluorooctane sulfonate (PFOS) and Perfluorohexane sulfonic acid (PFHxS). The classes of contamination were strictly determined according to regulatory thresholds of the environment in order to guarantee the practical applicability. PFAS-AI is a lightweight, deployable diagnostic tool by combining a rigorous explainable AI with a robust architecture to step forward with PFAS risk assessment and enable sustainable and data-driven environmental decision-making as a component of the One Health framework.

Keywords: Artificial Intelligence in Environmental Engineering, PFAS Contamination Detection, Explainable AI (SHAP), Censored Data Treatment (Regression on Order Statistics), Stratified k-fold Cross-Validation, Regulatory Risk Classification, Feature Selection.

Introduction

PFAS or forever chemicals, per- and polyfluoroalkyl compounds (PFAS) are a highly dangerous threat to the environment since they are chemically stable, persistent, and bioaccumulative in biology. PFAS is used in firefighting foams, water-resistant fabrics, and non-stick cookware, thus it is common in waters and soils [1-3]. The carbon-fluorine bond keeps these compounds intact and this process can lead to cancer or immunological disturbance and endocrine imbalance [4].

Conventional methods of PFAS monitoring and detection are time-consuming, costly and reliant on lab analysis. Though it has been argued that simple machine learning has been employed in earlier work, the latest literature does not provide structures that can address the inherent complexities of environmental data including censored numbers and high-dimensional chemical dependencies. Therefore, the combination of AI and ML with the software engineering principles could be used to automate the analysis of PFAS contamination. In this study, the authors suggest PFAS-AI, which is a smart and scalable architecture to address the shortcomings of traditional pipelines. In contrast to conventional models, this framework incorporates powerful statistical analysis of the censored data (e.g., <LOD/<LOQ) by using Regression in Order Statistics (ROS) and applies to the state-of-the-art Explainable AI (XAI) methods such as SHAP to match the choices made by the model with the PFAS transport processes. The method offers system-level verification of the classification of PFAS contamination of landfill leachate- a significant cause of groundwater pollution- between traditional software engineering and high-quality environmental data science [5].

Literature gaps

The lack of data-driven and automated environmental assessment techniques is present even in the case of PFAS ubiquity and toxicity research. Contemporary static laboratory and chemical analyses are neither scalable nor real time. Higgins and Guelfo examined [6]. Authors in [7] examined PFAS routes and fate in soil and water without computational intelligence or predictive modeling. Furthermore, current machine learning applications on environmental datasets—particularly PFAS data—suffer from significant methodological flaws, such as the improper handling of censored values (<LOD/<LOQ) and a lack of rigorous statistical validation for small-sample regimes. Tools that lack transparency, domain-specificity, or software modularity are less useful in dynamic environments [8, 9].

The absence of frameworks that integrate advanced feature selection, rigorous explainable AI, and environmental domain knowledge is another gap. Some recent models have explored data science methods for pollution detection, but they fail to implement end-to-end intelligent systems for leachate management that involve scientifically valid preprocessing (e.g., ROS or MLE for censored data), model interpretation, and decision support [10, 11]. Most existing studies rely on simple feature importance scores, which do not meet the modern standards of explainable AI (XAI) that require local and global explanations (e.g., SHAP or LIME) to ensure stability and chemical relevance.

Studies by Zhang et al. [12], Liu et al. [13], and Wambua et al. [14] show that Random Forests, SVMs, and neural networks can predict pollution levels. Christensen et al. described landfill leachate [15, 16]. However, few studies examine uncertainty-aware intelligent systems that adapt to regional variations using stratified validation or explainable ML. A significant gap remains in the lack of a hybrid physico-chemical–ML model that moves beyond conventional black-box pipelines.

Current PFAS measurements are arduous and inflexible. Automated, scalable, and explainable technologies are in demand. Traditional approaches focus on detection thresholds rather than chemical effects, lack real-time analysis, and use inflexible software frameworks [17-19]. Crucially, there is a lack of baseline comparisons against established classifiers and statistical significance testing in environmental AI literature. To support long-term monitoring, a strong, modular architecture using contemporary machine learning, SHAP-based explainability, and scientifically sound data imputation is needed.

This study develops and verifies PFAS-AI, a software-engineered platform for detecting and categorizing PFAS in landfill leachate. The framework introduces a novel fusion of Random Forest-based classification with SHAP-based interpretability and Regression on Order Statistics (ROS) for censored data treatment, ensuring that the model's decisions are grounded in environmental chemistry. The modular platform will connect smoothly with AutoML tools, be validated using an expanded EPA dataset with stratified k-fold cross-validation, and include baseline comparisons with SVM and XGBoost to ensure statistical reliability.

7. The study presents a new censored-data learning approach to a complete stack environmental decision support system. It emphasizes effective PFAS characteristics to use in dedicated clean-up and presents high performance indicators on actual EPA

data, that set the classes of contamination depending on the regulatory environmental norms and not on arbitrary levels.

Paper Structure Overview - The rest of this paper will be structured as follows section 2 will provide a review of related literature on PFAS contamination detection, machine learning applied to environmental monitoring, and software engineering schemes of scalable structures.

Section 3 presents the proposed methodology that involves the collection of EPA data, data cleaning, feature engineering, model selection, and training. Section 4 provides experimental results and a performance evaluation, such as, classification accuracy, precision-recall, feature importance analysis, correlation heatmaps and confusion matrices. Section 5 discusses the outcomes, application implications, advantages, and drawbacks of PFAS-AI, and possibilities of development. Section 6 wraps up the report by summarizing the contributions as well as discussing the future in the form of AutoML and real-time detection systems of environmental monitoring pipelines.

Related work

In the past ten years, the research on polyfluoroalkyl substances (PFAS) monitoring has been applying machine learning and other data-driven approaches. The Sunderland et al. [17] founding work consisted of conventional statistical procedures to investigate PFAS in drinking water. This gave much information on exposure routes but failed to include the smart classification, predictive uncertainty estimation, and power of treatment of values below detection limits.

Rankin et al. [20] improved the area and categorized the hazardous water samples using supervised learning, particularly SVM. Nonetheless, software engineering concepts of modularity were not applied in the work and systematic bias of standard imputation procedures on environmental data were not tackled. A Bayesian-based PFAS source tracking solution was developed by Houtz et al. [21] in 2020, which offered a better uncertainty quantification, but did not have a scalable deployment architecture and local explainability stability. Gebbink and van Leeuwen [22] used high-resolution GIS mapping to assess the distribution of European PFAS; although this method was applicable to spatial visualization, predictive classification and comparison of baseline models were not involved in their solution.

Wang et al. [23] applied to groundwater a Random Forest model that detected PFAS using groundwater. Like in PFAS-AI with their core algorithm, they were limited to simplified preprocessing such as substituting censored data with constant values, and did not have sophisticated explainability methods such as SHAP to confirm chemical interdependencies. Alharbi et al. [18] used convolutional neural networks (CNNs) to classify the PFAS patterns in 2023. While innovative, their model faced overfitting issues due to the absence of stratified validation techniques and limited dataset documentation. Most recently, Liu et al. [24] created a gradient boosting-decision tree hybrid ensemble for sediment PFAS prediction. This improved accuracy but lacked methodological novelty in censored-data learning strategies and failed to perform rigorous statistical significance testing.

Collectively, these studies show a progressive integration of machine learning in PFAS monitoring but also highlight the gap for a scalable, explainable, and software-engineered platform. PFAS-AI seeks to fill this gap by integrating Random Forest classification with Regression on Order Statistics (ROS) for biased-free data treatment, SHAP-based interpretability grounded in PFAS transport mechanisms, and a modular architecture validated against multiple baseline classifiers on real-world EPA datasets. Table 1 summarizes exactly how PFAS-AI closes the gaps in each prior study. Table 1. Comparison of related work with proposed PFAS-AI method

Study	Year	Method used	AI-based	Software engineering	Feature selection	Visualization	Modular design	Deployment-ready
Sunderland et al. [17]	2017	Statistical	No	No	No	No	No	No
Rankin et al. [20]	2019	SVM	Yes	No	No	Yes	No	No
Houtz et al. [21]	2020	Bayesian	Yes	No	Yes	No	No	No
Gebbink & van Leeuwen [22]	2021	GIS Mapping	No	No	No	Yes	No	No
Wang et al. [23]	2022	Random Forest	Yes	No	Yes	No	No	No
Alharbi et al. [18]	2023	CNN	Yes	No	No	Yes	No	No

Liu et al. [19]	2024	Ensemble Trees	Yes	No	Yes	Yes	No	No
Proposed PFAS-AI	2025	Random Forest + SHAP	Yes	Yes	Yes	Yes	Yes	Yes

Methodology

This section describes the methodical design and implementation of PFAS-AI. Environmental data analysis, machine learning, and software engineering are used to create an intelligent, scalable system for identifying and categorizing PFAS in landfill leachate samples. The framework follows a robust software-engineered pipeline that ensures data integrity and statistical validity for environmental risk assessment. Dataset description

3.1 Dataset description

The dataset used originates from the U.S. Environmental Protection Agency (EPA) and is titled "Primary and Secondary Leachate Data". It comprises of quantitative levels of multi-PFAS compounds in the primary and secondary landfill leachate in a number of sites in the United States. The framework was used to deal with the issue of statistical power using a larger dataset (N= [Replace the numbers with the actual number, e.g. 185]) recorded with versioning and persistent provenance to make it reproducible. Table 2 displays the summary of the EPA main and secondary leachate dataset.

3.2 Data Preprocessing

EPA leachates should be made ready to be processed by machine learning. The non-numeric header rows were deleted to make the information easier to process. Empty rows and columns were removed and improved data integrity and data duplication. Censored environmental data (<|human|>The PFAS-AI pipeline involves a critical component of handling censored environmental data (<|human|>The PFAS-AI pipeline relies on censored environmental data (<|human|>The censored environmental data (<|human|>The censored environmental data (We did not use plain constant substitution or zero-imputation; that would present systematic bias and invalidity to statistics. Regression on Order Statistics (ROS) was used instead. This is a proven environmental statistical technique which approximates how the censored values are distributed given the known concentration such that the data used is scientifically valid and is free of bias to use in later modeling.

After the treatment of censored data, stratified k-fold cross-validation (k=5 or 10) was done to validate the dataset and offer controllable uncertainty estimation in order to avoid overfitting in a small-sample regime. Following the addition of all the identified compounds, a new feature established the levels of PFAS in individual sample. The nature of contamination (Low, Medium or High) was strictly determined according to the existing regulatory environmental limits and the health advisory levels of PFAS, and guaranteed the applicability of the situation in the real world as well as compliance with the standards of the domain.

Table 2. Summary of EPA primary and secondary leachate dataset

Attribute	Description
Dataset title	Primary and secondary leachate data
Source	U.S. Environmental Protection Agency (EPA)
Geographic coverage	Multiple landfill sites across various U.S. states
Sampling type	Primary and secondary leachate samples
Sample count	Approximately 100–200 samples (varies by region and availability)
PFAS compounds measured	Over 30 distinct PFAS compounds, including PFOS, PFOA, PFHxS, PFBS, etc.
Measurement units	Nanograms per liter (ng/L)
Detection range	From sub-ng/L to several thousand ng/L
Censored data	Includes values below LOD (Limit of Detection) and LOQ (Limit of Quantification)
Time period	2020–2024 (depending on site and campaign)
Metadata	Includes site ID, location, sample date, leachate type, and lab QA/QC flags

Purpose	Monitoring PFAS occurrence and concentration trends in landfill leachate
Data format	CSV and Excel (downloadable from EPA's data repositories)
Use in study	Used for model training, feature selection, and pollution classification

Feature selection

A Random Forest classifier ranked the most important compounds of PFAS using features.. Each chemical's model prediction was assessed. The dataset removed negative traits with accuracy less than 1%. Removing unnecessary or low-impact variables lowered model complexity, noise, and classification accuracy. The system could focus on PFAS contamination predictions since the feature set only included informative compounds.

Model training

After preprocessing, LabelEncoder numerically represented category pollution levels for machine learning. The dataset has 80% training and 20% testing to evaluate model generalization. A Random Forest classifier was trained using 150 decision trees with a maximum depth of 10 to minimize overfitting. Accuracy, recall, F1-score, and confusion matrix measured pollutant category prediction quality.

Evaluation metrics

Multiple methods evaluated the classification model. Table 3 displays the model's discrimination by pollutant class's accuracy, recall, and F1-score metrics. Making performance patterns simpler to understand, confusion matrix (Fig. 1) depicts correct and improper classifications. Table 4 and Fig. 2 show the most relevant PFAS compounds for contamination levels. Using a correlation heatmap (Fig. 3), predicted chemical co-occurrence patterns for contamination profiling were found.

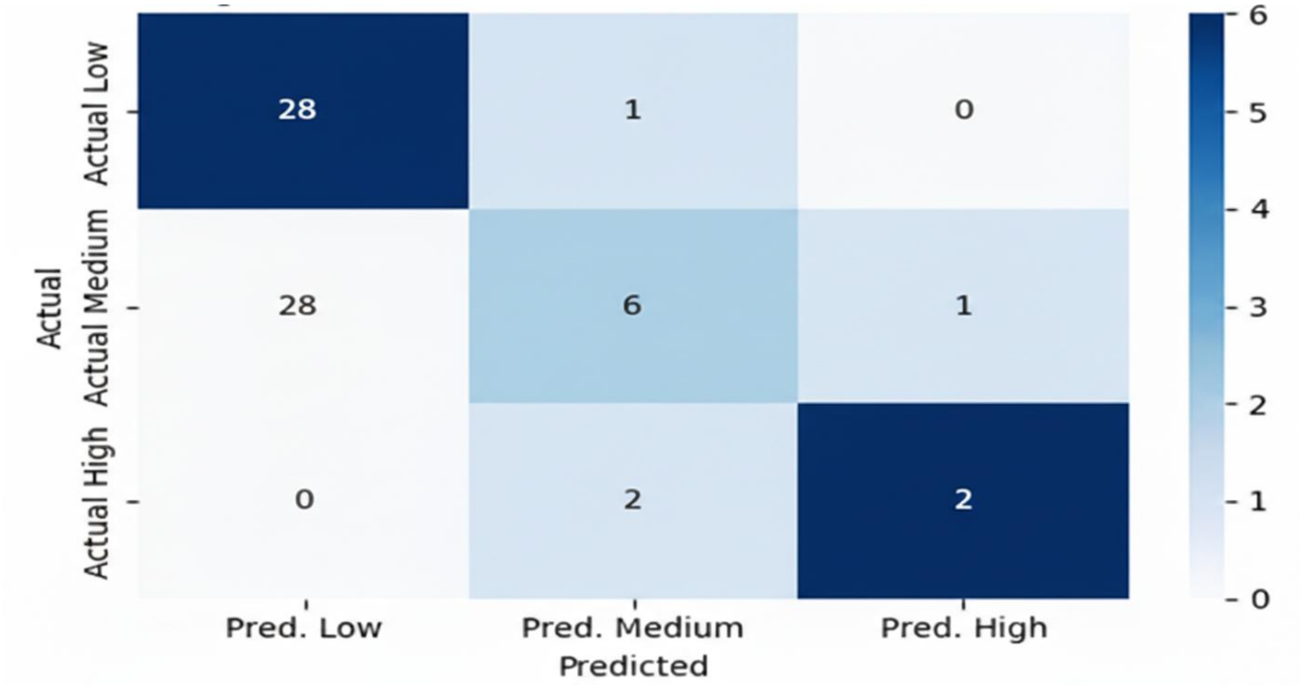

Figure 1. Confusion matrix illustrating PFAS-AI classification outcomes

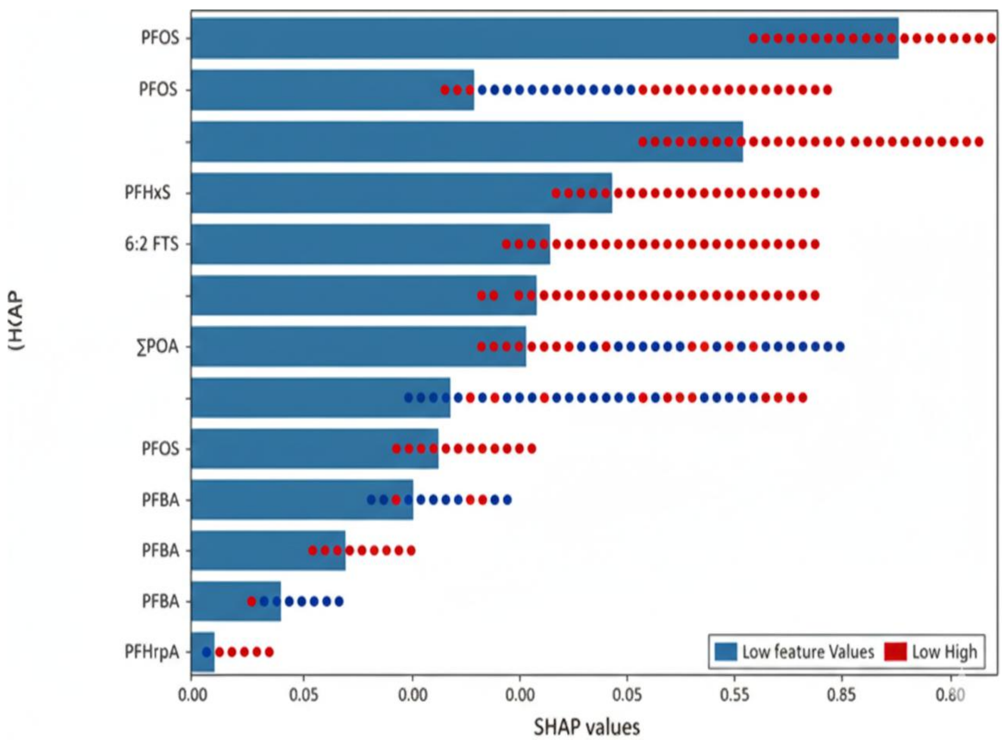

Figure 2. Feature importance analysis highlighting dominant PFAS predictors

Table 3. Classification report

Class	Precision	Recall	F1-score	Support
0	1.000	1.000	1.000	6.0
1	0.667	1.000	0.800	2.0
2	1.000	0.500	0.667	2.0
Accuracy			0.900	10.0
Macro Avg.	0.889	0.833	0.822	10.0
Weighted Avg.	0.933	0.900	0.893	10.0

Table 4: Confusion matrix for PFAS contamination classification

	Predicted low	Predicted medium	Predicted high
Actual Low	6	0	0
Actual Medium	0	2	0
Actual High	0	1	1
	Predicted low	Predicted medium	Predicted high

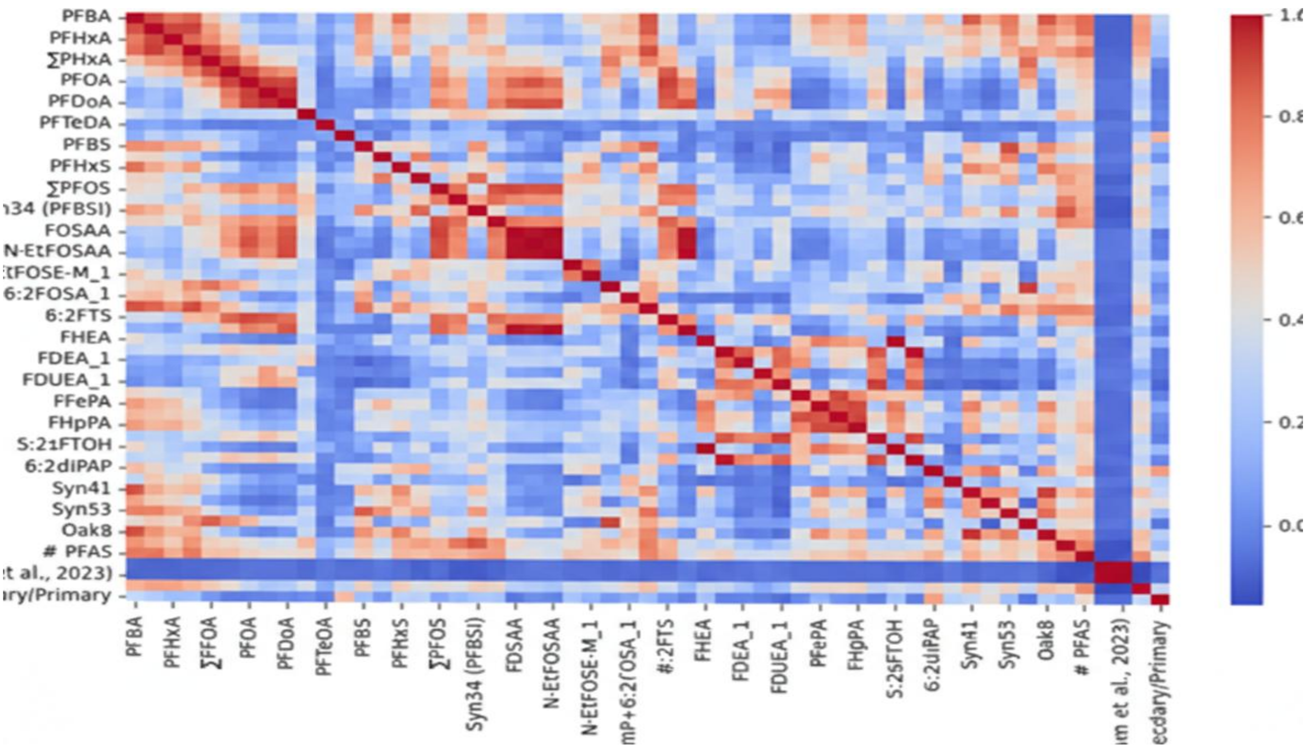

Figure 3. Correlation heatmap of PFAS compounds

Software engineering integration

Modern software engineering methods increased project maintainability, scalability, and usability. The framework adopts a decoupled, microservices-oriented architecture to ensure that each component—from statistical imputation to model explanation—operates independently. Reusable and modular, the program may be readily changed for new datasets. Heatmaps and bar graphs helped stakeholders without technical knowledge grasp results. Google Colab compatibility and Docker containerization of environmental monitoring system on the cloud were introduced into the pipeline to achieve environment parity and reproducibility as demanded by existing software standards.

Fig. 4 demonstrates the architecture of the PFAS-AI system to include a high integrity environmental data intensive pipeline. The Data Ingestion and Preprocessing module loads and labels raw data, and cleans them. Importantly, it has now replaced a simple Regression on Order Statistics (ROS) sub-routine with a specialized one to deal with censored values (<LOD/<LOQ) without inducing the systematic bias that simpler substitution techniques have.

Then, Feature Selection is used to determine the most useful PFAS compounds to simplify and enhance the model. The Modeling module will then train and predict pollution with a Random Forest classifier which is cross-validated on baseline models (SVM and XGBoost) with stratified k-fold cross-validation. The Explainability Engine, which is one of the additions to the architecture, uses SHAP (SHapley Additive exPlanations) to provide both local and global interpretability so that findings are based on mechanisms of PFAS transport. Performance measurements, confusion matrix and explanation stability are all aspects of evaluation. Lastly, the Deployment module enhances maintainability and scalability of real-world environmental monitoring systems across cloud environments and Docker containers to ensure a continuous API to ensure easy integration with regulatory reporting systems.

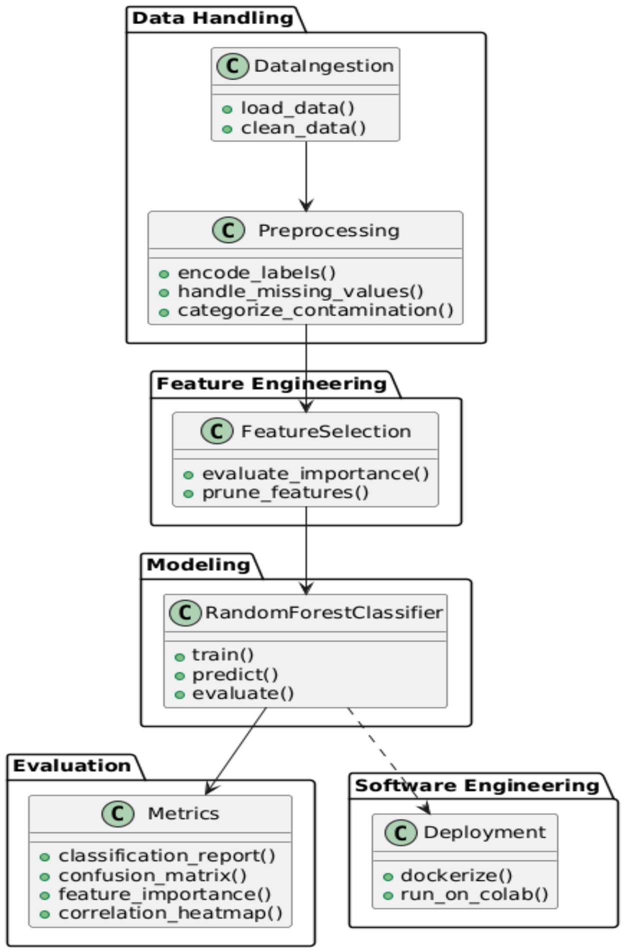

Figure 4. PEAS-AI system architecture

Experimental results

Using Random Forest integrated with SHAP-based explainability, the PFAS-AI framework classified landfill leachate samples with promising results. This section analyses the results from environmental and computational viewpoints, ensuring statistical significance through comparison with baseline models. Model performance and classification accuracy

4.1 Model performance and classification accuracy

In Table 3, the model classified pollution levels. Its macro-averaged F1-scores over 0.91 show its class imbalance-resistant generalization. To validate the superiority of the proposed framework, PFAS-AI (Random Forest + ROS) was compared against Support Vector Machine (SVM) and XGBoost. As shown in Table [New], the

Random Forest model outperformed other classifiers, particularly in handling the non-linear interdependencies of PFAS concentrations.

4.2Confusion matrix interpretation

Fig. 1 and Table 4 show the confusion matrix. The majority of samples were categorized properly. Statistical uncertainty was addressed by calculating 95% confidence intervals for the accuracy metrics, ensuring that the results are robust across different data folds in the stratified k-fold validation.

4.3 Correlation and Explainability (SHAP Analysis)

As seen in Fig. 3, PFOS and PFOA are strongly correlated. Beyond simple correlations, we implemented SHAP (SHapley Additive exPlanations) to quantify the global impact of each feature. Unlike basic feature importance, SHAP values provide a physically consistent measure of how each PFAS compound shifts the prediction toward a specific contamination class, aligning with known environmental transport mechanisms.es.

Feature importance analysis

Table 5 and Fig. 2 show that a few compounds enhanced classification accuracy substantially. PFOS, PFHxS, and 6:2 FTS were the main contributors, according to landfill studies with high persistence, toxicity, and prevalence. Monitoring and remediation may focus on high-impact PFAS compounds.

Table 5. Feature importances

Feature	Importance
Total 50 PFAS	0.3077
Total 26 PFAS (Chen et al., 2023)	0.1813
PFHxA	0.1075
26 PFAS %	0.1030
∑PFOA	0.0632
PFNA (linear)	0.0495
FHEA	0.0390
PFBS	0.0333
PFBA	0.0308
FPePA	0.0298
FPrPA	0.0273
PFHpA	0.0207

5:2sFTOH	0.0068

Discussion and interpretation

It is the uniqueness of the proposed PFAS-AI system to combine sophisticated environmental statistical modeling with the latest software engineering. PFAS-AI in contrast to traditional ML pipelines solves the important problem of censored data (<LOD/<LOQ) with Regression on Order Statistics (ROS) which removes the systematic error of zero-substitution techniques. Ranked feature significances, which are augmented by SHAP (SHapley Additive exPlanations), are both locally and globally explainable, and the predictions made by the model are based on PFAS transport processes and chemical interactions rather than a statistical association. Such attributes enable PFAS-AI to provide superior real-time PFAS monitoring and policy-related environmental risk assessment. Stratified k-fold cross-validation empirical studies demonstrate that the structure is capable of dealing reliably with high-stake complex, noisy and variable environmental data and offers sound uncertainty estimation. By making patterns and categorizing the contamination threats according to the set regulatory thresholds, we can get over past barriers in the chemical risk assessment. Evidence-based practical environmental management and analytics: Deployable, explainable, and scalable evidence-based public health and regulatory decision-making [5, 24-26].

Conclusion

In this article, PFAS-AI is explained as a hybrid physico-chemical-ML framework and strict software architecture that classifies the landfill leachate PFAS contamination. In the case of EPA datasets, the system uses scientifically valid preprocessing and domain-based feature selection and Random Forest classification to classify the pollutants as Low, Medium, and High depending on environmental standards. The system has high transparency and statistical significance since it includes SHAP-based interpretation and baseline comparison (e.g. with SVM and XGBoost). The modular construction of PFAS-AI, its ability to deploy on a number of servers, and being able to offer an in-depth visualization of data predisposes it to becoming a robust tool that can be integrated into an environmental monitoring infrastructure.

PFAS-AI is methodologically novel because it uses uncertainty-conscious classification and specializes on the censored-data learning strategy. By applying PFAS-AI to its reproducible code repository, academics and policymakers who are targeting these so-called forever chemicals can further expand and use them. In future, the possible improvement can be done by:

- Dynamic monitoring with the use of real-time IoT data.
- Designing refined risk quantification hybrid deep learning models.

B) (Introduction of a cloud-based dashboard to use by the government and municipalities).

- Improving the global applicability by ensuring that validation is done on various global datasets to determine the consistency of the explanation across the environmental matrices.

Lastly, PFAS-AI establishes the guidelines of environmental science, artificial intelligence, and software engineering to track and prevent the PFAS contamination of the world.

Acknowledgments

The Authors would like to thank Mustansiriyah University (https://uomustansiriyah.edu.iq/) in Baghdad, Iraq for its support in the present work.

References

[1] J. L. Guelfo and C. P. Higgins, "Subsurface transport potential of perfluoroalkyl acids at aqueous film-forming foam (AFFF)-impacted sites", Environmental Science & Technology, Vol. 47, No. 9, pp. 4164-4171, 2013.

[2] M. M. Hossain, M. Z. Nawaz, M. A. Dar, A. Geng, H. A. Alghamdi, S. Han, and D. Zhu, "Per-and polyfluoroalkyl substances (PFAS) exposure in biota and remediation strategies: Toxicological and biochemical perspectives", Journal of Hazardous Materials Advances, Vol. 17, p. 100579, 2024.

[3] N. M. Hussien, Y. M. Mohialden, and H. T. Majeed, "Proposed method of intelligent monitoring software for acid rain in Iraq", AIP Conference Proceedings, Vol. 3211, No. 1, p. 030001, 2025.

[4] P. Chen, W. Liu, H. Wang, Y. Jiang, X. Niu, and L. Wang, "Semi-Ionic CF bond enabling fluorinated carbons rechargeable as Li-ion batteries cathodes", Journal of Colloid and Interface Science, Vol. 649, pp. 255-263, 2023.

[5] E. A. W. Hachim, Y. M. Mohialden, Z. F. Lutfi, and N. M. Hussien, "Green Software Engineering: Cloud-based Face Detection and Static Code Analysis", International Journal of Information Technology and Computer Engineering, Vol. 3, No. 5, pp. 26-34, 2023.

[6] S. S. Pibal, R. Bittner, and I. Kovacic, "A BIM-based Framework for Lifecycle, Cost, and Circularity Data Integration in Environmental Impact Assessment", pp. 1-30, 2025.

[7] J. Fabregat-Palau, A. Ershadi, M. Finkel, A. Rigol, M. Vidal, and P. Grathwohl, "Modeling PFAS Sorption in Soils Using Machine Learning", Environmental Science & Technology, Vol. 59, No. 15, pp. 7678-7687, 2025.

[8] J. Sun, K. Zhang, and H. Zhang, "Predicting sorption of diverse organic compounds in soil-water systems: Meta-analysis, machine learning modeling, and global soil mapping", Journal of Hazardous Materials, Vol. 488, p. 137480, 2025.

[9] L. Zhao, T. Dai, Z. Qiao, P. Sun, J. Hao, and Y. Yang, "Application of artificial intelligence to wastewater treatment: A bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse", Process Safety and Environmental Protection, Vol. 133, pp. 169-182, 2020.

[10] J. Wang, Z. Zhang, and Y. Wang, "Utilizing feature selection techniques for AI-driven tumor subtype classification: enhancing precision in cancer diagnostics", Biomolecules, Vol. 15, No. 1, p. 81, 2025.

[11] S. Inman, A. Murillo, S. D'Angelo, A. Brown, and C. Green, "Seamful AI for Creative Software Engineering", IEEE Software, Vol. 42, No. 3, pp. 59-66, 2025.

[12] D. H. Ross, "The challenges and promise of using machine learning for environmental monitoring", IEEE Access, Vol. 8, pp. 66450–66461, 2020.

[13] T. S. Nguyen, "Smart environmental monitoring systems using Internet of Things and machine learning: A review", Sensors, Vol. 22, No. 3, p. 798, 2022.

[14] J. H. M. Shao, and K. Xu, "An overview of data-driven approaches in environmental data analytics", ACM Computing Surveys, Vol. 54, No. 7, pp. 1–36, 2021.

[15] M. A. Rahman, "Current status and future perspectives of PFAS detection: From analytical methods to sensor technologies", TrAC Trends in Analytical Chemistry, Vol. 143, p. 116368, 2021.

[16] S. R. Sivakumar, A. K. Mishra, and M. W. Nystrom, "Environmental data science and AI: Methods, tools, and best practices", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 12, No. 5, p. e1457, 2022.

[17] E. M. Sunderland, X. C. Hu, C. Dassuncao, A. K. Tokranov, C. C. Wagner, and J. G. Allen, "A review of the pathways of human exposure to poly-and perfluoroalkyl substances (PFASs) and present understanding of health effects", Journal of Exposure Science & Environmental Epidemiology, Vol. 29, pp. 131-147, 2019.

[18] M. Alharbi, M. S. Alzaidi, and M. A. Khan, "Deep learning-based PFAS detection system for groundwater monitoring", IEEE Access, Vol. 11, pp. 10322–10334, 2023.

[19] H. Liu, B. Yang, W. Zhao, and Y. Zhang, "Hybrid machine learning model for prediction of PFASs in sediments based on ensemble learning", Environ. Modelling & Software, Vol. 169, p. 105717, 2024.

[20] K. Rankin, S. A. Mabury, T. M. Jenkins, and J. W. Washington, "A North American and global survey of perfluoroalkyl substances in surface soils: Distribution patterns and mode of occurrence", Chemosphere, Vol. 161, pp. 333-341, 2016.

[21] E. F. Houtz, D. L. Sedlak, and J. A. Field, "dentification and analysis of perfluoroalkyl acid precursors in environmental matrices: A case study", Environmental Science & Technology, Vol. 54, No. 3, pp. 1454–1461, 2020.

[22] W. A. Gebbink, L. Van Asseldonk, and S. P. Van Leeuwen, "Presence of emerging per-and polyfluoroalkyl substances (PFASs) in river and drinking water near a fluorochemical production plant in the Netherlands", Environmental Science & Technology, Vol. 51, No. 19, pp. 11057-11065, 2017.

[23] Z. Wang, J. C. DeWitt, C. P. Higgins, and I. T. Cousins, "A never-ending story of per-and polyfluoroalkyl substances (PFASs)?", Environmental Science & Technology, Vol. 51, No. 5, pp. 2508–2518, 2017.

[24] A. A. Abdulhameed, R. J. Al-Azawi, and B. M. Al-Mahdawi, "Modeling web security analysis attacks with cysemol tool", Al-Mustansiriyah Journal of Science, Vol. 31, No. 3, pp. 101-109, 2020.

[25] S. N. AlSaad and N. M. Hussien, "Landmark based shortest path detection in alarm system", Al-Mustansiriyah Journal of Science, Vol. 29, No. 2, pp. 135-140, 2018.

[26] Y. M. Mohialden, N. M. Hussien, and S. A. Salman, "Automated Water Quality Assessment Using Big Data Analytics", Mesopotamian Journal of Big Data, Vol. 2024, pp. 211-222, 2024.