Environmental Research 215 (2022) 114286

Contents lists available at ScienceDirect

Environmental Research

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

XGBoost model as an efficient machine learning approach for PFAS removal: Effects of material characteristics and operation conditions

Elika Karbassiyazdi [a], Fatemeh Fattahi [b,1], Negin Yousefi [b,1], Amirhessam Tahmassebi [c], Arsia Afshar Taromi [d], Javad Zyaie Manzari [e], Amir H Gandomi [f], Ali Altaee [a], Amir Razmjou [g,h,*]

[a] Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, Australia
[b] Department of Chemical Engineering, Isfahan University of Technology, Isfahan, Iran
[c] Department of Scientific Computing, Florida State University, Tallahassee, FL, USA
[d] Petrochemicals Department, Iran Polymer and Petrochemical Institute, P.O. Box 14965/115, Tehran, Iran
[e] Department of Chemical Engineering, Iran University of Science and Technology, Tehran, Iran
[f] Faculty of Engineering & Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia
[g] School of Engineering, Edith Cowan University, Joondalup, Perth, WA, 6027, Australia
[h] UNESCO Centre for Membrane Science and Technology, School of Chemical Engineering, University of New South Wales, Sydney, NSW, 2052, Australia

ARTICLE INFO

Keywords:
PFAS Removal
Adsorbent
Machine learning
Wastewater treatment
Artificial intelligence

ABSTRACT

Due to the implications of poly- and perfluoroalkyl substances (PFAS) on the environment and public health, great attention has been recently made to finding innovative materials and methods for PFAS removal. In this work, PFAS is considered universal contamination which can be found in many wastewater streams. Conventional materials and processes used to remove and degrade PFAS do not have enough competence to address the issue particularly when it comes to eliminating short-chain PFAS. This is mainly due to the large number of complex parameters that are involved in both material and process designs. Here, we took the advantage of artificial intelligence to introduce a model (XGBoost) in which material and process factors are considered simultaneously. This research applies a machine learning approach using data collected from reported articles to predict the PFAS removal factors. The XGBoost modeling provided accurate adsorption capacity, equilibrium, and removal estimates with the ability to predict the adsorption mechanisms. The performance comparison of adsorbents and the role of AI in one dominant are studied and reviewed for the first time, even though many studies have been carried out to develop PFAS removal through various adsorption methods such as ion exchange, nanofiltration, and activated carbon (AC). The model showed that pH is the most effective parameter to predict PFAS removal. The proposed model in this work can be extended for other micropollutants and can be used as a basic framework for future adsorbent design and process optimization.

1. Introduction

Perfluoroalkyl and polyfluoroalkyl are water soluble due to their hydrophilic head. Strong carbon-fluorine covalent bonds are formed by replacing hydrogen atoms with fluorine atoms in alkyl chains. Perfluorooctane sulfonate (PFOS) and poly and perfluorinated alkyl substances (PFASs), known as long-chain PFAS, have been used more than other types of PFAS (Hassan et al., 2020). A significant amount of PFASs production is related to Perfluorooctane sulfonate (PFOS), Perfluorooctanoic acid (PFOA), and Perfluorohexane sulfonate (PFHxS) (Holmquist et al., 2020). PFASs, due to their high thermal, biological (Dolatabadi and Ahmadzadeh, 2019), and chemical stability, have a variety of applications in industries, including paper production, fire-fighting foams, waxes, and drink can-lining materials, wrapping materials, food, and non-stick cookware. Additionally, PFASs are used widely in the textile, carpet, and leather treatment industries (Bolan et al., 2021; Liu et al., 2020).

Global research demonstrates that PFAS disposal to the environment pollutes groundwater, aquatic environment, soil, and landfills (Liu et al., 2020). Generally leads to tragic consequences on plants, animals, and humans, for example showing toxic effects on the growth and reproduction of rodents and causing cancer in mice (Zhao et al., 2016).

PFASs are persistent and non-degradable chemicals, therefore,

* Corresponding author. School of Engineering, Edith Cowan University, Joondalup, Perth, WA, 6027, Australia.
E-mail addresses: gandomi@uts.edu.au (A.H. Gandomi), amirr@unsw.edu.au (A. Razmjou).
[1] Co-second authors.

https://doi.org/10.1016/j.envres.2022.114286
Received 5 June 2022; Received in revised form 19 August 2022; Accepted 3 September 2022
Available online 9 September 2022
0013-9351/

Acronyms

PFAS	perfluorinated alkyl substances
PFHxS	Perfluorohexane sulfonate
PFOS	Perfluorooctane sulfonate
PFOA	Perfluorooctanoic acid
GAC	Granular activated carbon
MI	machine learning
NF	nanofiltration
RO	reverse osmosis
AI	artificial intelligence
QSAR	quantitative structure-activity relationship
MNN	multitask neural network
AFFF	aqueous film-forming foam
IX	ion exchange
AC	activated carbon
RMSE	root-mean-squared-error
BAC	Bamboo-derived activated carbon
SHAP	Shapley additive explanations
AUC	area under the curve
PAC	powder activated carbon
C–F	Carbon-fluorine

remain in the environment for a long time. Furthermore, concerns about human and wildlife's health have increased in recent years, meaning that in addition to the inevitable effects on the human immune and thyroid system, there are pieces of evidence of serious health problems like the presence of harmful substances in the serum of living organisms (Bolan et al., 2021). Based on recent studies, people who are highly exposed to PFAS have a higher risk of death from Covid-19 than others in their community for various reasons, including weakened immune systems and lung tissue disease-related to PFASs (Catelan et al., 2021; Radfar et al., 2021).

Given the disposal of effluent and industrial wastewater to water sources, potable water is one of the most significant sources that are prone to contamination by PFASs. According to recent studies, ordinary water, and wastewater treatment methods, including sand ozonation, chlorination, and aeration, were not effective in removing PFAS. Other methods such as granular activated carbon (GAC) filtration, reverse osmosis (RO), nanofiltration (NF), advanced oxidation, and ion exchange have also been used to remove PFASs, which showed potential for PFAS removal. Many factors should be considered for choosing an appropriate water treatment method, such as long-term and high-performance efficiency, stability, and eco-friendly (Dolatabadi et al., 2021; Dolatabadi et al., 2021; Domingo and Nadal, 2019).

In recent years, (ML) and artificial intelligence (AI) in water and wastewater treatment have been employed to improve system performance (Li et al., 2021). AI can be defined as the machine's ability to reason, solve problems, understand, interpret, learn, communicate, and achieve goals as a skilled person in a particular field (Leo Kumar 2017). AI will play a significant role in technology and science and is used in water industries. ML, expert systems, fuzzy logic, robotics, and natural language processing are the six of the most vital areas in which AI is being used. The most widely used subsets in the water industry are ML, deep learning, and fuzzy logic (Leo Kumar 2017; Singh and Gupta, 2012). To assist plant operators, decision support systems adapted in plants for supporting operators utilize AI and non-AI based technologies to control and monitor operations. In other words, AI technologies are used in various process sections, including control, simulation, operation and maintenance, troubleshooting, and training (Uraikul et al., 2007).

AI has emerged as a beneficial implement for efficiently diagnosing and treating PFAS. A recent study has proposed an effective removal strategy for PFAS compounds and C–F bond evaluation based on an automated ML algorithm, which is more precise, user-friendly, and faster than manual processing. Reza et al. (Raza et al., 2019) exhibited precise prediction for dissociation energies of C–F bond through different ML algorithms such as least-absolute shrinkage, random forest, feed-forward neural networks, and selection operator regression with a significant saving in time and cost. They also asserted that training the data takes a few minutes, and less than 1 s is needed to predict the C − F bond dissociation energy (Raza et al., 2019). Since bioactivity is one of the potential hazards of PFAS, Cheng et al. (Cheng and Ng, 2019) have used ML-based quantitative structure-activity relationship (QSAR) models to predict \ the range of PFASs bioactivity. The author's trained new ML models, including conventional models like logistic regression and multitask neural network (MNN) and advanced graph-based models like weave models. Graph-based models and MNN achieved the best results. Kibbey et al. (Kibbey, Jabrzemski & O'Carroll, 2020) employed supervised ML to measure samples and recognize the source of PFAS components. Supervised ML algorithms use labeled training data to predict classes in the new test data. Of the four types of supervised MLs, one is the multilayer perceptron deep neural network. The rest are the common classifications of ML which are extremely-randomized trees, support vector machines, and K-neighbors. While these classifications work on completely different assumptions, they offer similar predictions, thus hypothesizing that patterns in the data can be used to identify the source of PFAS (Kibbey, Jabrzemski & O'Carroll, 2020). In subsequent studies, Kibby et al. (Kibbey, Jabrzemski & O'Carroll, 2021) examined data on an extensive set of samples from various sources to explore diverse pre-processing variables and the influence of feature selection on the classification of performances. In this set, 12 classification algorithms, including 11 conventional classification algorithms and a bunch of deep neural networks, were studied to allocate PFAS formulation in water samples, asserting the distinction between aqueous film-forming foam (AFFF) and non-AFFF compounds. Therefore, PFAS sources can be identified by combining supervised machine acquisition with PFAS sources (Kibbey, Jabrzemski & O'Carroll, 2021). Su et al. (Su & Rajan 2021) developed a database framework to discover systematics in structure-function relationships together with novel PFAS chemistries. The data framework maps are attributed to the SMILES standardization approach of encoding molecular structure with bioactivity and physicochemical property of PFAS. The PFAS map is an unsupervised tool to arrange new chemistries matching with the current properties of PFAS classification (Su & Rajan 2021).

ML usually involves automated computational procedures based on logical or binary operations and learns a task from a number of examples. Machine learning aims to produce simple, uncomplicated phrases categorized so that humans can easily understand them. They provide intuition into the decision-making process by imitating human reasoning. Like statistical approaches, the development of background knowledge is extracted, but it should be noted that it is assumed that humans are not involved in this operation (*Machine learning, neural and statistical classification* 1995).

In previous studies, engineers utilize AI for each process to improve their performance and safety, thus reducing maintenance costs and allowing the system to operate in more challenging conditions, however up to this point, no detailed studies have been conducted on material selection and discovery using AI (Merayo et al., 2019).

Although numerous types of research have been conducted to develop PFAS removal through various methods such as adsorption, ion exchange, nanofiltration, and AC, only a few prior researches have used ML methods for PFAS and environmental processes. Akber et al. applied machine learning to predict C–F bond dissociation energies and used theoretical data to predict them. They managed to classify and rationalize chemical trends in PFAS structures by their ML algorithm. Zhonge et al. have studied the importance of ML applications in data gathering in environmental science and engineering in future studies. (Raza et al., 2019; Sun, Fan & Bai). However, it is for the first time in this paper that

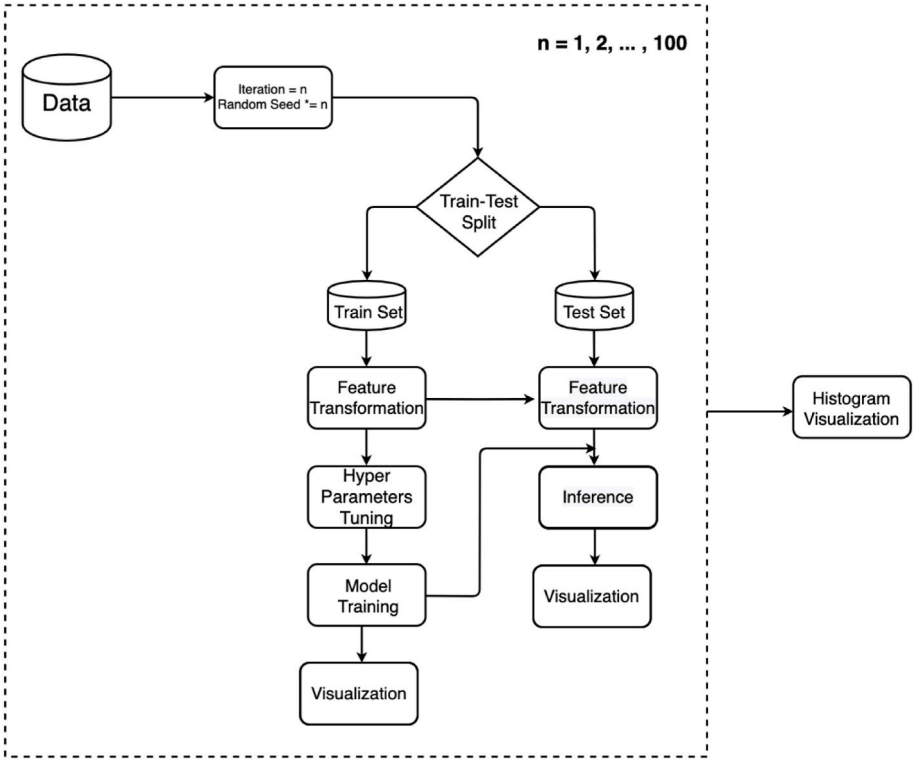

Fig. 1. Modeling and interpretation flowchart which used in this study.

the performance comparison of adsorbents, the function of AI, and different materials are studied and reviewed simultaneously; in other words, the combination of process and material contribution in PFAs removal are studied. In this work, we intend to study adsorbents' performances using ML models to improve material selection process for PFAS removal with the help of experimental data published in literature.

2. Modeling workflow

XGBoost (eXtreme Gradient Boosting), a popular type of shallow learning, was used in this study to predict PFAS removal by different adsorbents because of its advantages. Although deep learning models have demonstrated good performance in a wide range of variable situations, XGBoost is less prone to overfitting with a limited dataset and has lower processing requirements than deep learning (Jeong et al., 2021). ML techniques like XGBoost, which can manage missing information without the need for imputation pre-processing, are frequently utilized for classifying issues (Rusdah and Murfi, 2020). Compared to deep learning models, our study objectives and dataset do not require the extraction of larger data sets.

The XGBoost model, a scalable tree boosting method, is commonly utilized for data mining.

XGBoost is an enhanced gradient boosting algorithm that transforms many weak learners (decision trees) into strong learners. Fig. 1 demonstrates the workflow of the proposed framework. The figure shows that the data is being split into train/test sets at each iteration based on the given test size. The feature transformation step contains the feature encoding for categorical variables, feature standardization and scaling, and imputation of the null values. The feature transformer object is being trained only on the training set to ensure that the chance of data leak is zero. Then, the transformed data is passed to hyper-parameter tuning. Various methods include exhaustive grid-search, random search, and Bayesian optimization. In grid-search, all the possible combinations of the specified hyper-parameters will be checked. At the same time, the main idea of random search and a fixed number of the hyper-parameter grid can be sampled from the specified distributions. In principle, the learning slope will be positive. At the same time, similar outcomes would not be replicated, considering that employing exhaustive grid-search requires large enough computational time. Knowing specific regions for each hyper-parameter can help us ignore part of the search space. This solution is possible using Gaussian Process models for each new hyper-parameter combination.

The main assumption of Gaussian processes is that similar inputs give similar outputs, which can be used as an effective prior for the hyper-parameters tuning process. This weak prior is the basis of Bayesian optimization. Bayesian optimization is a constrained global optimization approach built upon Bayesian inference and Gaussian process models to find the maximum value of an unknown function in the most efficient ways (i.e. Fewer iterations) (Tahmassebi and Smith, 2021). The hyper-parameter tuning object is trained only on the training set, like features transformation. Therefore, all the components to train a model are ready. However, the main issue is to challenge the reliability of the trained model, considering the small sized database. Thus, a generalization pipeline is included in the proposed framework to validate the model's reliability by permuting the train/test data. The train/test split processcan be randomized by producing new random seed. For example, the process in the generalization workflow is repeated100 times. At each iteration, a new random seed is created (new random seed = random seed × iteration number) to split the data into train/test sets considering the test size as 10% of the data; to maximize the possibility of the coverage of the whole data in train/test sets. For each trained model, regression metrics such as coefficient of determination (R^2) can be calculated based on the testing set which comes from a unique random seed. For instance, for 100 iterations, 100 models are generated, and each model can be validated over its testing set using a regression metric. Therefore, an array of metric values of size 100 is produced, which can be used for statistical significance tests and confidence intervals (as shown in train/test histograms) for each target. In this way, we can consider statistics and, specifically, the central limit theorem to obtain the confidence intervals attached to each inference result and

Table 1
Input and output parameters for data collection.

Input/Output Parameters	Mean	Maximum value	Minimum value	Standard deviation
Particle size (nm)	332257.5	1000000	100	±454953.2
Initial concentration of adsorbent (mg. L^{-1})	3376.129	100000	10	±14966.88
Initial concentration of PFAS (mg. L^{-1})	550.5556	5000	10	±1022.28
Temperature (°C)	25.52991	70	20	±4.656716
pH	5.940427	10.3	2.3	±1.803938
Carbon chain	7.598291	11	2	±1.212892
Equiliuburm time (h) (output)	230.856	9312	0.016	±1012.786
Adsorption capacity (mg. g^{-1}) (output)	379.9393	3067	0.0014	±605.9524
Removal efficiency (%) (output)	66.18184	100	0.9	31.64933

reduce the possibility of stochastic effects. The pipeline design is model agnostic. In better words, any ML algorithm, including linear regression, support vector machines, decision trees, neural networks, or any ensemble method including Gradient boosting or random forest, can be employed as the primary training model. It depends on the use-case and data quality to choose the appropriate ML algorithm as the main model of the pipeline. In this paper, since we are dealing with short-skinny data (<500 observations and <15 features) and boosting algorithms perform better in this type of problem, the XGBoost algorithm was selected as the main algorithm. Boosting algorithms try to learn the observations that are harder to be learnt during multiple iterations (boosting rounds), considering the total number of observations is limited. Therefore, the chances are higher that the final model can predict the rare cases better (Chen et al., 2015; Tahmassebi et al., 2022). It should be noted that most of the steps in our proposed workflow are problem agnostic; therefore, it can be applied to any engineering problem (Tahmassebi et al., 2022). In principle, training an ML model for tubular data (like our data) mostly depends on how the data pre-processing and feature transformation steps were done. The more salient features are passed to the pipeline, the better results we can expect. This is also recently addressed in a great study by Scikit-Learn authors (Grinsztajn et al., 2022).

3. Data collection

A necessary step in ML is data gathering by analyzing which ML algorithm enables the best representation to the most complex problems. To build our model, 64 reports published from 2008 to 2021 were selected. In some of these articles, the authors studied more than one PFAS type; therefore, a total of 234 PFASs from those 64 papers were selected for training and testing our AI models. The collected data were used to analyze the adsorption capacity and removal percentage of different absorbents to find out the most appropriate adsorbent material.

To obtain the reference papers, keywords such as perfluoro octane sulfonate, polyfluoroalkyl sulfonate, removal, anion-exchange, and adsorption were used to search in Elsevier Scopus and Google Scholar.

After identifying the final works, various parameters were collected from each article, including adsorbent type, adsorbent concentration, average adsorbent particle size, PFAS type, PFAS initial concentration, pH, temperature, and equilibrium time, removal percentage. Percentage distribution and scatter plots were created for the extracted data that is collected in an Excel file before analysis and evaluation. The adsorbents used in the studies were classified based on their chemical component characteristics into 8 groups carbon family (materials within AC groups), resins, membrane, polymer (nanotubes and nanofibers with polymeric material base), clay, salt, MOF, and metal oxide.

In this study, missing removal percentage data (those not available in articles) were calculated from pure adsorption capacity data (Equations (1) and (2)).

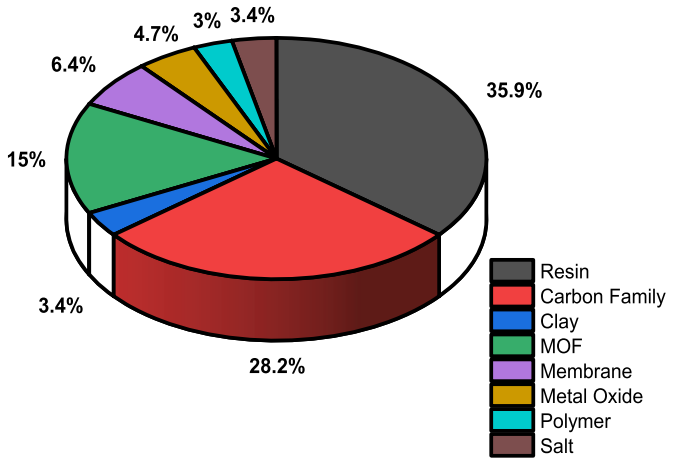

Fig. 2. Percentage of available data for different adsorbents such as clay, membrane, metal oxide, metal-organic framework, polymer, salt, resin, and carbon family.

$$\text{Removal efficiency } \% = \frac{(C_0 - C_e)}{C_0} \times 100 \quad (1)$$

$$\text{Adsorption capacity } q_e = \frac{(C_0 - C_e) \times V}{m} \quad (2)$$

C_0 and C_e are the initial concentration and equilibrium of adsorbent (mg. L^{-1}), respectively, V is the volume of the solution (L), and m is the adsorbent mass (g).

These data sets can be shared upon a request using a material-transfer agreement. Table 1 shows the minimum, maximum value, mean and standard deviation data according to the model parameters.

A pie chart was prepared to display the distribution of some parameters such as adsorbent types, particle size, PFAS initial concentration, and initial adsorbent concentration in the sorption process throughout the literature. Although the specific surface area is the key factor in the adsorption process, it was not included in the above parameters because of a lack of data. A few papers have stated surface area for adsorbent, while particle size was more available. Therefore, particle size has been chosen instead of surface area because these two parameters are inversely proportional; a smaller particle size results in more surface (Sun, Fan & Bai). As presented in Fig. 2, various types of ion exchange (IX) resins and carbon family data are available in experimental studies published in recent years compared to other adsorbents. Because of its efficacy, ease of operation, available exchange capacities, compact footprint, and regeneration characteristics, IX resin has received more attention than the other adsorbents (Liu et al., 2015; Zhang et al., 2021). Several studies have investigated the adsorption of PFOS and PFOA on ACs generated from natural or chemical sources via chemical or physical activation processes. Furthermore, AC was widely used to remove organic and inorganic contaminants from wastewater and drinking water due to its tunable textural properties and the possibility of decorating various chemical groups on its surface (25–27).

4. Results and discussions

Fig. 3a shows that AC and resin adsorption capacity distributions vary from 0.0014 to 3067 mg. $g^{-1,}$ while metal oxide-based adsorbents showed the lowest adsorption capacity of around 0.007 mg. g^{-1}. According to Figure (3), IRA67 has the highest adsorption capacity with 3067 mg/g at an initial concentration of 100 mg.L^{-1} PFOS at PH = 3 and 25 ▫C after 48 h compared to other resins. IRA67 is a polyacrylic gel-type resin with a tertiary amine group, and its particle size is 16–50 mesh (Du et al., 2015). Also, PFOS is described as a long-chained PFAS because of

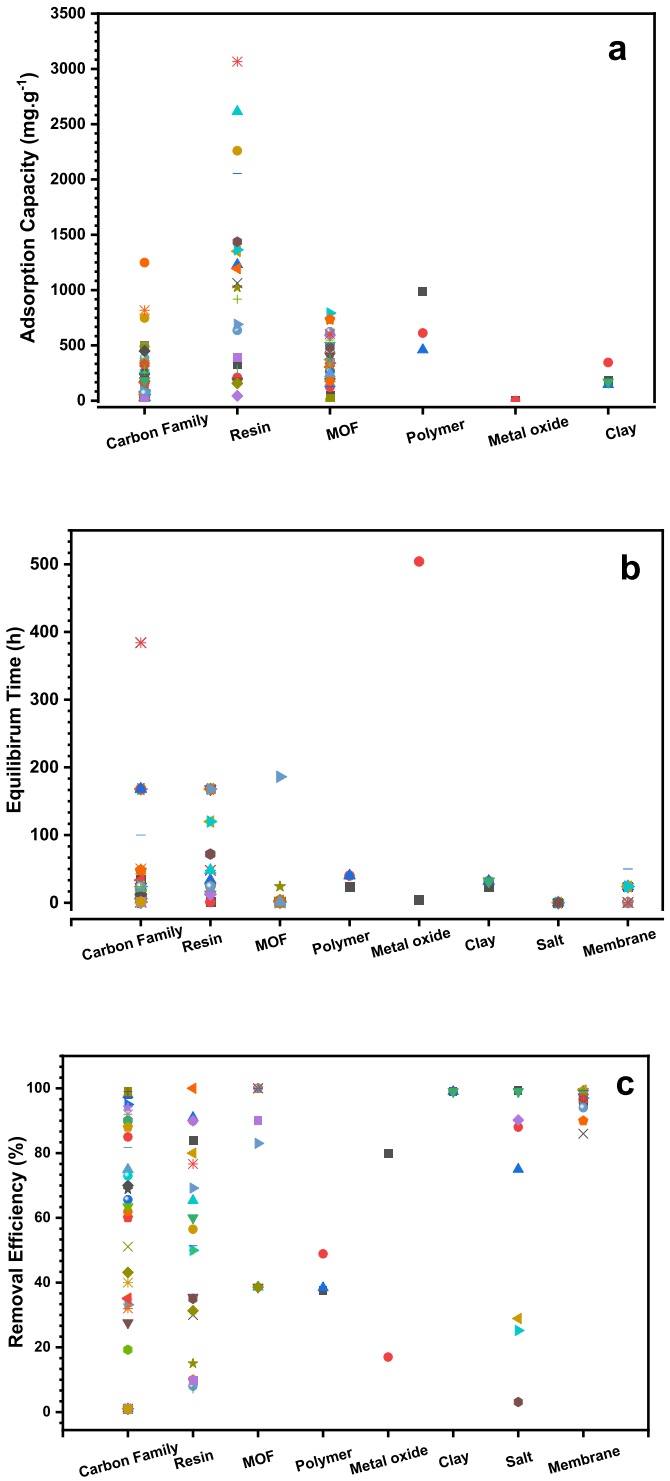

Fig. 3. (a) Adsorption capacity of different adsorbent, (b) Equilibrium time of different adsorption, and (c) Removal of PFAS on different adsorbent for short and long-chain PFAS the initial concentration of adsorbent from 10 mg. L^{-1} to 100000 mg. L^{-1}, temperature 25 °C.

the 8 carbons on its chain.

The higher performance of IX resins and AC might be related to the resin properties such as polymer matrix (acrylic vs. styrene), pore structure (gel-based or macroporous), and functional groups (quaternary ammonium or tertiary ammonium), which affect not only the adsorption capacity but also the adsorption rate (12). As presented in Fig. 3 (b), the equilibrium time for ACs is less than IX resins, suggesting a higher adsorption rate or lower surface area for AC than IX resins. This feature can also be attributed to their textural properties, improving surface-to-volume ratio, thus lowering equilibrium time.

Qu et al. showed that maximum adsorption occurs during the first hour; in this stage, 90% of the PFOA was adsorbed because the surface is free, and the reaction proceeds at a fast rate (Qu et al., 2009) Pramanik et al. investigated the effect of contact time; it was shown that AC led to a more significant reduction of PFOS and PFOA by increasing contact time. This reduction was more prominent for PFOA than PFOS. This was likely due to the higher solubility of PFOA and the degree of attraction by water molecules, which are prevented from making bonding links with the carbon surface (Pramanik et al., 2015)

The removal percentage from various adsorbents is summarized in Fig. 3 (c). The adsorption capacity of different resins and ACs is relatively higher than other adsorbents such as MOF, clay, metal oxide, polymer, salt, and polymeric membranes. Additional factors, including temperature, pH, adsorbent dosage, contact time, and textural properties, can affect the PFASs adsorption onto AC and IX resins. Table 2 depicts the minimum and maximum output parameters based on different adsorbents.

4.1. *XGBoost model for PFAS removal predictions*

Data is separated into test and train sets for each output parameter, with 80% of the data utilized for training and 20% for testing. Bayesian optimization was used to choose the hyper-parameters of the trained XGBoost model. Fig. 4 represents the development of the root-mean-squared-error (RMSE) as the selected evaluation measure for both the testing and training sets throughout the number of boosting rounds. As proved, the RMSE values for both training and testing sets decrease as the number of boosting cycles increases. This fact establishes the training model as a reasonably fitted model, eliminating any possibility of overfitting.

4.2. *Performance of XGBoost model in predicting PFAS removal*

The contribution of features to the training process is among the most critical aspects of any training model. It may be used to determine the relative relevance of different characteristics in a trained model. Fig. 5 shows the trained model's feature significance based on the Total Gain. Among the features presented in Fig. 5, those with the most significant Total Gian influence the PFAS removal prediction. In general, total gain refers to a feature's overall improvement in an evaluation measure (RMSE in this case) compared to all other elements on the sections it is on.

As shown in Fig. 5, pH has the most significant influence on the model, with a total gain of more than 3.19, 10.26, and 1.47 that of particle size for adsorption, equilibrium time, and removal, respectively the model's next most important characteristic. The first most influential factor is pH followed by particle size (Fig. 5a), type 2 (resin and carbon chain), followed by type 3 (MOF) and type 1(metal oxide, salt, and polymer), initial concentration of adsorbent and PFAS, carbon-chain, temperature and finally kinetic of model for adsorption capacity. In Fig. 5b, the first important parameter is pH; the second one is particle size, followed by type 2 (metal oxide, membrane, polymer, and salt), carbon-chain, type 4 (MOF, Carbon family), initial concentration of adsorbent, and PFAS, type 3 (resin), type 1 (clay), temperature and kinetic models. In Fig. 5c, the result shows pH and particle size still are the most significant parameters on PFAS removal. The impact of initial concentrations of adsorbent and PFAS are roughly the same on the PFAS removal. It should point out here that the effect of type 3 (carbon family, membrane), type 4 (resin, clay), carbon-chain, and type 2 (polymer, salt) on the PFAS removal are similar. The contribution of type1(MOF) and temperature and kinetic of the model is also similar.

Overall, pH and particle size predominate PFAS removal, while the kinetics of adsorption and temperature does not play an essential role in

Table 2
Minimum and maximum output parameters of different adsorbents used for PFAS removal.

Adsorbent	Maximum adsorption capacity	Minimum adsorption capacity	Maximum equilibrium time	Minimum equilibrium time	Maximum removal efficiency	Minimum removal efficiency
Membrane	–	–	50	0.33	99.3	86
Resin	3066	43.7	168	1	100	8
Clay	344.8	146.27	32	24	99	99
Carbon family	343.782	0.03	168	4	99	27.6
MOF	734.7	0.17	186	0.016	100	38.6
Polymer	1667.8	460	40	4	83.16	37.42
Metal oxide	5.3	0.007	504	4	80	17
Salt	4.975	0.155	0.083	0.083	99.5	3.1

prediction. The type of adsorbent, especially carbon family and resin, could be crucial features in the adsorption capacity; in contrast, for PFAS removal percentage, the initial concentration of adsorbent and PFAS are more dominant parameters.

4.2.1. Effect of pH on the PFAS removal

The pH of solutions has shown a significant impact on various PFASs removal. PFOS and PFOA dissociate totally in temperatures lower than 25 °C. The PFOS acid dissociation constant (pKa) is 3.27, while pKa for PFOA falls in the range of 0.5–3.8. In a solution with a pH ranging from 3.0 to 11.0, PFOS has mostly existed in anionic forms. The effect of pH on PFOS adsorption is notable, and at pH below 6, PFOS adsorbed was at a high rate while its adsorption slowed down at pH above 6. For example, the PFOS and PFOA equilibrium adsorption capacities reduced from 0.85 to 0.60 mmol. g^{-1} with the increase of the pH from 2.0 to 10.0 (Deng et al., 2015; Pramanik et al., 2015; Qian et al., 2017).

To keep the charge balanced, a cation must join any PFOS anions that adsorb without moving a chloride ion. This process may be aided by increasing the concentration of H^+ in the aqueous phase. As a result of the electrostatic contact (attraction) between the anionic PFOA/PFOS and the protonated amine groups, the enhanced adsorptions of PFOS and PFOA were feasible.

At pH over 6, however, only the quaternary ammonium cation may impact PFOS and PFOA sorption via anion exchange, leading to deprotonation of tertiary amine groups, resulting in continuous adsorption at the remarkably high pH solution. (Schepelina and Zharov, 2008).

The best PFOS adsorption function of IRA67 as an adsorbent has occurred at pH = 3.0 (Pan et al. 2008, 2009). When solution pH is above 10.0, the adsorption capacity on IRA67 decreases significantly as amine groups on the resin had deprotonated and converted into the base form; therefore, their anion exchangeability had been reduced, indicating that anion exchange was the main factor of adsorption (Pan et al. 2008, 2009). Since the adsorption capacity did not decrease to zero when the pH was high, electrostatic adsorption was not the only adsorption mechanism (Dolatabadi et al., 2020). Isotherm observations indicated other mechanisms involved in the adsorption process. At pH higher than 10.0, micelles or hemi-micelles formation could be another mechanism, however, with the rise in pH, the formation is reduced. Hydrophobic interaction was also involved in the adsorption, and at high pH, it has been the primary adsorption mechanism (Gao et al., 2017; Pan et al., 2009).

Studies on AC revealed that significant H^+ congregate at the AC-water interface at low pH. The AC surface was positive, which magnified electrostatic interplays with the PFOS anions, promoting the adsorption of PFOS at the AC-water interface. Conversely, the positive interface between AC and water continuously became neutral or negative while increasing the pH. The electrostatic interaction changed from attraction to repulsion between the AC and PFOS surfaces, dropping the adsorption capacity of PFOS (Deng et al., 2015; Meng et al., 2019; Zhao et al., 2011).

Studies on the adsorption of activated carbons under acidic and basic conditions revealed that adsorption was more favorable in acidic solutions (@ pH 4.0) than basic solutions (@ pH 9.0). The accessible area of adsorbent in acid media was remarkably higher than the sorption of PFOA and PFOS in the basic media. A higher tendency that causes an increase in electrostatic attraction between adsorbate and adsorbent is likely to elucidate the sorption difference at lower pH. Also, it was concluded that the abundance of OH^- ions in the basic media could limit the diffusion of organic molecules onto the active sites present on the surface of activated carbons; as a result, the chances of adsorption are reduced (Fagbayigbo et al., 2017).

Bamboo-derived activated carbon (BAC) was another adsorbent studied in our database. Its removal percentage of PFOA sharply decreased when the pH increased from 2.0 to 4.0 but became stable at pH above 5.0. Since the pKa value of PFOA is −2.0, it exists as anions at pH above 2.0. While increasing the pH, the electrostatic attraction between PFOA and BAC at pH < 3.2 decreased, resulting in a notable decrease of PFCAs removal. At pH > 5.0, the electrostatic repulsion hindered the adsorption of PFOA on BAC, whereas hydrophobic interplay performed a principal role in the sorption process. In comparison, PFOA removal percentage by resin IRA67 decreased progressively with increasing wastewater pH from 2.0 to 9.0. As the amine groups on IRA67 have a pKa of 9.5, they protonated at pH < 9.5, and the anionic PFOA can be adsorbed via anion exchange. Increasing the pH of the solution decreased the number of protonated amine groups, and some amine groups lost the anion exchange capacity for PFOA, resulting in decreased elimination of PFOA (Du et al., 2015).

Increasing pH from 3.0 to 9.0 led to a reduction in PFAS sorption onto (GAC); the amounts of PFOA and PFOS adsorbent on GAC were reduced by 27.5% and 32.0%, respectively (Zhang et al., 2021). The same trend was observed for biochar. On the other hand, the decrease in pH due to increased electrostatic adsorption among adsorbents and PFAS intensified PFAS sorption. The PFAS adsorption was mainly related to carbon surface basicity, meaning that the capacity of anion exchange or high acid neutralizing was crucial for abundant adsorption of PFOA and PFOS. At the pH values observed in this study, PFOS and PFOA primarily were in anion forms. According to Zhang et al., (2021) zero-point charge (pHzpc) was around 7.2 and 6.9 for GAC and biochar, respectively. As a result, at pH 7.0 in the solution, the GAC surface has a positive charge, while biochar has a negative charge. By decreasing the pH of the solution, the electrostatic attraction between anionic PFAS and positively charged adsorbents (GAC at pH < 7.2; biochar at pH < 6.9) increased, which resulted in a significant increase in PFAS adsorption. When the pH was increased, the adsorbent surface change altered from positive to negative (biochar at pH > 6.9; GAC at pH > 7.2), and there was a significant electrostatic repulsion between negatively charged absorbents and anionic PFAS (Zhang et al., 2021).

4.2.2. Effect of particle size on the PFAS removal

Type 1, particle size of adsorbent, and the initial concentration of PFAS in Fig. 5 are also important factors, such that their greater values could result in counter-predictive responses in the trained model for adsorption capacity, and equilibrium time, and removal percentage. In addition, the surface area of adsorbents generally increases with

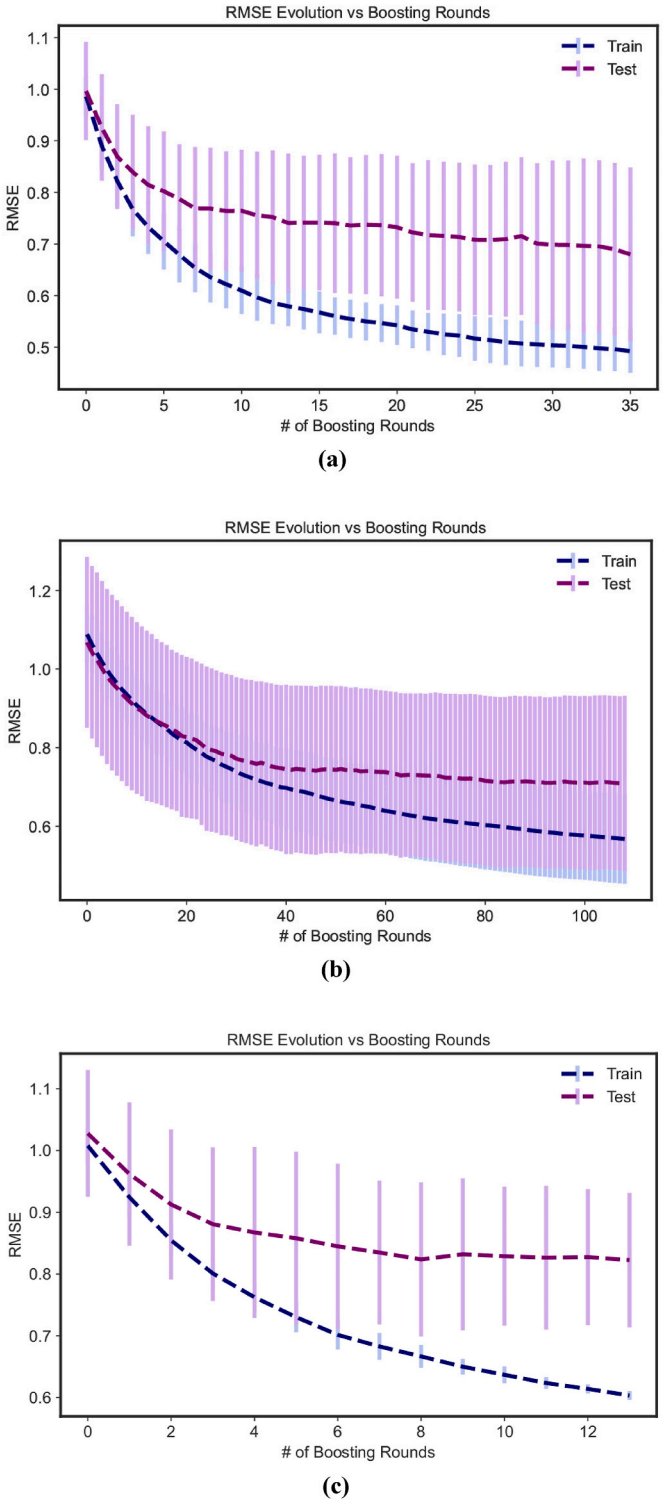

Fig. 4. - XGBoost performance curves evolution of RMSE for the train/test set through a number of boosting rounds (a) Adsorption capacity (b) Equilibrium time (C) Removal efficiency.

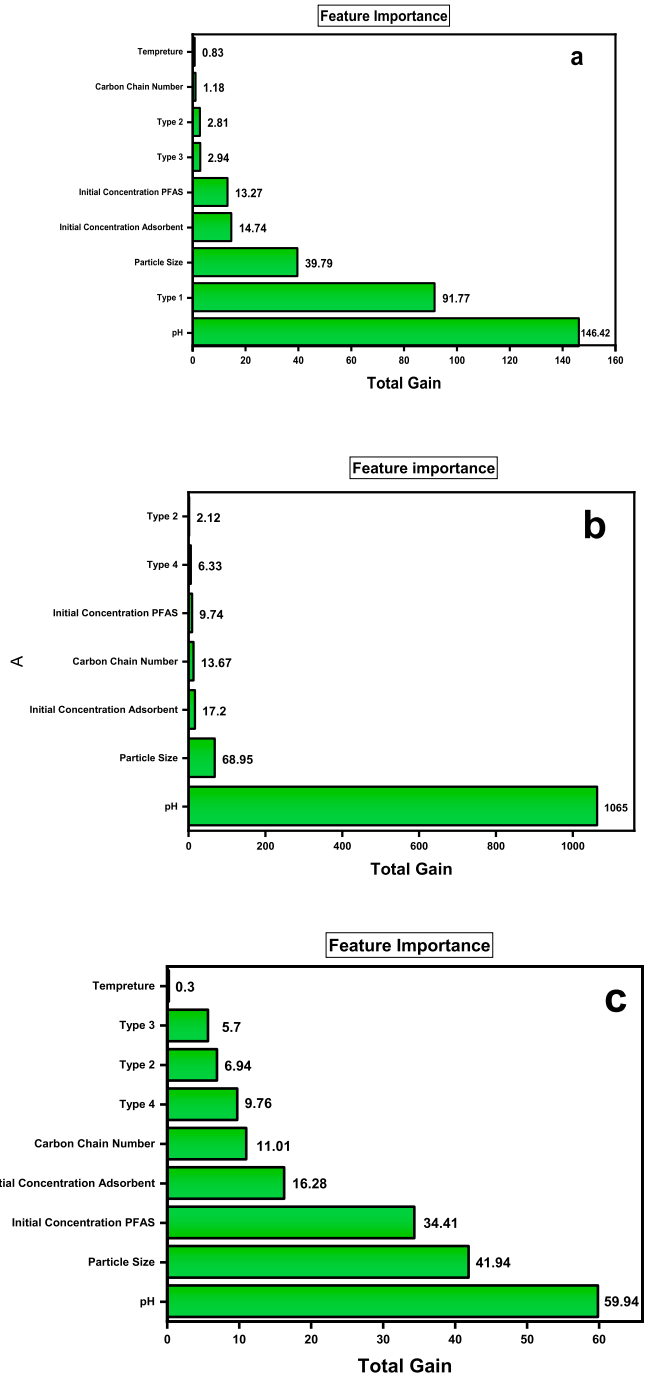

Fig. 5. Feature importance model for PFAS removal (a) Adsorption capacity (b) Equilibrium time (c) PFAS removal efficiency.

decreasing particle size, resulting in a significantly higher number of available adsorbent sites(Sun, Fan & Bai). More adsorbing sites lead to more surface reactions, and adsorption capacity will improve, and equilibrium time will decline significantly (Zhang et al., 2021). However, the main drawback is the difficulty of regenerating smaller particle sizes while synthesizing adsorbents (Hassan et al., 2020).

Hassan et al., (2020) studied the effect of adsorbent particle size on PFOS adsorption. The adsorbents used in this study were red mud modified sawdust (RMSDN600) and unmodified sawdust (SDN600), both with three different particle sizes (<2 mm, <1 mm, and <0.5 mm). To study the effect of particle size, the experiments were carried out at an initial PFOS concentration of 248.48 mg. L^{-1}, equilibration time of 24 h, and pH of 5.5 ± 0.2 for RMSDN600 and SDN600 at an adsorbent concentration of 1.0 g. L^{-1}. The results showed that the adsorption capacity of SDN600 was increased while the particle size was decreased. It was 17 (mg. g^{-1}), 90 (mg. g^{-1}), and 108 (mg. g^{-1}) for SDN600 with <2 mm, <1 mm, and <0.5 mm particle size, respectively. Likewise, RMSDN600 adsorption capacity exhibited the same trend. Its adsorption capacity was 130 (mg. g^{-1}), 155 (mg. g^{-1}), and 150(mg. g^{-1}) with <2 mm, <1 mm, and <0.5 mm particle size, respectively. At the same mass,

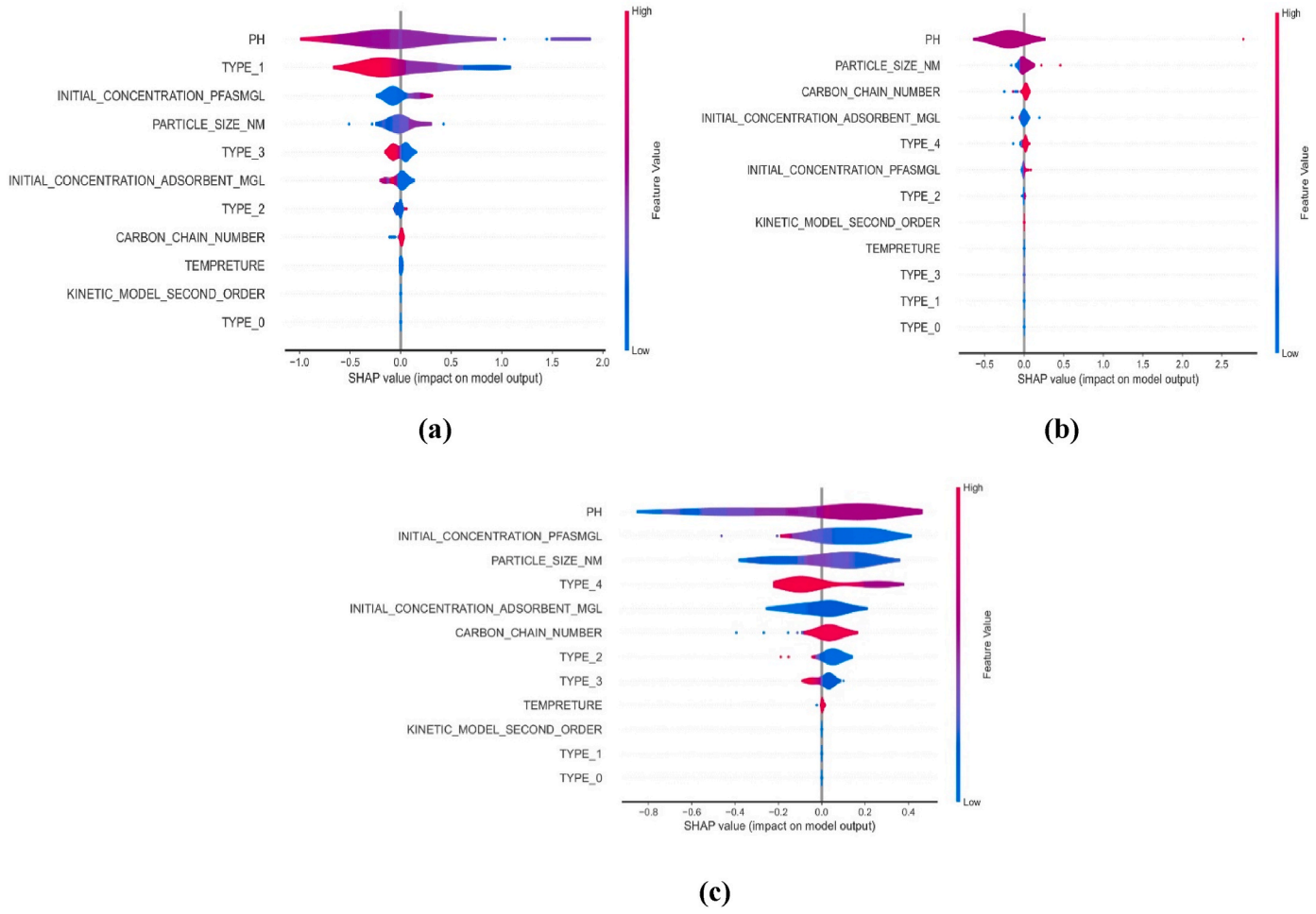

Fig. 6. SHAP summary plot of input variables. The higher absolute SHAP values represent more contribution of the variable to the model predictions. The color of points indicates the magnitude of the variable values. SHAP summary plots of the "tested" set for various adsorbents: (a) Adsorption capacity, (b) Equilibrium time, and (c) Removal efficiency. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the active sites of the adsorbent with smaller size are higher than that of the larger one resulting in higher PFOS removal efficacy (Hassan et al., 2020).

Wang et al., (2019) showed that decreasing the particle size of ACs significantly improves their adsorption for OBS sodium p-perfluorinated nonenoxybenzene sulfonate, while the small particles sizes were unfavorable for adsorption due to the destruction of the pore structure during the mechanical grinding process. They reported that by decreasing the particle size of ACs from 150 mesh to 25 mesh (pH = 7.0), the adsorbed amount of OBS is increased from 0.05 (mmol. g^{-1}) to 0.25 (mmol.g^{-1}), respectively (Wang et al., 2019).

Du et al., (2015) argue that the small size of the adsorbent could speed up the replacement process. They apply powder-activated carbon (PAC) and GAC granular activated carbon to remove perfluoro-2-propoxypropanoic acid (GenX). It is noted that the adsorbed amounts of GenX on PAC (0.2 mmol. g^{-1}) were more significant than that of GAC (0.12 mmol. g^{-1}) during the first 3 h. Although their experiments suggested that the particle size of ACs influenced the initial adsorption rate, after 48 h of adsorption, PACs and GACs had similar adsorbed amounts for GenX (Du et al., 2015).

Textural properties of the adsorbent are an essential parameter that should be considered when comparing two samples. Zhang et al., (2021) showed that PFOS could be efficiently removed by GAC having a particle size ranging from 0.1 to 0.3 mm (S_{BET} = 757 m^2 g^{-1} and pore size of 1.39 nm). The other experimental variables were pH = 7, T = 25 °C, and the initial PFOS, and GAC concentrations were 10 mg.L^{-1} and 100 mg. L^{-1}, respectively. They reported that the adsorption capacity and equilibrium time were 66.4 mg g^{-1} and 24 h (Zhang et al., 2021). In a similar study by Steigerwald et al. (Steigerwald and Ray, 2021) with a GAC with roughly the same textural properties as that of Zhang et al., (2021), they applied the same conditions for PFOS elimination by the GAC, namely SCGKOH sample. The particle size of the adsorbent was between 0.3 and 0.6 mm, (SBET = 858 m^2 g^{-1}, and pore size 1.42 nm). Their results showed that the adsorption capacity was 43.4 mg g^{-1}, and the equilibrium time was 24 h (Steigerwald and Ray, 2021). Their results can be followed by our output results (Fig. 5) obtained from our model and can reveal decreasing particle size's role in improving adsorption capacity.

4.2.3. *Effect of PFAS and adsorbent concentration on the PFAS removal*

Fig. 5 (a, b) revealed that the third important factor affecting adsorption capacity and equilibrium time is adsorbent concentration. Increasing adsorbent dosage increases the adsorption capacity due to more available sites. In other words, increasing the adsorbent dosage increases the available surface area for adsorption, which results in higher removal efficiency (Deng et al., 2015; Qian et al., 2017). The study of Pramanik et al. (Pramanik et al., 2015) showed that by increasing adsorbent dosage, the removal efficiency increases. Increasing the adsorbent dosage from 5.0 to 30.0 mg L^{-1} (mg.L^{-1}) increased the removal efficiency from 35% to 64% and 27%–58% for PFOS and PFOA, respectively (Pramanik et al., 2015). Due et al.(Du et al., 2015) studied the effect of adsorbent dosage on adsorption capacity using two different adsorbents, BAC and IRA67. They found that

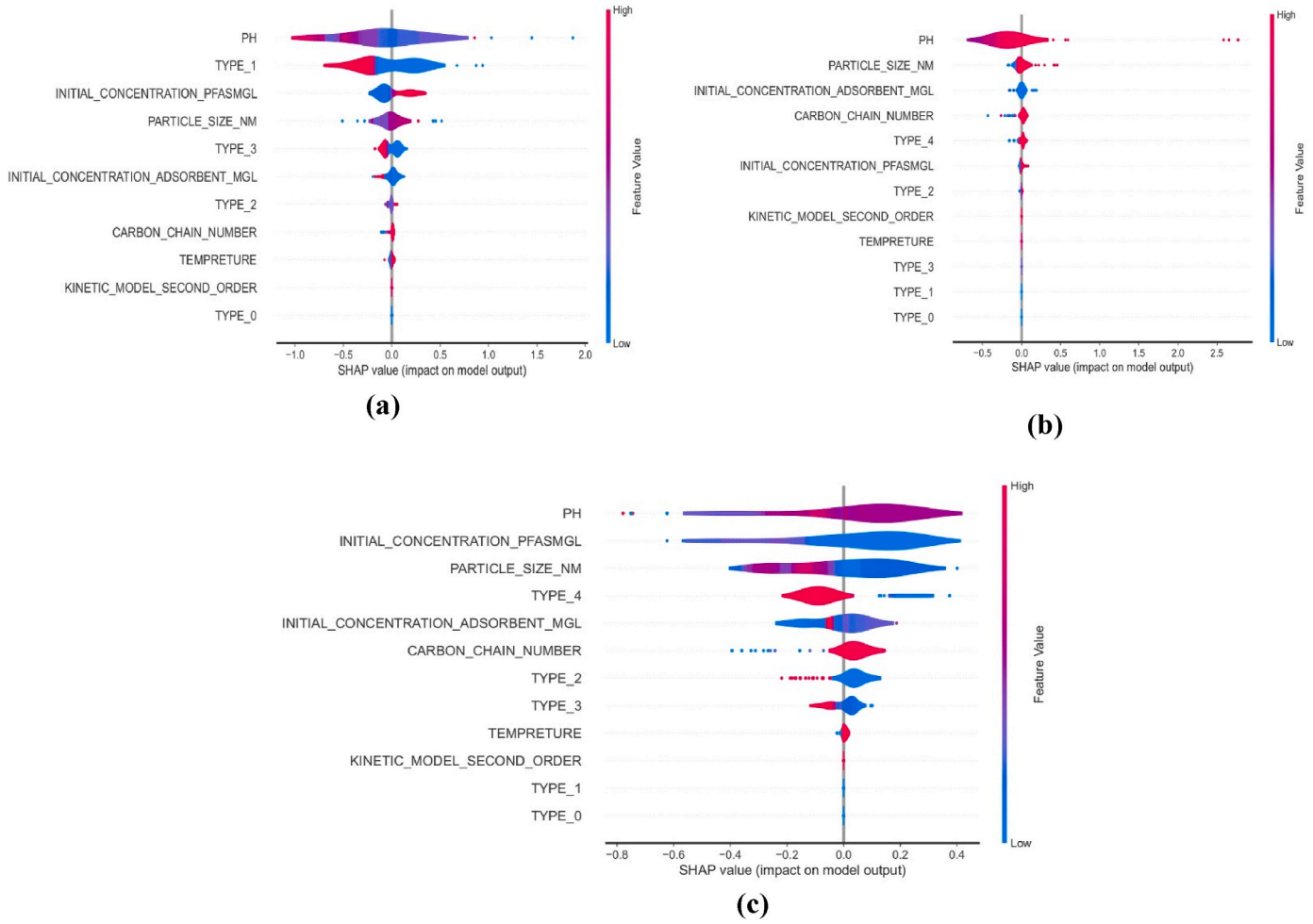

Fig. 7. SHAP summary plot of input variables. The higher absolute SHAP values represent more contribution of the variable to the model predictions. The color of points indicates the magnitude of the variable values. SHAP summary plots of the "**trained**" set for various adsorbents of adsorbent: (a) Adsorption capacity, (b) Equilibrium time, and (c) Removal efficiency. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

regardless of the type of adsorbent, the adsorption capacity increases while the adsorbent concentration increases. When BAC doses increased from 0.1 to 0.5 gL^{-1}, the removal of PFOA increased from 33.2% to 96.6%. A similar trend is perceived for the other adsorbate types, PFHxA and PFHpA. By increasing the BAC dosage, PFHxA and PFHpA removal increased. When the BAC dose is as low as 0.3 $g.L^{-1}$, it can adsorb more PFOA PFHxA. PFHxA's removal percentage at this BAC concentration was only 8.8%, while PFOA's was 88.4%, representing PFOA selective adsorption at low BAC concentration. There is a hydrophobic interaction between long-chain PFCAs and BAC during the adsorption process, which leads to competitive adsorption of PFHxA, PFH, and PFOA on BAC. In addition, when IRA67 doses increased, the removal percent of the three PFCAs all increased since anion exchange is the main reason for the sorption of PFCAs on IRA67. Notwithstanding intra-particle diffusion, the removal percent on IRA 67 still follows the increasing order of PFHxA < PFHpA < PFOA, indicating that the formation of hemi-micelles and micelles may be the dominant reason for the high removal of long-chain PFCAs (Du et al., 2015). Studies have shown that different adsorbents show approximately the same removal at the same concentration, revealing that the concentration is more important in adsorption than the type of adsorbent. For instance, for GAC and AI400, their removal is slightly the same at an initial concentration of 50 mg. L^{-1}. At pH = 3.0 and pH = 7.0, their PFOS removals are 35.5% and 43.15% and 31.31% and 35.158% for AI400 and GAC, respectively. On the other hand, the adsorption capacity increased too when the concentration increased. For instance, when the concentration of GAC increased from 50 $mg.L^{-1}$ to 100 $mg.L^{-1}$, its removal increased from 35.158% to 98% at pH = 7 (Deng et al., 2012). Considering PAC as another adsorbent showed the same trend. Its removal rate increased from 51.1% to 99.9%, as the (PAC) concentration increased from 0.1 gL^{-1} to 10.0 gL^{-1}. This trend may be due to PACs' available high surface area (SSA: 1000 m^2g^{-1}) (Qu et al., 2009).

By comparing data from different articles, the same trend is found. Under the same condition but at different initial concentrations, the effect of the initial concentration of adsorbent is apparent at pH = 7.0 for removing PFOA. Two adsorbents of MIEX and Purolite®A860 were used with an initial concentration of 100 mg. L^{-1} and 1000 mg. L^{-1}, respectively. PFOA removal is increased as the adsorbent concentration increases. The elimination of PFOA using MIEX is 52%, and Purolite®A860 is 84% (Dixit et al., 2019; Park et al., 2020). As shown above and refer to data from ML output presented in Fig. 5 (a, b), adsorbent type does not have high significance compared to adsorption concentration.

Comparing two different types of activated carbon, AC-KOH and PACF showed a similar trend as above. AC-KOH and PACF with the same size of 50 μm are used as adsorbents, with different initial concentrations to remove 100 mg. L^{-1} of PFOA at pH @ 4.0, AC-KOH with an initial concentration of 5000 mg. L^{-1} removed 90% of PFOA. At the same time, the removal of PACF was only 32% at 100 mg. L^{-1} of the initial concentration of PACF (Chen et al., 2017; Fagbayigbo et al., 2017).

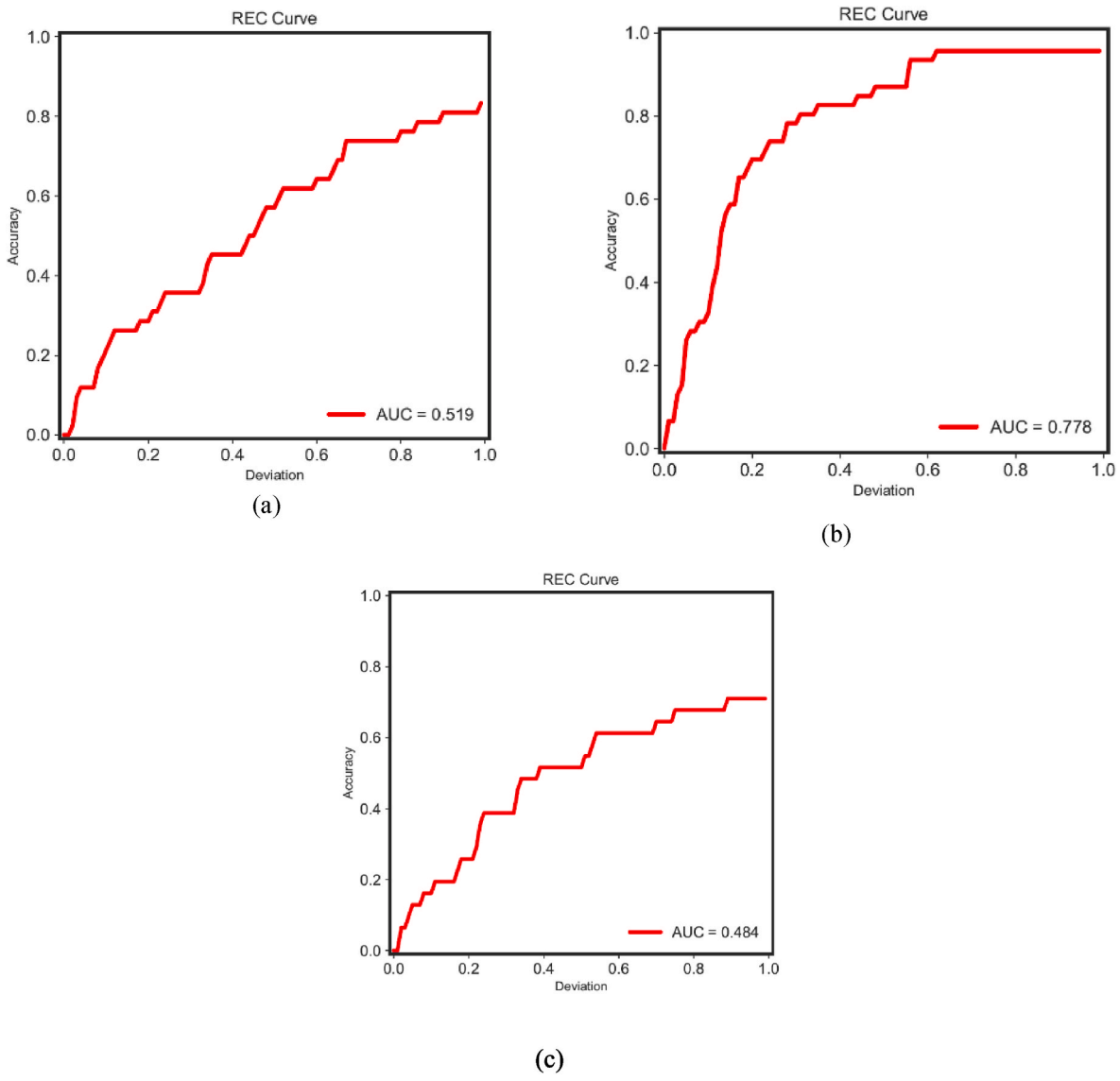

Fig. 8. XGBoost performance curves AUC (a) Adsorption capacity (b) Equilibrium time (c) Removal efficiency.

4.3. Can the XGBoost model understand the importance of parameters for PFAS removal?

Shapley additive explanations (SHAP) focus on cooperation game theory and measure the relevance of each variable to explain the contribution of every participant (input parameter) to the output values. In addition to the XGBoost feature importance plot, SHAP can be a valuable tool for revealing the interpretation of each of the tested model's predictions. SHAP is an approach for interpreting the output of any ML model using game theory and the local impact of each feature (Lundberg and Lee, 2017). The importance of the SHAP feature is straightforward: features with significant fundamental SHAP values are valuable. This is the feature's overall impact on the testing set for PFAS removal.

The SHAP values of the variables were calculated using the XGB model with all 236 data received from different adsorbents to understand the influence of each variable on model prediction. Fig. 6 shows the SHAP summary charts of the trained XGBoost model for (a) adsorption capacity, (b) equilibrium time, and (c) removal efficiency using a tree explainer, which combines feature significance and feature effect considerations. The SHAP values and their affecting value (scale and distribution) to the model high (red) and low (blue) are presented for each attribute. The density of the dots in the summary graphic shows the correct dispersion of exemplars in the testing data set. The outcome is consistent with what we saw in Fig. 5 (feature importance), representing that the model has been appropriately trained and the potential of overfitting has been reduced.

For example, red and blue indicate big and tiny molecules, respectively, in pH. The SHAP value can also show if the related variable positively or negatively impacts PFAS removal. Positive and negative SHAP scores suggest that taking the variable into account leads to PFAS absorption. As a result, the SHAP summary graphic depicts the direction and distribution of each variable's contribution. This is the feature's overall impact on the testing set for PFAS removal. The order of input variables on the y-axis in the SHAP summary graphic illustrates the priority of variable relevance based on the average of total SHAP values (Figs. 6 and 7).

Fig. 7 should be in agreement with the XGBoost feature relevance (Fig. 5). As can be seen, the attributes are ranked the same in both Figures (Figs. 5 and 7) and have a similar impact/importance on the model output. Variable significance is evaluated using the average of absolute SHAP values, reflecting the model's overall variable contribution. As a result, the ordering of varying importance is inconsistent across all data. Under certain circumstances, operating conditions such

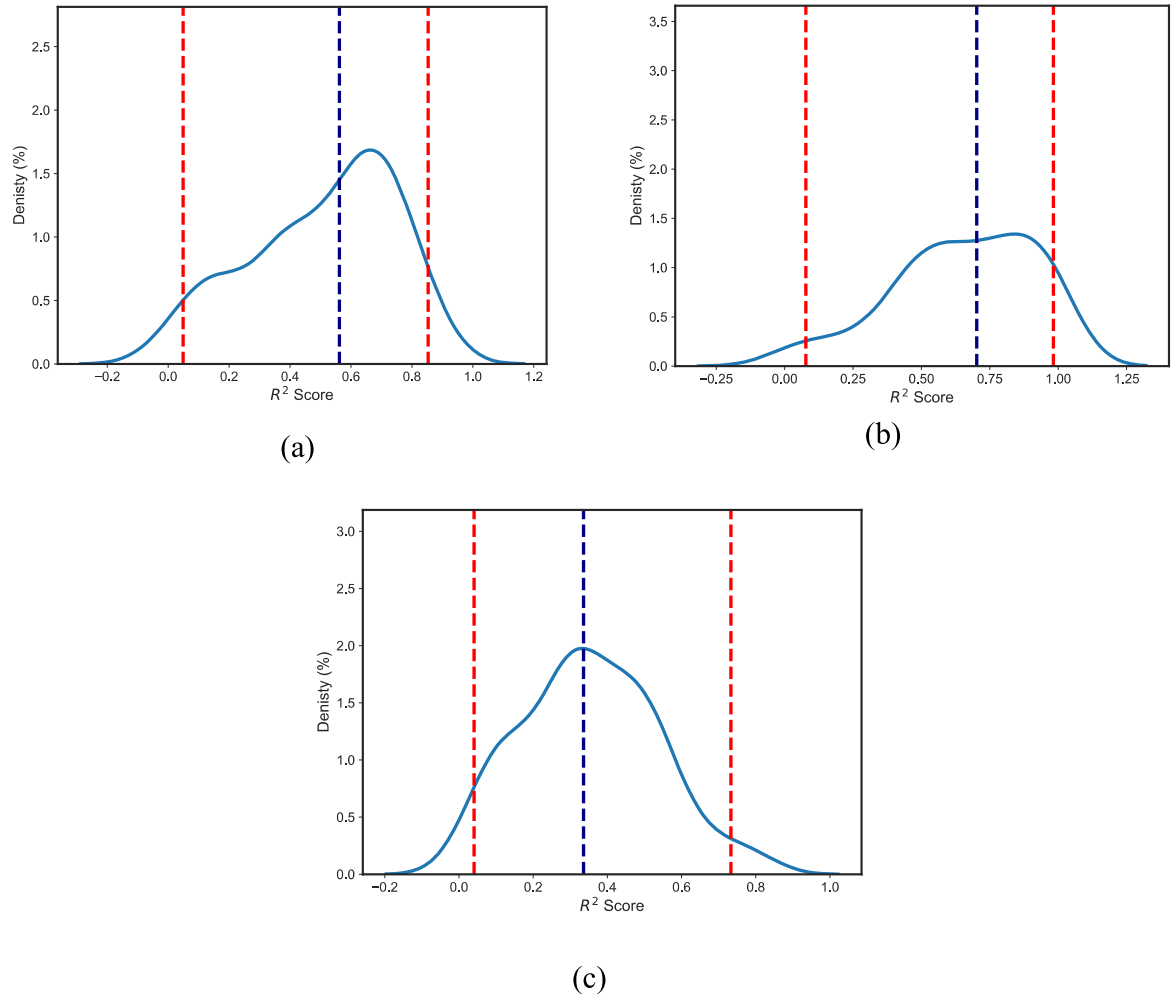

Fig. 9. Histograms of the micro-average R^2 of test sets according to the generalization pipeline for (a) Adsorption capacity, (b) Equilibrium time, and (c) Removal efficiency.

as particle size (red dots with negative SHAP values) can considerably contribute. It's worth noting that the XGBoost total gain is evaluated using training data. In contrast, the SHAP summary plot is produced over the testing data using the trained model.

As can be seen in Fig. 7a, pH exhibits the largest impact on the adsorption capacity. Furthermore, Type 1 is the second most important feature in predicting the adsorption capacity. In contrast, it is not important in predicting the other equilibrium time and removal, as shown in Fig. 5. The top five parameters with the greatest influence in predicting PFAS removal are pH, initial concentration of PFAS, particle size, Type 4, and initial adsorbent concentration.

As shown in Fig. 8, equilibrium time has the most significant area under the curve (AUC) values of 0.77, whereas removal has the lowest performance with an AUC of 0.484. The probability values may be calculated using their AUC values based on the REC curves shown in Fig. 8, where equilibrium time has an AUC of 0.77, adsorption capacity has an AUC of 0.51, and removal has the lowest AUC of 0.41. Fig. 5 shows the trained model's feature contribution using the XGBoost total gain metric, with the pH having the greatest total gain of all the features. As a result, pH plays the most significant role in adsorption and equilibrium time to remove PFAS.

The histograms of the micro-average coefficient of determination (R^2) of the testing set (10 percent of data) for 100 iterations (using random seed) are shown in Fig. 9, with the navy and red dashed lines representing the median (50th percentile) and lower and upper confidence intervals (based on 95 percent significance levels), respectively. The presented generalization results and the REC results are shown in Fig. 7.

5. Conclusions

In this study, we proposed a scalable-interpretable modeling framework that could handle various engineering problems, such as PFAS removal using different adsorbents. We showed that ML models could reliably estimate PFAS removal using different adsorbents and removal conditions as input variables. Based on knowledge, most literature focus on the adsorption removal of PFOS and PFOA, and only a few reports onC2-C11 PFAS adsorption. Some technologies such as advanced oxidation process (AOP) and photocatalytic oxidation have not been investigated due to unrelated parameters for this research.

The conclusions of the study are as follows:

- The adsorption capacity of different adsorbents regarding material parameters was investigated. The result showed that pH (146.2), Type 1 (91.77), and particle size (39.79), are crucial variables in predicting PFAS adsorption.
- The equilibrium time of PFAS based on different material parameters was also evaluated. A similar order of impact importance as pH (1065), particle size (68.95), and initial concentration (17.2) on adsorption capacity was observed.
- Removal of PFAS based on different material parameters was studied; after the pivotal impact of pH (59.94) and particle size (41.94)

on PFAS removal, the initial concentrations of adsorbent (34.41) is critical to know for predicting PFAS removal.

We showed through our case studies that these models might be interpreted by employing SHAP values, at least in the domain of feature significance. As data becomes more readily available, ML algorithms will have more opportunities to find answers to real-world issues. The outcome of this work can be used as a platform for designing advanced materials as absorbents for PFAS removal and optimizing process operating conditions.

Author Statement

Elika Karbassiyazdi: Writing-Original draft preparation, data miming, leading author. Fatemeh Fattahi: data mining and writing, discussion. Negin Yousefi: data mining and writing, discussion. Amirhessam Tahmassebi: Simulation, and writing, discussion. Arsia Afshar Taromi: data mining and writing, discussion. Javad Zyaie Manzari: data mining and writing, discussion. Amirhossein Gandomi: Reviewing and Editing, Supervision, Ali Altaee: Reviewing and Editing, Supervision, Amir Razmjou: Original idea and Concept development, Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

A. R. appreciates the CPDRF fund (PRO20-11072)) that is provided by UTS.

References

Bolan, N., Sarkar, B., Yan, Y., Li, Q., Wijesekara, H., Kannan, K., Tsang, D.C.W., Schauerte, M., Bosch, J., Noll, H., Ok, Y.S., Scheckel, K., Kumpiene, J., Gobindlal, K., Kah, M., Sperry, J., Kirkham, M.B., Wang, H., Tsang, Y.F., Hou, D., Rinklebe, J., 2021. Remediation of poly- and perfluoroalkyl substances (PFAS) contaminated soils – to mobilize or to immobilize or to degrade? J. Hazard Mater. 401, 123892.

Catelan, D., Biggeri, A., Russo, F., Gregori, D., Pitter, G., Da Re, F., Fletcher, T., Canova, C., 2021. Exposure to perfluoroalkyl substances and mortality for COVID-19: a spatial ecological analysis in the veneto region (Italy). Int. J. Environ. Res. Publ. Health 18 (5), 2734.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., 2015. Xgboost: extreme gradient boosting. R package version 0.4-2 1 (4), 1–4.

Chen, W., Zhang, X., Mamadiev, M., Wang, Z., 2017. Sorption of perfluorooctane sulfonate and perfluorooctanoate on polyacrylonitrile fiber-derived activated carbon fibers: in comparison with activated carbon. RSC Adv. 7 (2), 927–938.

Cheng, W., Ng, C.A., 2019. Using machine learning to classify bioactivity for 3486 per- and polyfluoroalkyl substances (PFASs) from the OECD list. Environ. Sci. Technol. 53 (23), 13970–13980.

Deng, S., Zheng, Y.Q., Xu, F.J., Wang, B., Huang, J., Yu, G., 2012. Highly efficient sorption of perfluorooctane sulfonate and perfluorooctanoate on a quaternized cotton prepared by atom transfer radical polymerization. Chem. Eng. J. 193–194, 154–160.

Deng, S., Nie, Y., Du, Z., Huang, Q., Meng, P., Wang, B., Huang, J., Yu, G., 2015. Enhanced adsorption of perfluorooctane sulfonate and perfluorooctanoate by bamboo-derived granular activated carbon. J. Hazard Mater. 282, 150–157.

Dixit, F., Barbeau, B., Mostafavi, S.G., Mohseni, M., 2019. PFOA and PFOS removal by ion exchange for water reuse and drinking applications: role of organic matter characteristics. Environ. Sci. J. Integr. Environ. Res.: Water Research & Technology 5 (10), 1782–1795.

Dolatabadi, M., Ahmadzadeh, S., 2019. A rapid and efficient removal approach for degradation of metformin in pharmaceutical wastewater using electro-Fenton process; optimization by response surface methodology. Water Sci. Technol. 80 (4), 685–694.

Dolatabadi, M., Ahmadzadeh, S., Ghaneian, M.T., 2020. Mineralization of mefenamic acid from hospital wastewater using electro-Fenton degradation: optimization and identification of removal mechanism issues. Environ. Prog. Sustain. Energy 39 (3), e13380.

Dolatabadi, M., Naidu, H., Ahmadzadeh, S., 2021a. A green approach to remove acetamiprid insecticide using pistachio shell-based modified activated carbon; economical groundwater treatment. J. Clean. Prod. 316, 128226.

Dolatabadi, M., Świergosz, T., Ahmadzadeh, S., 2021b. Electro-Fenton approach in oxidative degradation of dimethyl phthalate - the treatment of aqueous leachate from landfills. Sci. Total Environ. 772, 145323.

Domingo, J.L., Nadal, M., 2019. Human exposure to per- and polyfluoroalkyl substances (PFAS) through drinking water: a review of the recent scientific literature. Environ. Res. 177, 108648.

Du, Z., Deng, S., Chen, Y., Wang, B., Huang, J., Wang, Y., Yu, G., 2015. Removal of perfluorinated carboxylates from washing wastewater of perfluorooctanesulfonyl fluoride using activated carbons and resins. J. Hazard Mater. 286, 136–143.

Fagbayigbo, B.O., Opeolu, B.O., Fatoki, O.S., Akenga, T.A., Olatunji, O.S., 2017. Removal of PFOA and PFOS from aqueous solutions using activated carbon produced from Vitis vinifera leaf litter. Environ. Sci. Pollut. Control Ser. 24 (14), 13107–13120.

Gao, Y., Deng, S., Du, Z., Liu, K., Yu, G., 2017. Adsorptive removal of emerging polyfluoroalky substances F-53B and PFOS by anion-exchange resin: a comparative study. J. Hazard Mater. 323, 550–557.

Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why Do Tree-Based Models Still Outperform Deep Learning on Tabular Data?.

Hassan, M., Liu, Y., Naidu, R., Du, J., Qi, F., 2020. Adsorption of Perfluorooctane Sulfonate (PFOS) onto Metal Oxides Modified Biochar, vol. 19. Environmental Technology & Innovation, 100816.

Holmquist, H., Fantke, P., Cousins, I.T., Owsianiak, M., Liagkouridis, I., Peters, G.M., 2020. An (Eco)Toxicity life cycle impact assessment framework for per- and polyfluoroalkyl substances. Environ. Sci. Technol. 54 (10), 6224–6234.

Jeong, N., Chung, T.-h., Tong, T., 2021. Predicting micropollutant removal by reverse osmosis and nanofiltration membranes: is machine learning viable? Environ. Sci. Technol. 55 (16), 11348–11359.

Kibbey, T.C.G., Jabrzemski, R., O'Carroll, D.M., 2020. Supervised machine learning for source allocation of per- and polyfluoroalkyl substances (PFAS) in environmental samples. Chemosphere 252, 126593.

Kibbey, T.C.G., Jabrzemski, R., O'Carroll, D.M., 2021. Source allocation of per- and polyfluoroalkyl substances (PFAS) with supervised machine learning: classification performance and the role of feature selection in an expanded dataset. Chemosphere 275, 130124.

Leo Kumar, S.P., 2017. State of the art-intense review on artificial intelligence systems application in process planning and manufacturing. Eng. Appl. Artif. Intell. 65, 294–329.

Li, L., Rong, S., Wang, R., Yu, S., 2021. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: a review. Chem. Eng. J. 405, 126673.

Liu, K., Zhang, S., Hu, X., Zhang, K., Roy, A., Yu, G., 2015. Understanding the adsorption of PFOA on MIL-101(Cr)-Based anionic-exchange metal–organic frameworks: comparing DFT calculations with aqueous sorption experiments. Environ. Sci. Technol. 49 (14), 8657–8665.

Liu, L., Liu, Y., Gao, B., Ji, R., Li, C., Wang, S., 2020. Removal of perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS) from water by carbonaceous nanomaterials: a review. Crit. Rev. Environ. Sci. Technol. 50 (22), 2379–2414.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

Machine learning, 1995. Neural and Statistical Classification. Ellis Horwood.

Meng, P., Fang, X., Maimaiti, A., Yu, G., Deng, S., 2019. Efficient removal of perfluorinated compounds from water using a regenerable magnetic activated carbon. Chemosphere 224, 187–194.

Merayo, D., Rodríguez-Prieto, A., Camacho, A.M., 2019. Comparative analysis of artificial intelligence techniques for material selection applied to manufacturing in Industry 4.0. Procedia Manuf. 41, 42–49.

Pan, B., Zhang, W., Pan, B., Qiu, H., Zhang, Q., Zhang, Q., Zheng, S., 2008. Efficient removal of aromatic sulfonates from wastewater by a recyclable polymer: 2-naphthalene sulfonate as a representative pollutant. Environ. Sci. Technol. 42 (19), 7411–7416.

Pan, G., Jia, C., Zhao, D., You, C., Chen, H., Jiang, G., 2009. Effect of cationic and anionic surfactants on the sorption and desorption of perfluorooctane sulfonate (PFOS) on natural sediments. Environ. Pollut. 157 (1), 325–330.

Park, M., Daniels, K.D., Wu, S., Ziska, A.D., Snyder, S.A., 2020. Magnetic ion-exchange (MIEX) resin for perfluorinated alkylsubstance (PFAS) removal in groundwater: roles of atomic charges for adsorption. Water Res. 181, 115897.

Pramanik, B.K., Pramanik, S.K., Suja, F., 2015. A comparative study of coagulation, granular- and powdered-activated carbon for the removal of perfluorooctane sulfonate and perfluorooctanoate in drinking water treatment. Environ. Technol. 36 (20), 2610–2617.

Qian, J., Shen, M., Wang, P., Wang, C., Li, K., Liu, J., Lu, B., Tian, X., 2017. Perfluorooctane sulfonate adsorption on powder activated carbon: effect of phosphate (P) competition, pH, and temperature. Chemosphere 182, 215–222.

Qu, Y., Zhang, C., Li, F., Bo, X., Liu, G., Zhou, Q., 2009. Equilibrium and kinetics study on the adsorption of perfluorooctanoic acid from aqueous solution onto powdered activated carbon. J. Hazard Mater. 169 (1), 146–152.

Radfar, P., Bazaz, S.R., Mirakhorli, F., Warkiani, M.E., 2021. The role of 3D printing in the fight against COVID-19 outbreak. J. 3d Print. Med. https://doi.org/10.2217/3dp-020-0028.

Raza, A., Bardhan, S., Xu, L., Yamijala, S.S.R.K.C., Lian, C., Kwon, H., Wong, B.M., 2019. A machine learning approach for predicting defluorination of per- and

polyfluoroalkyl substances (PFAS) for their efficient treatment and removal. Environ. Sci. Technol. Lett. 6 (10), 624–629.
Rusdah, D.A., Murfi, H., 2020. XGBoost in handling missing values for life insurance risk prediction. SN Appl. Sci. 2 (8), 1–10.
Schepelina, O., Zharov, I., 2008. Poly(2-(dimethylamino)ethyl methacrylate)-modified nanoporous colloidal films with pH and ion response. Langmuir 24 (24), 14188–14194.
Singh, K.P., Gupta, S., 2012. Artificial intelligence based modeling for predicting the disinfection by-products in water. Chemometr. Intell. Lab. Syst. 114, 122–131. Complete.
Steigerwald, J.M., Ray, J.R., 2021. Adsorption behavior of perfluorooctanesulfonate (PFOS) onto activated spent coffee grounds biochar in synthetic wastewater effluent. J. Hazard. Mater. Lett. 2, 100025.
Su, A., Rajan, K., 2021. A database framework for rapid screening of structure-function relationships in PFAS chemistry. Sci. Data 8 (1), 14.
Sun, Z., Fan, W. & Bai, Y., 'A flexible method to fabricate exsolution-based nanoparticle-decorated materials in seconds', Adv. Sci., vol. n/a, no. n/a, p. 2200250.
Tahmassebi, A., Motamedi, M., Alavi, A.H., Gandomi, A.H., 2022. An explainable prediction framework for engineering problems: case studies in reinforced concrete members modeling. Eng. Comput. 39 (2), 609–626.
Tahmassebi, A., Smith, T., 2021. Slickml: Slick machine learning in python. https://github.com/slickml/slick-ml.
Uraikul, V., Chan, C.W., Tontiwachwuthikul, P., 2007. Artificial intelligence for monitoring and supervisory control of process systems. Eng. Appl. Artif. Intell. 20 (2), 115–131.
Wang, W., Maimaiti, A., Shi, H., Wu, R., Wang, R., Li, Z., Qi, D., Yu, G., Deng, S., 2019. Adsorption behavior and mechanism of emerging perfluoro-2-propoxypropanoic acid (GenX) on activated carbons and resins. Chem. Eng. J. 364, 132–138.
Zhang, D., He, Q., Wang, M., Zhang, W., Liang, Y., 2021. Sorption of perfluoroalkylated substances (PFASs) onto granular activated carbon and biochar. Environ. Technol. 42 (12), 1798–1809.
Zhao, D., Cheng, J., Vecitis, C.D., Hoffmann, M.R., 2011. Sorption of perfluorochemicals to granular activated carbon in the presence of ultrasound. J. Phys. Chem. 115 (11), 2250–2257.
Zhao, P., Xia, X., Dong, J., Xia, N., Jiang, X., Li, Y., Zhu, Y., 2016. Short- and long-chain perfluoroalkyl substances in the water, suspended particulate matter, and surface sediment of a turbid river. Sci. Total Environ. 568, 57–65.