Machine learning-based modeling and analysis of PFOS removal from contaminated water by nanofiltration process

Ahmad Hosseinzadeh[a], John L. Zhou[a*], Javad Zyaie[b], Nahawand AlZainati[a], Ibrar Ibrar[a], Ali Altaee[a]

[a] Centre for Green Technology, School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia

[b] Chemical Engineering Department, Iran University of Science and Technology, Tehran, Iran

Corresponding author:

Prof John Zhou, email: junliang.zhou@uts.edu.au

Abstract

Per- and polyfluoroalkyl substances (PFASs) are hazardous chemicals that have been widely used in different industries and released into the environment through polluted effluents. Nanofiltration (NF) membrane is regarded as a promising process for removing PFAS from the effluents. This study aimed to model and analyze the performance of the NF membrane process in perfluorooctanesulfonic acid (PFOS) removal from polluted effluents using machine learning (ML) algorithms. The modelling output of seven ML algorithms was evaluated using statistical indexes of determination coefficient (R^2) and mean squared error (MSE) for robustness. The results demonstrated that random forest (RF), gradient boosting machine (GBM), and AdaBoost models were the most robust ones for the NF process. Accordingly, the optimization of these procedures was accomplished using a grid search. The optimized models were deeply analyzed using permutation variable importance (PVI) to quantify the relative importance of variables. The three ML procedures (RF, GBM, AdaBoost) presented high prediction strength for PFOS removal from polluted effluents with low MSE values (4.726, 2.45, 2.879) and high R^2 values (0.93, 0.975 and 0.968) respectively. In addition, PVI-RF showed decreasing importance of pressure, PFOS initial concentrations, membrane type, trivalent cations, pH, divalent cations and monovalent cations consecutively.

Keywords: Machine learning; Nanofiltration membrane; PFOS; Process modelling

1. Introduction

Per- and polyfluoroalkyl substances (PFASs) are extensively used chemicals in a wide range of industries, e.g. food packaging, paints, fire retardants, lubricant production, surfactants, metal coating and production of waterproof substances owing to particular characteristics, e.g. low friction coefficients, thermal chemical stability and low carbon-fluorine bond polarizabilities [1]. However, PFASs are biologically toxic and environmentally persistent, posing a major health risk to the public and wildlife. In 2009, perfluorooctane sulfonate (PFOS) was included in the persistent organic pollutants (POPs) list under Stockholm Convention as the first representative of PFASs. Sun et al. reported that approximately 45250 tons of PFOS have been indirectly and directly released into the environment since 1970 [2]. Too many detrimental effects of PFOSs on human health, e.g., endocrine distribution, hepatotoxicity, immunotoxicity, developmental toxicity and epigenetic changes, have been recognized [2]. Trudel et al. reported that the intake of PFOS and perfluorooctanoic acid (PFOA) received by European and North American people is 1-130 and 3-220 ng per kg body weight per day, respectively. In addition, polluted water and food are the most important sources of these pollutants threatening human health [3]. Furthermore, water scarcity is another important global challenge [4-6]. Therefore, the purification of the contaminated water with PFAS is urgently needed to protect public health.

Conventional water treatments processes such as sedimentation, coagulation and chlorination are not significantly efficient in the PFASs removal from water [7, 8]. Even though other processes like UV-mediated photodegradation [9, 10] or sonochemical decomposition [11] showed better performance for the removal of the PFASs from water, these processes are still under investigation and require certain conditions, long operational period and high energy consumption [12], and sometimes with limited efficiency [3]. More importantly, it is impossible to completely degrade these pollutants due to different hazardous by-products generation [13]. At the same time, membrane filtration processes have been widely used for drinking water and

wastewater treatment by promising efficiency in a wide spectrum of pollutants rejection. Generally, nanofiltration (NF) membrane possesses higher rejection capability than the ultrafiltration and microfiltration membranes and higher water flux than reverse osmosis. Therefore, the performance of NF membranes as a successful process in the removal of a wide range of contaminants, especially PFOS, has been investigated [14-16]. The effectiveness of NF membrane in the PFOS removal from aqueous solutions is affected by several parameters such as solution pH, the type of membrane, the applied pressure, the initial PFOS concentration and the availability of various cations in the feed water. The solution pH can impact the charge of the solute coupled with the membrane surface charge which leads to differences in the total repulsive forces between the solute and the membrane surface. Stelinle-Darling and Reinhard [17] demonstrated that as the pH value increased from 2.8 to 6, the molecular weight cut-off that produced 90% rejection of the perfluorochemical compounds decreased from 550 g/mol to 300 g/mol. Moreover, the transmission rate of PFOS in the NF270 membrane was 2.5% when the pH was between 5 and 6, while it was jumped to 30% when the pH value was reduced to 2.8. As a result, the rejection rate of PFOS was reduced with the reduction in solution pH (from 6 to 2.8). The results could be due to the neutral membrane surface at pH 2.8, which causes a reduction in the impact of the repulsive forces at the membrane surface, which, in turn, allows the high transmission of PFOS through the membrane. Typically, the negatively charged NF membrane because less negative with decreasing the solution pH. As a results, the NF rejection rate to the negatively charged PFOS decreases.

The rejection rate is also affected by the phenomena of concentration polarization. As the concentration polarization increase due to the accumulation of pollutants in the fouling layer on the membrane surface, the rejection rate is reduced. Additionally, membrane type will affect the rejection rate of perfluorochemicals by NF. For instance, it has been noticed that 15 perfluorochemicals, including PFOS, were below the detection limit in the permeate stream when the NF200 membrane was used. In contrast, the average transmission rate for these

compounds was higher when more loose NF membranes (NF270 and DK) were used [17]. Therefore, the type of membrane is another key factor in this process. Furthermore, Tang et al. [18] observed that there is a direct relationship between the applied pressure and the rejection rate of PFOS. Moreover, the PFOS initial concentration increased its rejection rate by NF membrane [19]. Increasing the PFOS rejection rate with the feed concentration increase could be related to blocking the membrane pores [20], which reduces the water flux.

Regarding the applied pressure effects on NF membrane efficiency, as the differences between the surface and sweeping forces can be changed under various pressure affecting membrane performance in PFOS rejection rate, the applied pressure is another effective factor in this process [21]. Moreover, the temperature is also a major factor in the NF process. The higher the feed temperature, the larger the membrane pore size, affecting the rejection rate of the pollutant [22]. The other pivotal factor affecting this process is the presence of different ions, i.e. mono, di and trivalent ions, in the contaminated waters with PFOS, affecting the membrane filtration efficiency. For instance, divalent calcium ions can be found in various water sources at concentrations up to 0.007 M [23], which can impact the rejection rate of the PFOS, as observed by Zhao et al. [19]. The authors found that the feed water with 100 ppb of PFOS had a rejection rate increased from 94% to 99.3% due to the increase in Ca^{2+} ion concentration from 0 to 0.002 M. That would be attributed to the electrostatic interaction between PFOS and calcium ions forming large molecules that were highly rejected by the membrane. ~~Consequently, the availability of Ca^{2+} with PFOS in the wastewater stream will reduce the permeation rate through the membrane and the PFOS rejection~~. Additionally, Wang et. al [24] noticed similar behaviour as they proposed that the higher the concentration of Ca^{2+} in the feed water at 25 ˚C (from 0.0001M to 0.002 M), the higher the rejection rate of the PFOS from the water (from 97.1% to 99.4%). Generally, divalent cations as Ca^{2+} and Mg^{2+} are responsible for forming cation bridges with the negatively charged PFASs, enhancing their rejection rate on the membrane surface [25]. Differently, the presence of monovalent cations as Na^{+} has a reverse

impact on the adsorption of PFASs [26]. However, Zhao et al. [3] found that the presence of monovalent ions and divalent and trivalent cations in feed water increased the PFOS rejection rate by the NF membrane. However, the increase in the rejection rate was higher when divalent and trivalent cations (Ca^{2+} and Fe^{3+}) were used compared to monovalent cations (Na^{+}). The authors revealed that one sodium cation could be linked to one molecule of PFOS, while the divalent and trivalent cations preferred to link with two molecules of PFOS. The highest rejection rate of 97.94% was reached in the presence of Fe^{3+} cations.

Therefore, optimising the operation condition is crucial in successfully applying the NF process. There are two general procedures, i.e. experimental and numerical, for optimising such processes. Experimental procedures are essential to generate fundamental data for key parameters, but they are often limited by human resources, equipment and time [27]. One-factorial design experimental procedures have been extensively applied to optimise the NF process [28]. However, this optimization procedure cannot consider all individual and interactive effects of the independent factors on the process performance. It will be too time-consuming and costly to simultaneously consider all the parameters above [29, 30]. Compared to the experimental procedures, the numerical modelling procedures are faster, more cost-effective and highly complementary to experimental methods [27]. To the best of our knowledge, there is a lack of study to apply numerical procedures for assessing the performance of the NF process in PFOS removal. Furthermore, to our knowledge, no study has been reported on the NF membrane performance in PFOS removal by considering all of the effective parameters which are highly valuable for predesigning the process, either before or during experimental investigations. In addition, determining the relative importance of the parameters affecting this process can help to improve experimental design and optimize the process with fewer experiments.

Machine learning (ML) procedures are the most powerful tools to model complex processes and are more capable of learning the relationship between inputs and output in complex

processes with a high quantity of effective factors. To establish the relationship between the inputs and outputs, there is no necessity for ML approaches to comprehend the complex mechanisms of the phenomenon/process to model. Different algorithms have been developed to model processes; nonetheless, different algorithms demonstrate different performances in different applications [27, 29, 31-33]. To date, ML methods have not been used for modelling the NF membrane performance in PFOS removal from contaminated waters. Additionally, no study has systematically considered different ML procedures in NF membrane applications for PFOS removal to choose the most appropriate models for deep analysis and modelling intentions.

Therefore, the present work aims to use various ML approaches, including ridge regression (RR), linear regression (LR), multilayer perceptron (MLP), AdaBoost, random forest (RF), support vector regression (SVR) and gradient boosting machine (GBM) in PFOS removal from contaminated waters using NF membranes. The study investigated the impact of environmental and technical factors on removing PFOS by the NF. The independent variables, including the type of the NF membrane, operating temperature, PFOS initial concentration, pressure, pH, concentrations of monovalent, divalent and trivalent cations, were studied to choose the most appropriate approaches for this application were subsequently applied for deeply modelling and analysing process performance. Ultimately, the performances of the selected models are compared by the results, and the permutation variable importance (PVI) is used to determine the relative importance of the independent variables.

2. Materials and Methods

2.1. Data collection and processing

In the present work to develop appropriate models for NF process performance in PFOS rejection, the experimental results in the published literature were extracted after a detailed review [3, 14, 19, 20, 24, 34-36]. After careful selection, 290 data points were extracted using

Plot Digitizer. The considered inputs were membrane type, operating temperature, PFOS initial concentration, pH, pressure, divalent cations, monovalent cations and trivalent cations. Moreover, to simplify the complexity of the computation and avoid overfitting, the extracted experimental data were randomized in a range from 0 to 1 using Eq. (1) [37]:

$$\text{Normalized value } (X) = \frac{\text{x}_\text{i} - \text{minimum value of data}}{\text{maximum value of data} - \text{minimum value of data}} \times (1 - 0) + 0.1 \quad \text{(Eq. 1)}$$

where x_i is any data.

2.2. Selection of ML procedures and modelling generality

Different ML approaches demonstrate various performances in applications; hence, choosing the soundest approach to model different processes will be critical. Using the default hyperparameters, several ML approaches, i.e. MLP, RR, LR, AdaBoost, RF, SVM and GBM from *Scikit-learn* library, were prescreened to select the best ones for deep modelling. After that, the most suitable ML approaches were selected based on the obtained values of R^2 (Eq. 2) and mean squared error (MSE) (Eq. 3) for the models. To develop prescreening and deep models for the performance of the NF process in PFOS rejection, the dataset was randomly divided into 20% and 80% as train and test datasets, respectively. Moreover, cross-validation with 5-folds was applied to validate the models developed to prevent overfitting and wasting the data. Furthermore, the generalization strength of the models was assessed using the test dataset. A grid search approach was considered to tune the hyperparameters of the selected ML approaches. Eventually, the hyperparameters tuned were taken into account in modelling and testing the models. The evaluation of the model strengths was conducted based on the values of R^2 and MSE:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_{prd,i} - y_{Act,i})}{\sum_{i=1}^{N}(y_{prd,i} - y_m)} \quad \text{(Eq. 2)}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_{prd,i} - y_{Act,i}\right)^2 \quad \text{(Eq. 3)}$$

where $y_{Act,i}$ and $y_{prd,i}$ are the actual and forecasted proportions of PFOS rejection, respectively; N and y_m are the total numbers of data points and the mean of actual PFOS rejection. It should be stressed that the validation of the models developed was accomplished based on the mean of the R^2 and MSE during the modelling.

2.3. Gradient boosting machine

GBM is a classification and regression ML method that uses decision trees to generate a prediction model. The final prediction can be reached after combining the weak learners of the decision trees iteratively, generating a single strong learner. The GBM aims to find a function to predict the output value for a set of inputs where it minimizes the loss function. This can occur by adding new weak learners trained to reduce the loss function where the previous weak learners will not be changed [38, 39].

2.5. Random forests

RF is one of the known ML prediction/classification methods and is considered an extremely powerful algorithm that relies on bagging. RF is a combination of N number of varied trees predictions where each one represents the random vector of an independent sample. All the forest's trees have the same distribution, and the intensity of each tree in one forest of trees classifiers how they correlate to each other impact the generalization error of that forest. The RF determines the deviation of an individual response parameter by continuous and binary division of the data to gradually make it more homogenous, relating to single or more explanatory parameters. The data division is finished for the parameter that minimizes the errors. Each tree results in a predicted parameter (y_n) based on the same input parameters (x_{new}). The final prediction parameter (y_{new}) is the average of all individual predictions from all trees, as shown in Eq. 4 [40-42].

$$y_{new}(x_{new}) = \frac{1}{N}\sum_{n=1}^{N} y_n(x_{new}) \qquad \text{(Eq. 4)}$$

2.6. AdaBoost

AdaBoost is another ML statistical algorithm, which can collaborate with various algorithms or weak learners. In Ada Boost, the new weak learners will be adjusted to enhance the misclassified cases classified by the previous weak learners where the error of the predicted model is updated post to each stage. The weak learners are improved through the process and become more favourable than random guessing. Consequently, the model will reach the stage of a stronger learner. As the AdaBoost is utilized with weak learners (decision trees/classifiers), the collected information regards the hardness of the individual training samples will be introduced to the updated algorithm. Accordingly, the next weak learners will focus on the cases that are harder to categorize. The basic learners with small classification errors will have large weights, while those with a large classification rate will have small weights. The final predictive model resulting from AdaBoost is given by equation (5) [43, 44]:

$$f(x) = \sum_{n=1}^{N} \alpha_n G_n(x) \qquad \text{(Eq. 5)}$$

where N is the number of the weak learners, α_n is the parameter of the nth weak learner, and $G_n(x)$ is the nth weak learner.

2.8. Relative importance of variables

PVI is an approach by which inspection of any model fitted in the tabular data is applicable [27]. This approach randomly permutates the independent variables (inputs) in the prepared model and takes into account the errors in the output prediction; thus, the higher the error, the higher the importance of the variables. To compute the importance of the variables, the MSE was considered. Some of the advantages of this approach is counting both interactive and single effects of the inputs, a general approach and being quick and simple to apply. Therefore, this approach was applied for all three models used in the present study to indicate the relative

importance of variables affection PFOS rejection by NF membrane from polluted aqueous solution.

2.9. Comparison of model performance

Mean absolute error (MAE) (Eq. 6), MSE and R^2 as statistical indexes were applied to compare the good fitness of the constructed models, i.e. AdaBoost, RF and GBM in forecasting the PFOS rejection by NF process. It is worth mentioning that the mentioned statistical indexes were calculated using a test dataset.

$$MAE = 1 - \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \qquad \text{(Eq. 6)}$$

where n, x_i and y_i are the total number of data points, experimental and predicted proportions of the output consecutively.

3. Results and discussion

3.1. Selection of ML procedures

Different ML approaches performances in modelling PFOS rejection from contaminated waters by NF membrane were evaluated, with the outcomes listed in Table 1. According to the values of R^2 and MSE (Table 1) displaying the robustness of the models, RF, GBM and AdaBoost was chosen as the most appropriate ones. Moreover, different studies have recognized the great robustness of RF, GBM and AdaBoost in various modelling processes [45-47]. Thus, these three approaches were selected for deep modelling NF filtration in the present study.

Table 1. Various ML approaches performances in modelling PFOS removal from contaminated waters by NF membranes

	GBM	RF	AdaBoost	SVR	MLP	LR	RR
Total-Train R^2	0.98	0.98	0.95	0.23	0.30	0.51	0.50

Total-Test R^2	0.94	0.94	0.92	0.10	0.23	0.46	0.46
Train MSE	1.59	1.98	5.59	96.16	88.79	62.73	63.48
Test MSE	8.9	9.03	11.13	133.94	112.88	77.78	77.25

3.2. Random Forests

The hyperparameters were manipulated in a grid search to design the RF model. The best conditions of the number of the boosted trees, the best split features numbers, the least samples number in a leaf, and the least samples number in a split were considered and equal to 200, 7, 1 and 3, respectively. The R^2 values (0.977 and 0.936) and MSE values (2.852 and 8.299) were determined for training and cross-validation. Through the test phases, the R^2 value for the total train (train and cross-validation) was 0.984 and 0.930 for the test, while the corresponding MSE value was 2.305 for the total training and 4.726 for the test. As shown in Fig. 1, the PFOS rejection rates predicted by the RF model were close to their actual values, which suggests that the model can reliably predict the rejection rate of PFOS during the NF process.

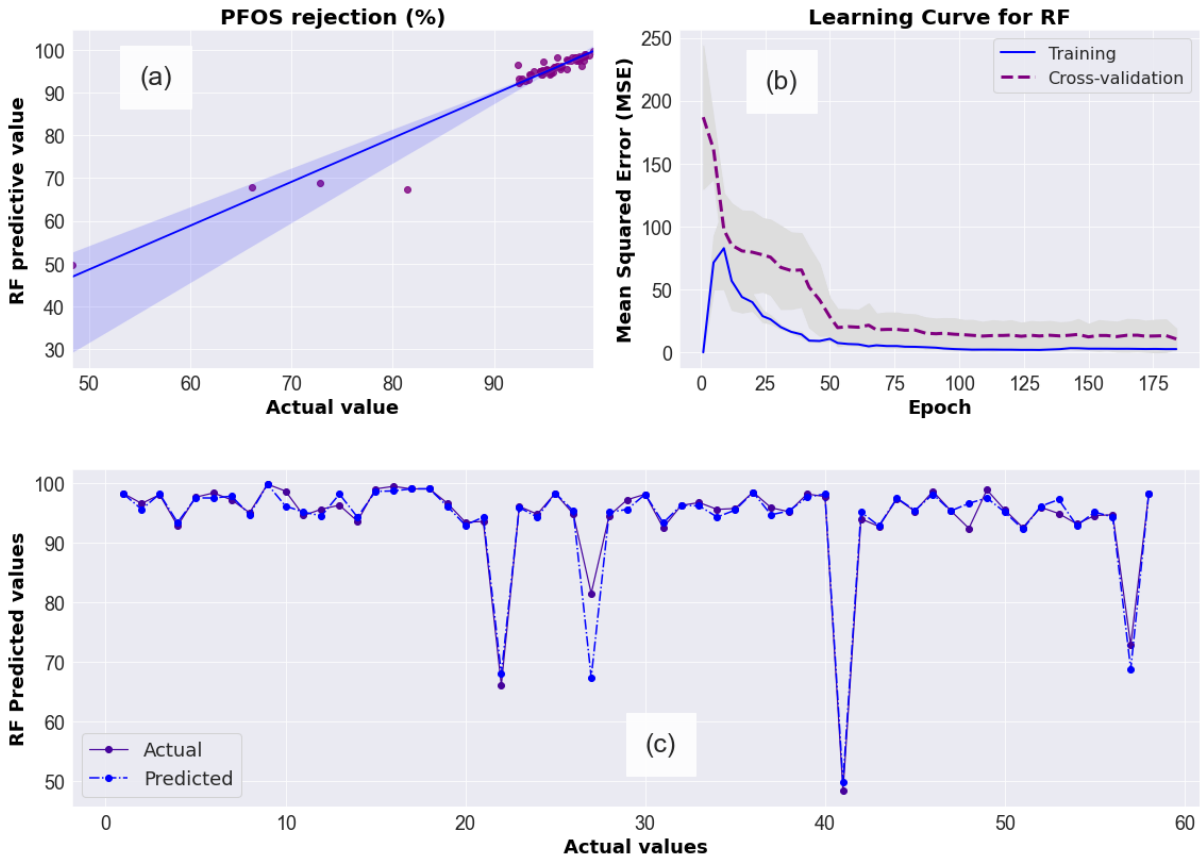

Fig. 1. RF model presentation as (a) scatter plots of the prepared model in the testing stage, (b) learning curve for the model prepared, and (c) the model accuracy in the testing stage.

Furthermore, potential overfitting, underfitting, and goodness of fitting are among the most key factors in modelling procedures that the learning curve can monitor. As observed in Fig. 1, the MSE trends in training and cross-validation phases experience stable conditions with a low difference from epoch ≥50. Therefore, the prepared model is not suffering from overfitting and underfitting issues.

3.2. Gradient boosting machine

The hyperparameters manipulated in a grid search to optimize the GBM model were a minimum number of samples per split, a minimum number of samples per leaf, several features for the best split, and a maximum depth of GBM and the number of gradients boosted trees. The best values of these parameters are 4, 4, 6, 5 and 800, respectively. Within these best parameters values, both training and cross-validation were taking place and the values of the R^2 (0.99 and 0.91) and the MSE (0.74 and 9.72) were specified. Moreover, the values of the R^2 under the test phases for the total train and test were 0.992 and 0.975, while the MSEs at these conditions were 0.992 and 2.45, respectively. Fig. 2 represents the GBM prediction performance in the test phase and clarifies the considerable prediction ability of the GBM model for the PFOS rejection rate from aqueous solutions by NF membranes.

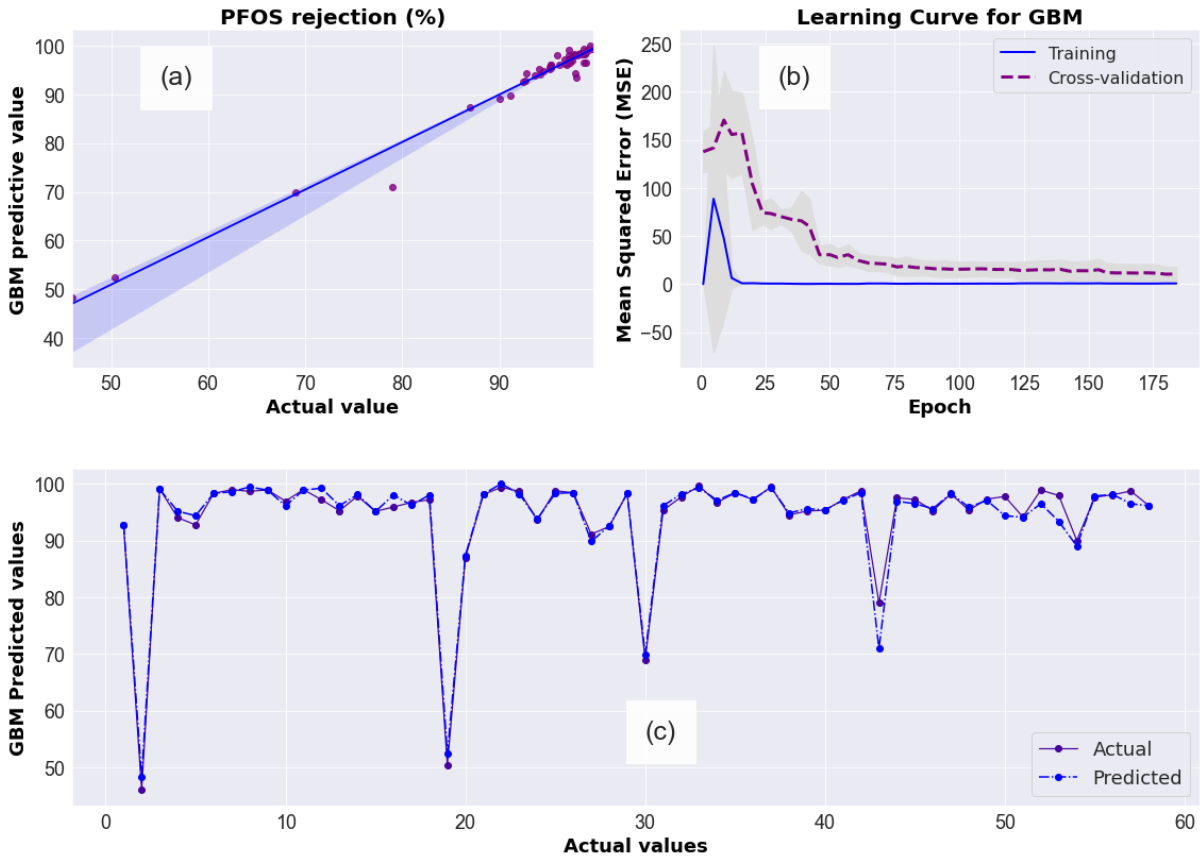

Fig. 2. GBM model presentation as (a) scatter plots of the prepared model in the testing stage, (b) learning curve for the model prepared, and (c) the model accuracy in the testing stage.

Furthermore, Fig. 2(b) shows the changing patterns of the MSE over different epochs of the prepared model during both training and cross-validation phases identifying that the prepared model deserves no overfitting and underfitting. The results indicate that at an epoch higher than 75, the MSE value is approximately 25 for both datasets with a smaller difference, emphasising the goodness of the model fitting.

3.3. AdaBoost

Various loss functions were tested with different hyperparameters values manipulated by a grid search to choose the best loos function. Table 3 shows different hyperparameters values, R^2 and MSE values at various phases; the square function was the best loss function with an n-estimator of 20 and a learning rate of 0.1. The R^2 was 0.945 for the training phase and 0.927 for the

validation phase. The MSE values were 7.955 and 8.194 for the training and validation phases. Additionally, the MSE values (7.241 and 2.879) and R^2 values (0.948 and 0.968) for the total training and testing phases demonstrated a prediction strength of 96.8% for this model. Moreover, the learning curve (Fig. 3(b)) verifies no overfitting and underfitting in this prepared mode.

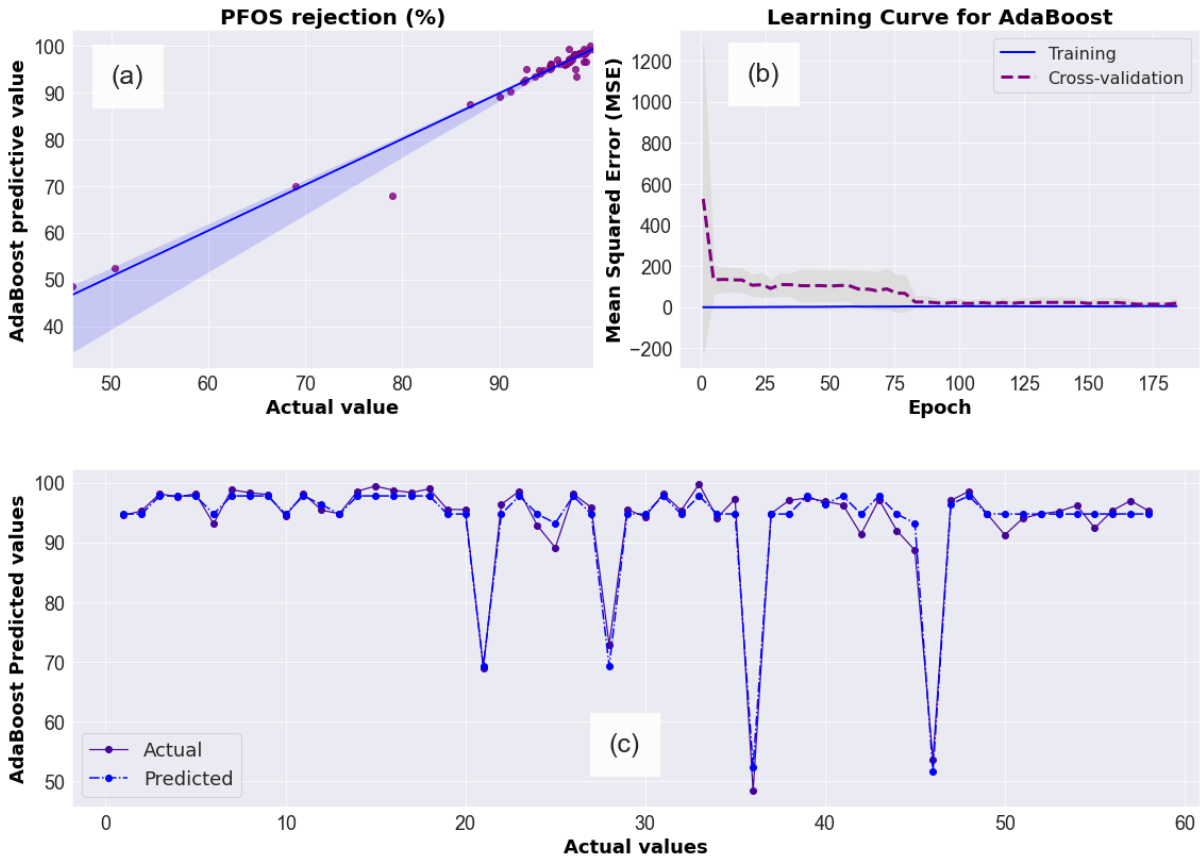

Fig. 3. AdaBoost model presentation as (a) scatter plots of the prepared model in the testing stage, (b) learning curve for the model prepared, and (c) the model accuracy in the testing stage.

Table 3. The manipulated condition and the results of the AdaBoost model under different loss functions.

	Grid search		R^2	R^2	R^2	R^2	MSE	MSE	MSE	MSE
	n-estimator	learning rate	Train	validation	Total-Train	Test	Train	validation	Total-Train	Test
Linear	120	1	0.945	0.885	0.889	0.888	8.005	12.956	0.945	0.942
Square	**20**	**0.1**	**0.945**	**0.927**	**0.948**	**0.968**	**7.955**	**8.194**	**7.241**	**2.879**
Exponential	80	0.1	0.951	0.913	0.953	0.967	7.187	9.845	6.574	2.981

3.4. Permutation variable importance

The relative importance of the inputs provided by the RF-PVI, GBM-PVI and AdaBoost-PVI are depicted in Fig. 4. The relative importance attained for the inputs differs from the different nature of the algorithms applied for modelling the process. Accordingly, the most important factors in RF-PVI are pressure, initial concentration of PFOS, membrane type, trivalent cations concentrations, pH, divalent cation concentrations, monovalent cation concentrations and temperature in decreasing order. GBM-PVI model has the following important factors order; temperature, membrane type, pH, monovalent cations concentrations, divalent cations concentrations, pressure, PFOS initial concentration, trivalent cations concentrations. In contrast, the order of the most important ones for the AdaBoost-PVI is divalent cations concentrations, membrane type, pressure, pH and PFOS initial concentration in decreasing order. The cations with one valence like Na^{+}, can react with only one PFOS molecule. The cations with higher valences can react with more PFOS molecules, resulting in larger compounds production and higher rejection efficiency. Besides, the more effects of Fe^{3+} as a trivalent cation than Ca^{2+} as a divalent cation can be attributed to the higher charge density and size of Fe^{3+} with 0.66 $Å^{-1}$ than Ca^{2+} with 0.48 $Å^{-1}$ [3, 34]. Yu et al. (2016) demonstrated that PFOS rejection is decreased approximately 15% with increasing 0.8 MPa pressure by HYDRA-CORE membrane [14]. Whilst, another study by Zhao et al. (2018) showed that the effect of the Fe^{3+} on the PFOS rejection was less than 2% with increasing 20 folds Fe^{3+} concentration from 0.1 mM to 2 mM [3]. Besides, the efficiency of the HYDRA-CORE membrane in PFOS rejection was increased approximately seven times more than NF270 one under the same condition [14]. In addition, in another experimental study, it has been reported that pH increase from almost 3 to 10 has led to roughly 5% more rejection [24]. Therefore, the PVI-RF has demonstrated more logical results than the others in this regard and is selected as the most proper procedure to indicate the relative importance of the inputs in this process.

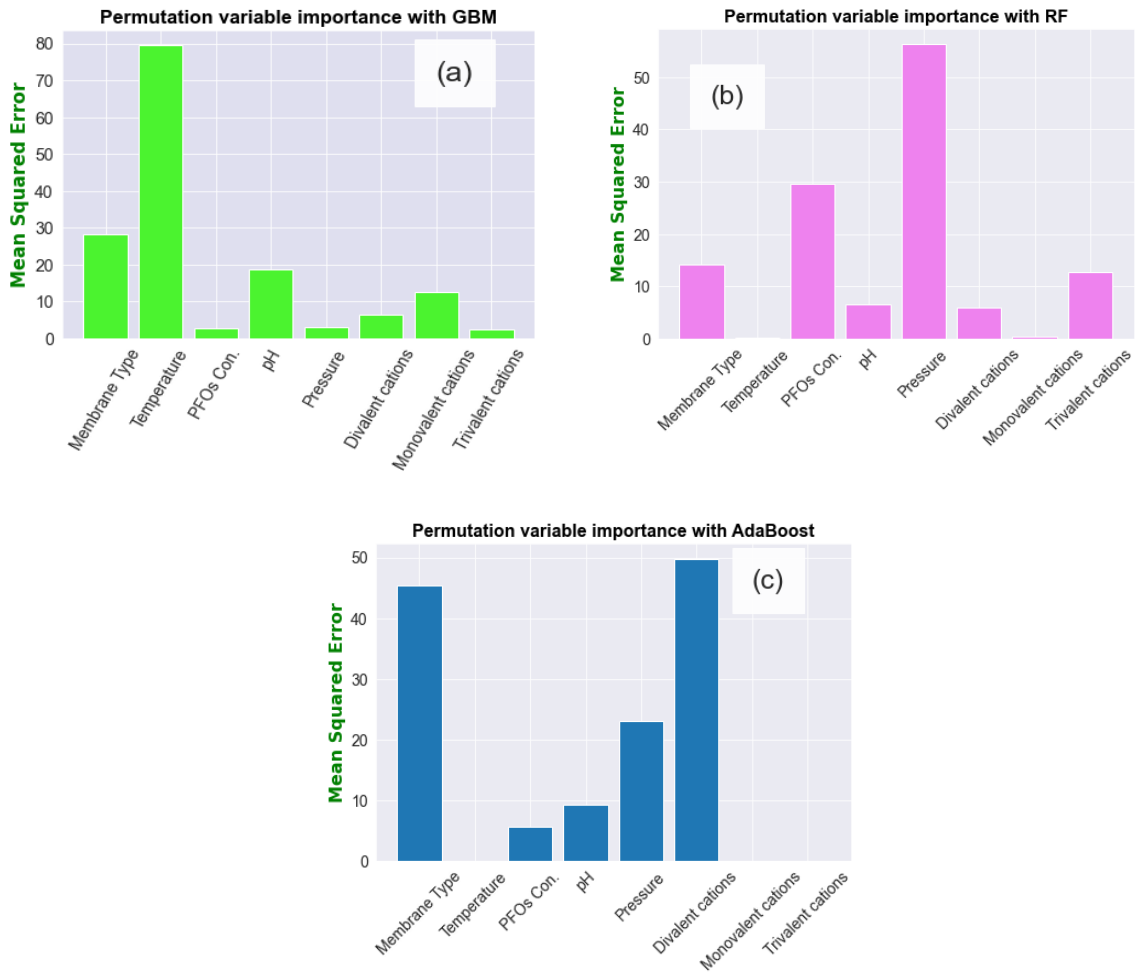

Fig. 4. The relative importance of the features by (a) GBM, (b) RF, and (c) AdaBoost models.

3.5. Model performance comparison

The models' performances in the PFOS rejection by NF membrane were evaluated by MAE, MSE and squared-R indexes presented in Table 3. As observed in R2, the results attained were very close together; however, the one for GBM was slightly better than the others, followed by AdaBoost and RF in decreasing order. In addition, both error indexes showed the same condition for these models. In a way that the lowest proportions of MAE and MSE belonged to the GBM; nonetheless, MEA for the RF model was rarely less than the AdaBoost one. In addition, the residual errors of these three models in the test phase are depicted in Fig. 5.

Ultimately, the GBM indicated better performance than AdaBoost and RF concerning all the mentioned indexes.

Table 3. Comparison of AdaBoost, RF and GBM models for modelling PFOS rejection from polluted aqueous solutions by NF membranes

Model	Statistical indices		
	MAE	MSE	R^2
RF	0.948	4.726	0.930
GBM	**0.861**	**2.45**	**0.975**
AdaBoost	1.165	2.879	0.968

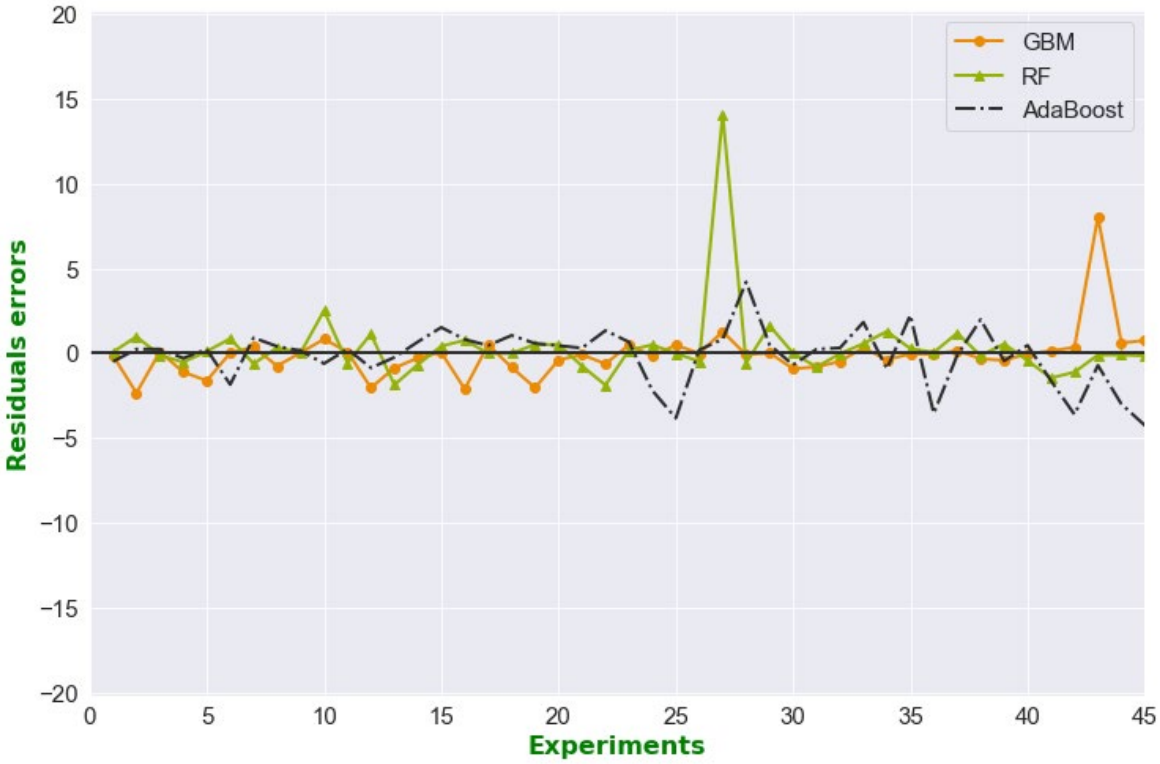

Fig. 5. The residual errors of the prepared AdaBooost, RF and GBM models for PFOS rejection by NF membrane from aqueous solutions.

4. Conclusions

NF membrane is a promising process applied for PFOS removal from contaminated water sources. This study has modelled and analyzed the NF membrane process performance in PFOS removal from polluted water using ML methods. Of various prescreened ML algorithms, RF, GBM and AdaBoost demonstrated the best potential for modelling and analyzing the performance of the NF process in PFOS removal with prediction strengths of 93%, 97.5% and 96.8%, respectively. In addition, the results of RF-PVI showed the increasing relative importance of the monovalent cations, divalent cations, pH, trivalent cations, membrane type, PFOS initial concentrations and pressure, as operating parameters in the NF process.

Acknowledgements

The authors would like to express their sincere appreciation for the University of Technology Sydney for a UTS President's scholarship and an International Research Scholarship.

References

[1] P.-Y. Chen, B. Wang, S. Zhuang, Y. Chen, Y.-P. Wei, Polyacrylonitrile fiber functionalized with fluorous hyperbranched polyethylenimine for selective removal of perfluorooctane sulfonate (PFOS) in firefighting wastewaters, Colloids and Surfaces A: Physicochemical and Engineering Aspects 619 (2021) 126539.

[2] Z. Sun, C. Zhang, J. Jiang, J. Wen, Q. Zhou, M.R. Hoffmann, UV/FeIINTA as a novel photoreductive system for the degradation of perfluorooctane sulfonate (PFOS) via a photoinduced intramolecular electron transfer mechanism, Chemical Engineering Journal 427 (2022) 130923.

[3] C. Zhao, G. Hu, D. Hou, L. Yu, Y. Zhao, J. Wang, A. Cao, Y. Zhai, Study on the effects of cations and anions on the removal of perfluorooctane sulphonate by nanofiltration membrane, Separation and Purification Technology 202 (2018) 385-396.

[4] A. Hosseinzadeh, J.L. Zhou, A.H. Navidpour, A. Altaee, Progress in osmotic membrane bioreactors research: Contaminant removal, microbial community and bioenergy production in wastewater, Bioresource Technology (2021) 124998.

[5] M. Afsari, H.K. Shon, L.D. Tijing, Janus membranes for membrane distillation: Recent advances and challenges, Advances in Colloid and Interface Science 289 (2021) 102362.

[6] A. Hosseinzadeh, J.L. Zhou, X. Li, M. Afsari, A. Altaee, Techno-economic and environmental impact assessment of hydrogen production processes using bio-waste as renewable energy resource, Renewable and Sustainable Energy Reviews 156 (2022) 111991.

[7] J. Li, Q. Li, L.-s. Li, L. Xu, Removal of perfluorooctanoic acid from water with economical mesoporous melamine-formaldehyde resin microsphere, Chemical Engineering Journal 320 (2017) 501-509.

[8] O. Quiñones, S.A. Snyder, Occurrence of perfluoroalkyl carboxylates and sulfonates in drinking water utilities and related waters from the United States, Environmental science & technology 43 (2009) 9089-9095.

[9] S. Wang, Q. Yang, F. Chen, J. Sun, K. Luo, F. Yao, X. Wang, D. Wang, X. Li, G. Zeng, Photocatalytic degradation of perfluorooctanoic acid and perfluorooctane sulfonate in water: A critical review, Chemical Engineering Journal 328 (2017) 927-942.

[10] L. Jin, P. Zhang, Photochemical decomposition of perfluorooctane sulfonate (PFOS) in an anoxic alkaline solution by 185nm vacuum ultraviolet, Chemical Engineering Journal 280 (2015) 241-247.

[11] J. Cheng, C.D. Vecitis, H. Park, B.T. Mader, M.R. Hoffmann, Sonochemical degradation of perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) in landfill groundwater: environmental matrix effects, Environmental science & technology 42 (2008) 8057-8063.

[12] Y. Zhou, Z. He, Y. Tao, Y. Xiao, T. Zhou, T. Jing, Y. Zhou, S. Mei, Preparation of a functional silica membrane coated on Fe3O4 nanoparticle for rapid and selective removal of perfluorinated compounds from surface water sample, Chemical Engineering Journal 303 (2016) 156-166.

[13] L. Yang, L. He, J. Xue, Y. Ma, Z. Xie, L. Wu, M. Huang, Z. Zhang, Persulfate-based degradation of perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS) in aqueous solution: Review on influences, mechanisms and prospective, Journal of Hazardous Materials 393 (2020) 122405.

[14] Y. Yu, C. Zhao, L. Yu, P. Li, T. Wang, Y. Xu, Removal of perfluorooctane sulfonates from water by a hybrid coagulation–nanofiltration process, Chemical Engineering Journal 289 (2016) 7-16.

[15] T.D. Appleman, E.R.V. Dickenson, C. Bellona, C.P. Higgins, Nanofiltration and granular activated carbon treatment of perfluoroalkyl acids, Journal of Hazardous Materials 260 (2013) 740-746.

[16] C. Bellona, D. Heil, C. Yu, P. Fu, J.E. Drewes, The pros and cons of using nanofiltration in lieu of reverse osmosis for indirect potable reuse applications, Separation and Purification Technology 85 (2012) 69-76.

[17] E. Steinle-Darling, M. Reinhard, Nanofiltration for trace organic contaminant removal: structure, solution, and membrane fouling effects on the rejection of perfluorochemicals, Environmental science & technology 42 (2008) 5292-5297.

[18] C.Y. Tang, Q.S. Fu, C.S. Criddle, J.O. Leckie, Effect of flux (transmembrane pressure) and membrane properties on fouling and rejection of reverse osmosis and nanofiltration

membranes treating perfluorooctane sulfonate containing wastewater, Environmental science & technology 41 (2007) 2008-2014.

[19] C. Zhao, J. Zhang, G. He, T. Wang, D. Hou, Z. Luan, Perfluorooctane sulfonate removal by nanofiltration membrane the role of calcium ions, Chemical Engineering Journal 233 (2013) 224-232.

[20] J. Wang, L. Wang, C. Xu, R. Zhi, R. Miao, T. Liang, X. Yue, Y. Lv, T. Liu, Perfluorooctane sulfonate and perfluorobutane sulfonate removal from water by nanofiltration membrane: The roles of solute concentration, ionic strength, and macromolecular organic foulants, Chemical Engineering Journal 332 (2018) 787-797.

[21] N.-E. Belkhouche, M.A. Didi, S. Taha, N.B. Farès, Zinc rejection from leachate solutions of industrial solid waste — effects of pressure and concentration on nanofiltration membrane performance, Desalination 239 (2009) 58-65.

[22] R. Xu, M. Zhou, H. Wang, X. Wang, X. Wen, Influences of temperature on the retention of PPCPs by nanofiltration membranes: Experiments and modeling assessment, Journal of Membrane Science 599 (2020) 117817.

[23] J.A. Baig, T.G. Kazi, M.B. Arain, H.I. Afridi, G.A. Kandhro, R.A. Sarfraz, M.K. Jamal, A.Q. Shah, Evaluation of arsenic and other physico-chemical parameters of surface and ground water of Jamshoro, Pakistan, Journal of Hazardous Materials 166 (2009) 662-669.

[24] T. Wang, C. Zhao, P. Li, Y. Li, J. Wang, Fabrication of novel poly(m-phenylene isophthalamide) hollow fiber nanofiltration membrane for effective removal of trace amount perfluorooctane sulfonate from water, Journal of Membrane Science 477 (2015) 74-85.

[25] C. Zhang, H. Yan, F. Li, X. Hu, Q. Zhou, Sorption of short- and long-chain perfluoroalkyl surfactants on sewage sludges, Journal of Hazardous Materials 260 (2013) 689-699.

[26] S.P. Lenka, M. Kah, L.P. Padhye, A review of the occurrence, transformation, and removal of poly- and perfluoroalkyl substances (PFAS) in wastewater treatment plants, Water Research 199 (2021) 117187.

[27] A. Hosseinzadeh, J.L. Zhou, A. Altaee, D. Li, Machine learning modeling and analysis of biohydrogen production from wastewater by dark fermentation process, Bioresource Technology 343 (2022) 126111.

[28] G. Zeng, Y. He, Y. Zhan, L. Zhang, Y. Pan, C. Zhang, Z. Yu, Novel polyvinylidene fluoride nanofiltration membrane blended with functionalized halloysite nanotubes for dye and heavy metal ions removal, Journal of Hazardous Materials 317 (2016) 60-72.

[29] A. Hosseinzadeh, A.A. Najafpoor, A.J. Jafari, R.K. Jazani, M. Baziar, H. Bargozin, F.G. Piranloo, Application of response surface methodology and artificial neural network modeling to assess non-thermal plasma efficiency in simultaneous removal of BTEX from waste gases: Effect of operating parameters and prediction performance, Process Safety and Environmental Protection 119 (2018) 261-270.

[30] A.A. Najafpoor, A.J. Jafari, A. Hosseinzadeh, R.K. Jazani, H. Bargozin, Optimization of non-thermal plasma efficiency in the simultaneous elimination of benzene, toluene, ethyl-benzene, and xylene from polluted airstreams using response surface methodology, Environmental Science and Pollution Research 25 (2018) 233-241.

[31] A. Hosseinzadeh, M. Baziar, H. Alidadi, J.L. Zhou, A. Altaee, A.A. Najafpoor, S. Jafarpour, Application of artificial neural network and multiple linear regression in modeling nutrient recovery in vermicompost under different conditions, Bioresource Technology 303 (2020) 122926.

[32] A. Hosseinzadeh, J.L. Zhou, A. Altaee, M. Baziar, D. Li, Effective modelling of hydrogen and energy recovery in microbial electrolysis cell by artificial neural network and adaptive network-based fuzzy inference system, Bioresource Technology 316 (2020) 123967.

[33] A. Hosseinzadeh, J.L. Zhou, A. Altaee, M. Baziar, X. Li, Modeling water flux in osmotic membrane bioreactor by adaptive network-based fuzzy inference system and artificial neural network, Bioresource technology 310 (2020) 123391.

[34] C. Zhao, T. Zhang, G. Hu, J. Ma, R. Song, J. Li, Efficient removal of perfluorooctane sulphonate by nanofiltration: Insights into the effect and mechanism of coexisting inorganic ions and humic acid, Journal of Membrane Science 610 (2020) 118176.

[35] H. Toure, A. Anwar Sadmani, Nanofiltration of perfluorooctanoic acid and perfluorooctane sulfonic acid as a function of water matrix properties, Water Supply 19 (2019) 2199-2205.

[36] C. Zhao, C.Y. Tang, P. Li, P. Adrian, G. Hu, Perfluorooctane sulfonate removal by nanofiltration membrane—the effect and interaction of magnesium ion/humic acid, Journal of Membrane Science 503 (2016) 31-41.

[37] H. You, X. Zhang, Sustainable livelihoods and rural sustainability in China: Ecologically secure, economically efficient or socially equitable?, Resources, Conservation and Recycling 120 (2017) 1-13.

[38] J. Zhou, E. Li, S. Yang, M. Wang, X. Shi, S. Yao, H.S. Mitri, Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories, Safety Science 118 (2019) 505-518.

[39] H. Zulfiqar, S.-S. Yuan, Q.-L. Huang, Z.-J. Sun, F.-Y. Dao, X.-L. Yu, H. Lin, Identification of cyclin protein using gradient boost decision tree algorithm, Computational and Structural Biotechnology Journal 19 (2021) 4123-4131.

[40] S. Schorr, M. Möller, J. Heib, D. Bähre, Quality Prediction of Drilled and Reamed Bores Based on Torque Measurements and the Machine Learning Method of Random Forest, Procedia Manufacturing 48 (2020) 894-901.

[41] S. Oppel, A. Meirinho, I. Ramírez, B. Gardner, A.F. O'Connell, P.I. Miller, M. Louzao, Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds, Biological Conservation 156 (2012) 94-104.

[42] L. Zhang, F. Huettmann, S. Liu, P. Sun, Z. Yu, X. Zhang, C. Mi, Classification and regression with random forests as a standard method for presence-only data SDMs: A future conservation example using China tree species, Ecological Informatics 52 (2019) 46-56.

[43] L. Li, C. Wang, W. Li, J. Chen, Hyperspectral image classification by AdaBoost weighted composite kernel extreme learning machines, Neurocomputing 275 (2018) 1725-1733.

[44] J.-Q. Chen, H.-Y. Chen, W.-j. Dai, Q.-J. Lv, C.Y.-C. Chen, Artificial intelligence approach to find lead compounds for treating tumors, The journal of physical chemistry letters 10 (2019) 4382-4400.

[45] X. Zhu, Z. Wan, D.C.W. Tsang, M. He, D. Hou, Z. Su, J. Shang, Machine learning for the selection of carbon-based materials for tetracycline and sulfamethoxazole adsorption, Chemical Engineering Journal 406 (2021) 126782.

[46] C.S.H. Yeo, Q. Xie, X. Wang, S. Zhang, Understanding and optimization of thin film nanocomposite membranes for reverse osmosis with machine learning, Journal of Membrane Science 606 (2020) 118135.

[47] Q. Tan, W. Li, X. Chen, Identification the source of fecal contamination for geographically unassociated samples with a statistical classification model based on support vector machine, Journal of Hazardous Materials 407 (2021) 124821.