

# Machine Learning and Neural Networks for Modelling the Retention of PPhACs by NF/RO

Y. Ammi,\* C. Si-Moussa, and S. Hanini

Laboratory of Biomaterials and Transport Phenomena (LBMP), University of Médéa,  
26 000 Médéa, Algeria

This work is licensed under a  
Creative Commons Attribution 4.0  
International License



## Abstract

The retention of polar pharmaceutical active compounds (PPhACs) by nanofiltration and reverse osmosis (NF/RO) membranes is of paramount importance in membrane separation processes. The retention of 21 PPhACs was correlated using artificial intelligence techniques: multi-layer perceptron (MLP), feedforward neural network with radial basis function (RBF), and support vector machine (SVM). A database of 541 retention values has been collected from the literature. The results showed a high predictive capacity of the MLP model for the retention of PPhACs by NF/RO with a very high correlation coefficient ( $R = 0.9714$ ) and a very low root mean squared error (RMSE = 3.9139 %) for the entire data set. The comparison between the three models showed the superiority of the MLP model. The sensitivity analysis emphasised that the retention of PPhACs is governed by three interactions arranged in descending order: polarity interactions (hydrophobicity/hydrophilicity), electrostatic repulsion, and steric hindrance. This research suggests that the PPhACs retention on the NF/RO membrane strongly depends on the topological polar surface area.

## Keywords

Machine learning, neural networks, modelling, retention, PPhACs, nanofiltration, reverse osmosis

## 1 Introduction

The presence of organic compounds in natural water sources is regarded as a high-priority environmental issue. In particular, polar pharmaceutical active compounds (PPhACs) can be highly mobile in water, potentially reaching source waters and even treated drinking water. The PPhACs are highly polar chemicals that tend to accumulate in nanofiltration and reverse osmosis membranes due to their low degradation properties and slow adsorption kinetics in water.<sup>1,2</sup>

Pharmaceutical active compounds can be removed using advanced technologies. Membrane processes such as reverse osmosis (RO) and nanofiltration (NF) are excellent technological solutions for removing pharmaceutical active compounds and protecting both humans and the environment. In fact, several studies have demonstrated that nanofiltration and reverse osmosis (NF/RO) processes can effectively remove pharmaceutical active compounds. Globally, studies have assessed retention efficiency with complex interactions between solutes and membranes, including steric hindrance, electrostatic repulsion, and hydrophobic-adsorption interactions. The interactions between solutes and membranes are, in turn, influenced by the characteristics of the compounds (including molecular size, polarity, charge, and hydrophobicity), the properties of the membrane (porosity, polarity, and electrostatic charges), and the conditions under the which filtration takes place (such as pH, pressure, permeate flux, temperature, membrane fouling, recovery, and cross-flow velocity).<sup>2-4</sup>

Modelling the retention of organic compounds by NF/RO membranes is a very important tool for developing robust

high-pressure membrane technologies. However, there have been fewer models available to accurately predict the retention of organic compounds due to the complexity of the underlying mechanisms.<sup>5-9</sup> These research studies were conducted to investigate the utilisation of various modelling approaches, such as multiple linear regression (MLR), artificial neural network (ANN), bootstrap aggregated neural network (BANN), partial least squares (PLS), and support vector machines (SVM), based on quantitative structure-activity relationship (QSAR) to establish correlations, models, and predictions for the retention of organic compounds (neutral and ionic) by NF/RO membranes. There is no modelling study conducted on the retention of organic compounds by NF/RO membranes using feedforward neural networks with radial basis function (RBF).

In this study, we attempted to estimate the retention of the PPhACs by NF/RO membranes using three artificial intelligence approaches: multi-layer perceptron (MLP), RBF, and SVM. These three approaches will consider a similar set of inputs, including the effective diameter of the organic compound in water  $d_c$ ,  $\log D_{ow}$ , dipole moment, molecular length, molecular equivalent width, molecular weight cut-off, sodium chloride salt rejection, zeta potential, contact angle, pH, pressure, temperature, and recovery.

## 2 Prediction of PPhACs retention using hybrid learning architecture

### 2.1 Feedforward neural networks

The feedforward neural network (FNN) is presented graphically by a set of connected neurons (Fig. 1). Information flows in one direction only, from the input layer to the output layer without turning back. It can be seen that the FNN

\* Corresponding author: Yamina Ammi, PhD  
Email: [ammi.yamina@univ-medea.dz](mailto:ammi.yamina@univ-medea.dz)

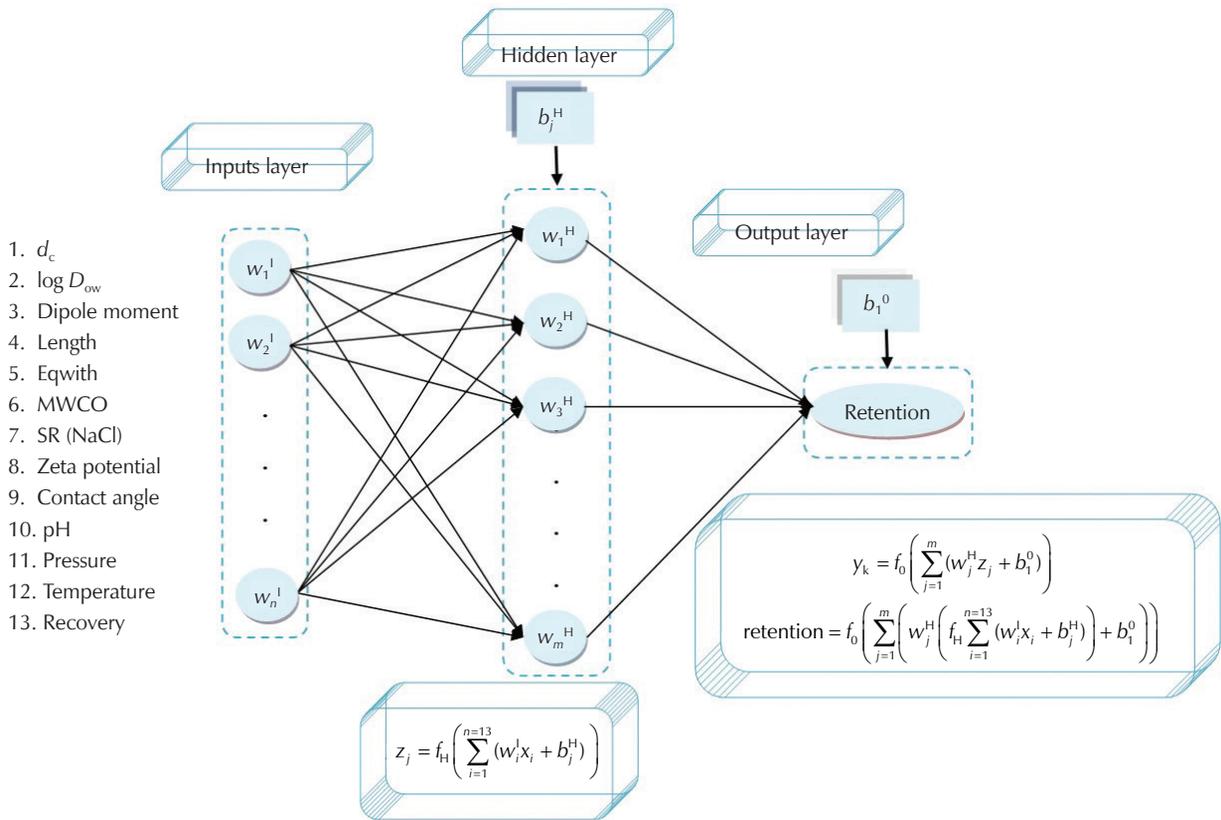


Fig. 1 – Structure of feedforward neural network:  $i, j$ , and  $k$  are numbers of neurons in the input, hidden, and output layers, respectively;  $x_i$  is the input to the  $i^{\text{th}}$  neuron of the input layer;  $w_i^I$  and  $w_j^H$  are the weights of input and hidden layers, respectively;  $b_j^H$  and  $b_1^0$  are the bias of the hidden and output layers, respectively; and  $f_H$  and  $f_0$  are the transfer functions in the hidden and output layer, respectively

is acyclic and structured in such a way that neurons within the same layer are not connected. Each layer receives signals from the previous layer and transmits the results of its processing to the next layer. The two outermost layers correspond to the input layer, which receives its inputs from the external environment, and the output layer, which provides the results of the performed treatments. The intermediate layers, known as hidden layers, can vary in number.<sup>10</sup>

MLP and RBF are the most commonly used FFNs. MLP can solve nonlinearly separable and complicated logic problems.<sup>11</sup> It requires supervised training.<sup>12</sup> RBF networks were proposed by Moody and Darken. RBF can be used in similar types of problems as MLP, such as classification and function approximation.

### 2.2 Support vector machine

Similar to the neural network, the primary concept underlying SVM is that neurons are organised in two layers (Fig. 2). However, unlike FFNs, SVM is based on statistical learning theory, which implements structural risk minimisation theory and various error optimisation techniques.<sup>12</sup>

The nonlinear relationship between inputs and outputs in SVM model can be utilised by the regression function. The outputs of the SVM model are obtained using the following equation:<sup>12</sup>

$$f(x_i) = \omega^T \varphi(x_i) + b, \quad i = 1, 2, \dots, n \quad (1)$$

$f(x_i)$  is the predicted value of the SVM model,  $\varphi(x_i)$  represents the nonlinear function that maps input finite-dimensional space into a higher-dimensional space which is implicitly created,  $\omega$  is the weight vector of the SVM model to be optimised, and  $b$  represents the bias of the SVM model to be optimised.

The database has a  $D$ -dimensional input vector  $x_i \in R^D$  and a scalar output  $y_i \in R$ .

The SVM model for the training database is given as follows:

$$\min R(w, \xi, \xi^*, \varepsilon) = \frac{1}{2} w^2 + C \left[ v\varepsilon + \frac{1}{N} \sum_{i=1}^N (\xi_i + \xi_i^*) \right]$$

subjective to :

$$y_i - w^T \varphi(x_i) - b \leq \varepsilon + \xi_i \quad (2)$$

$$w^T \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i$$

$$\xi_i^*, \varepsilon \geq 0$$

$C$  is the parameter used to balance the empirical risk and model complexity term  $w^2$  and  $\xi_i^*$  represents the slack variable to denote the distance of the  $i^{\text{th}}$  sample outside of the  $\varepsilon$ -tube.

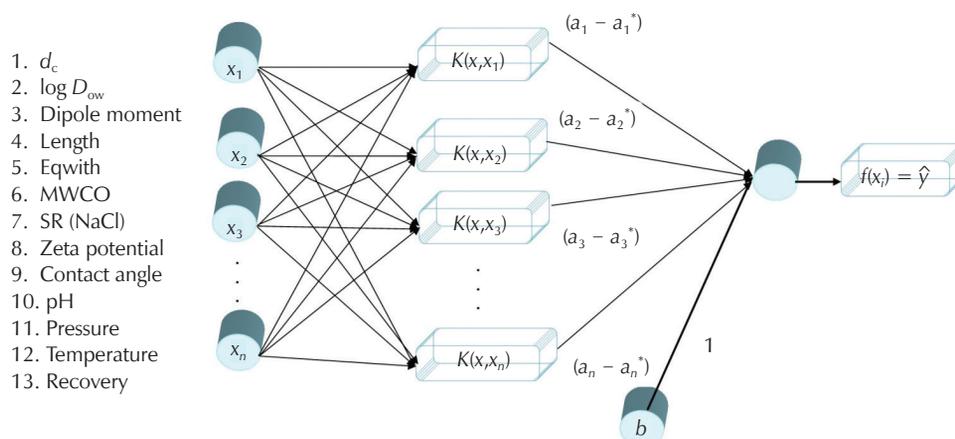


Fig. 2 – Sketch of the support vector machine model

Table 1 – Most important information on the data adjusted in this work<sup>17,26</sup>

Variable category	Factors	Unit	Domain	SD
Inputs	$d_c$	$\text{g mol}^{-1}$	[0.5855, 0.8899]	0.0859
	$\log D_{ow}$	–	[–9.6200, 3.7200]	2.3319
	dipole moment	D	[1.2200, 16.7300]	4.0622
	length	nm	[0.1500, 1.9970]	0.3946
	eqwidth <sup>17</sup>	nm	[0.1100, 1.1290]	0.1862
	MWCO	Da	[100.0000, 405.0000]	87.7269
	SR(NaCl)	–	[0.1210, 0.9950]	0.2432
	zeta potential	mV	[–77.8600, 3.1000]	17.4296
	contact angle	°	[15.1200, 80.0000]	15.5034
	pH	–	[3.0000, 9.0000]	1.1355
	pressure	kPa	[280.0000, 4100.0000]	570.1369
	recovery	%	[0.0300, 98.4800]	23.8393
	temperature	°C	[0.0000, 45.0000]	8.1209
Output	retention	%	[1.2900, 100.0000]	16.4552

$$\text{eqwidth} = \sqrt{\text{molecular width} \cdot \text{molecular depth}}$$

$$d_c = 0.065 \cdot \text{molecular weight}^{0.438}$$

### 2.3 Experimental data of polar pharmaceutical active compounds

Scientific literature<sup>6,7,13–25</sup> were reviewed to select 541 retention data on 21 PPhACs. The PPhACs were assigned to three categories based on their physicochemical properties: hydrophobicity ( $\log D_{ow}$ ), polarity (dipole moment), and size (effective diameter of PPhACs in water  $d_c$ , molecular length, and molecular equivalent width “eqwidth”). The retention mechanisms of PPhACs by NF/RO membranes are based on the sieving effect, electrostatic interactions, and hydrophobic/adsorption interactions between solutes and membranes. These solute–membrane interactions are

determined by properties of PPhACs, membrane characteristics (molecular weight cut-off “MWCO”, sodium chloride salt rejection “SR(NaCl)”, surface membrane charge “zeta potential” and membrane hydrophobicity “contact angle”), as well as filtration conditions (pH, pressure, temperature, and recovery). The first step in constructing a prediction model, such as MLP, RBF, and SVM, is to determine the input variables or factors. The input and output variables considered in this study and statistical analysis data are mentioned in Table 1. The statistical analysis of the input and output data was performed in terms of the domain studied and their standard deviation “SD”.

### 3 Modelling details

#### 3.1 Model development

Three models: MLP, RBF, and SVM, were implemented and evaluated for the prediction of the retention of PPhACs by NF/RO membranes. We developed all models using the STATISTICA software. Fig. 3 illustrates the procedure used to develop and optimise the architecture of the models. The samples were divided into two subsets: training and testing data set.

All MLP models tested in this study had one hidden layer. To create the optimal MLP models, we utilised the BFGS quasi-Newton (trainbfg) training algorithm, with four activation functions in the hidden layer: hyperbolic tangent sigmoid (tansig), log sigmoid (logsig), sine (sin), and exponential function, and one pure linear (purelin) activation function in the output layer. We varied the number of hidden neurons from 3 to 25 until the best model was obtained. A trial-and-error method was employed to obtain the optimal MLP models.

To construct the RBF model, we utilised the Gaussian radial basis activation function for the hidden layer and the linear activation function for the output layer. We adjusted the number of hidden neurons, ranging from 3 to 25, until the best model was obtained.

As stated previously in this article, the selection of the kernel functions is a crucial role in the performance of the

model. STATISTICA offers several kernel functions that can be utilised in SVM models. We calculated over the penalty term of the Gaussian radial basis function parameters with  $C = 10.0000$ ,  $nu = 0.5000$ , and  $Gamma = 3.51$ , for the SVM model and determined the optimal values for the target parameters.

#### 3.2 Evaluation criteria

In this paper, we used three measures to evaluate the quality of prediction models: correlation coefficient ( $R$ ), mean absolute error (MAE), and root mean squared error (RMSE).<sup>10</sup>

### 4 Results and discussion

In this study, we developed and evaluated nine models: three MLP, three RBF, and three SVM models, to predict the retention of PPhACs by NF/RO membranes; all models had similar inputs.

Fig. 4 illustrates the influence of the division of the data set into two subsets (the training and the testing set) on the coefficient of correlation. This was done to determine the optimal division of the data set for tested models (MLP, RBF, and SVM). We used three different divisions of the data set: division 1 (325 data for the training data set

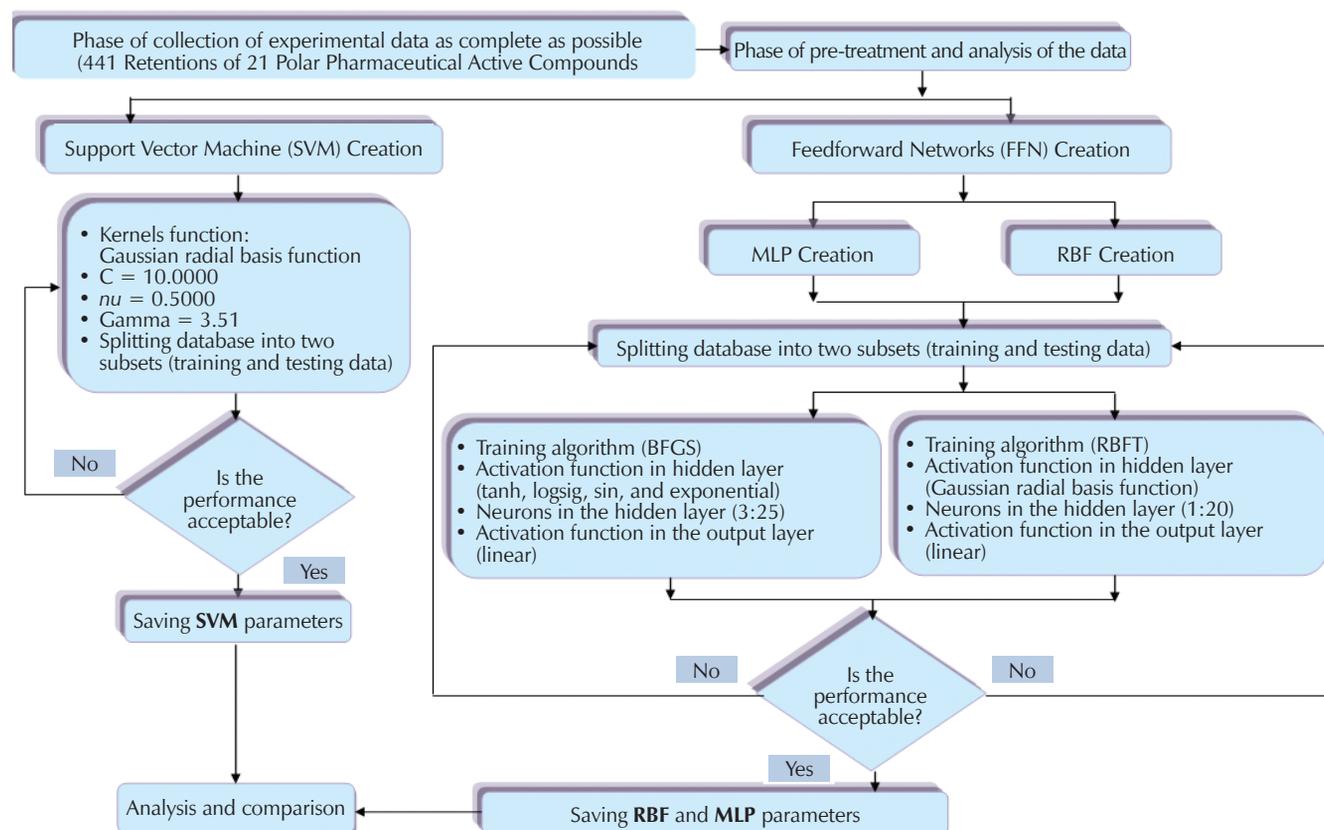


Fig. 3 – Flow diagram of the development of multi-layer perceptron (MLP), feedforward neural network with radial basis function (RBF), and support vector machine (SVM) models

(60 %), and 216 data for the testing data set (40 %)), division 2 (379 data for the training data set (70 %), and 162 data for the testing data set (30 %)), and division 3 (433 data for the training data set (80 %), and 108 data for the testing data set (20 %)). The results suggested that models using division 3 had better ability to predict the retention of PPhACs. MLP and SVM models showed high correlation coefficients, while those for the RBF model were lower, indicating that RBF models may not be as accurate in predicting the retention of PPhACs as the other two models.

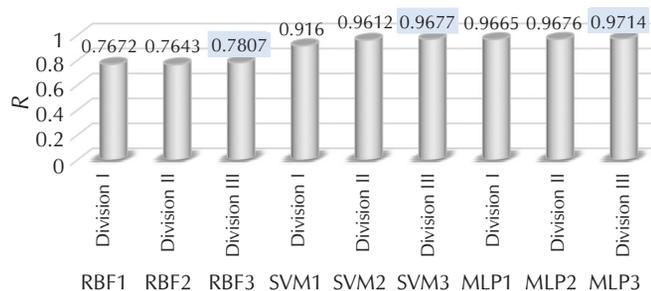


Fig. 4 – Effect of dividing the entire data set

The structures of the MLP models used are given in Table 2.

Table 2 – Structures of the optimised MLP models

	Training algorithm	Input layer	Hidden layer		Output layer	
		No. of neurons	No. of neurons	Activation function	No. of neurons	Activation function
MLP 1	trainbfg	13	9	tansig	1	purelin
MLP 2			10	tansig		
MLP 3			15	logsig		

Table 3 displays the results of RBF models comprising various numbers of hidden neurons. Among the training, the testing, and the entire data sets of the RBF 3 model, 15 hidden neurons gave the best *R*, MAE, and RMSE values (0.7901, 7.0832, and 0.1025 for the training data set and 0.7448, 8.2519, and 10.9558 for the testing data set, 0.7807, 7.3165, and 10.2785 for entire data set, respectively). Based on the performance statistics in the testing data set and the entire data set, the RBF 3 (13–15–1) model was selected as the best RBF model for predicting the retention of polar PPhACs.

Table 3 – Comparative performance of FNN-RBF models for various data sets

		RBF 1	RBF 2	RBF 3
Number of hidden neurons		20	16	15
Training data set	<i>R</i>	0.7935	0.7811	0.7901
	MAE/%	7.3671	7.2760	7.0832
	RMSE/%	10.0334	10.1888	10.1025
Testing data set	<i>R</i>	0.7272	0.7256	0.7448
	MAE/%	8.3071	8.6494	8.2519
	RMSE/%	11.3000	11.5152	10.9558
Entire data set	<i>R</i>	0.7672	0.7643	0.7807
	MAE/%	7.7424	7.6873	7.3165
	RMSE/%	10.5573	10.6034	10.2785

Table 4 summarises the accuracy of the SVM models for the training, the testing, and the entire data set. It is observed that in the various data sets, SVM 3 had the lowest MAE and RMSE, and the highest *R* (2.9686, 4.1598, and 0.9677, respectively). However, in the entire data set, SVM 1 got the highest MAE and RMSE and lowest *R* (3.3278, 6.6650, and 0.9160, respectively).

Table 4 – Comparative performance of SVM models for various data sets

		SVM 1	SVM 2	SVM 3
Number of support vectors		156 (30 bounded)	159 (22 bounded)	169 (21 bounded)
Training data set	<i>R</i>	0.9770	0.9759	0.9766
	MAE/%	2.7243	2.6525	2.6889
	RMSE/%	3.3866	3.4656	3.5399
Testing data set	<i>R</i>	0.8513	0.9373	0.9342
	MAE/%	5.8435	4.3916	4.0772
	RMSE/%	9.9717	6.4406	6.0183
Entire data set	<i>R</i>	0.9160	0.9612	0.9677
	MAE/%	3.3278	3.1765	2.9686
	RMSE/%	6.6650	4.5706	4.1598

The contribution of the input variables ( $d_c$ ,  $\log D_{ow}$ , length, eqwidth, *MWCO*, *SR*(NaCl), contact angle, zeta potential, pH, pressure, recovery, and temperature) on the output (retention of PPhACs) was determined by a sensitivity analysis for MLP 3 model.

The results on the relative importance (*RI*) and contribution analysis are presented in Fig. 5. Contact angle, zeta potential, *SR*(NaCl), *MWCO*,  $d_c$ , pressure, recovery, temperature, dipole moment, length,  $\log D_{ow}$ , eqwidth, and pH that may influence the retention of PPhACs. All input variables were

found to have a significant contribution ( $RI > 2$ ). Sensitivity analysis identified the importance of variables used for the modelling of the retention of PPhACs and confirmed the correctness of the variables selected in this study.

Fig. 5 shows that the membrane polarity interactions (hydrophobicity/hydrophilicity) measured by contact angle were even more important than surface membrane charge measured by zeta potential, steric hindrance measured by  $SR(NaCl)$ , and  $MWCO$ , with  $RI$  values of 75.68, 58.88, 54.73, and 32.25 %, respectively. These results confirm that the retention of PPhACs by NF/RO membranes depends on three interactions arranged in descending order: polarity interactions (hydrophobicity/hydrophilicity), electrostatic repulsion (charge effect), and steric hindrance (sieving effect). This research suggests that the PPhACs retention on NF/RO membranes strongly depends on the topological polar surface area.

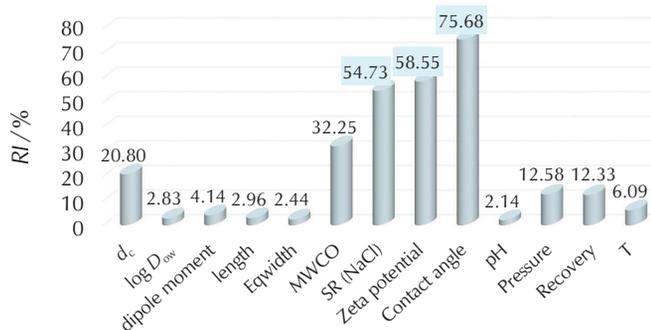


Fig. 5 – Plot of the relative importance ( $RI$ ) of the MLP 3 model for modelling the retention of PPhACs by NF/RO membranes

To assess the predictive ability of the MLP 3, RBF 3, and SVM 3 models, the matching between predicted and experimental retentions was analysed for the entire data set. The MATLAB function “postreg” was used to obtain the plot of the predicted versus experimental values and the parameters of the linear regression.

Fig. 6 shows the agreement plot between predicted and experimental values of PPhACs retention. The results indicated acceptable robustness of RBF model and the potential for prediction of different parameters that characterise the retention of PPhACs by NF/RO membranes. The values for SVM demonstrated the strong reliability of the model and the potential to predict accurately various parameters that characterise the retention of PPhACs. Fig. 6c demonstrates the predictive ability and accuracy of MLP model, revealing the ideal modelling of the entire data set by the optimised MLP.

## 5 Analysis and comparison of prediction models

We assessed the MLP model’s predictive performance for PPhACs retention by NF/RO membranes, and compared it with the performance of RBF and SVM models. We evaluated the accuracy of models using  $R$ , MAE, and RMSE. Table 5 compares performance of optimal models in terms of the errors ( $R$ , MAE, RMSE) for the entire data set. It is clear that the MLP model achieved the lowest errors in the entire data set (0.9714, 2.4410 %, 3.9139, respectively). These results demonstrate that the MLP model outperformed the RBF and SVM models. Our findings suggested that the MLP model might serve as a viable alternative to the RBF and SVM models for predicting the retention of PPhACs by NF/RO membranes.

Table 5 – Comparative performance of optimal models for the entire data set

	MLP	RBF	SVM
$R$	0.9714	0.7807	0.9677
MAE/%	2.4410	7.3165	2.9686
RMSE/%	3.9139	10.2785	4.1598

## 6 Applicability domain

In the present study, the Leverage mathematical technique was used to find outliers. This technique uses the residual values and a Hat matrix. To calculate the normalised residuals, we compare the experimental retention data with the values predicted by the model.

$$R\_Norm_i = \frac{\text{retention}_i^{\text{exp}} - \text{retention}_i^{\text{cal}}}{\sqrt{\text{var}(\text{retention}^{\text{exp}} - \text{retention}^{\text{cal}})}}, i = 1, \dots, m \quad (3)$$

Normalised residuals falling within the range of  $\pm 3$  are considered acceptable for validated data, while those outside this range are classified as suspected data. Furthermore, any data point with a  $Hat$  value greater than the threshold  $H^*$  is also flagged as suspected data. The results of the outlier analysis are presented in Fig. 7. As shown in the figure, the critical threshold for the  $Hat$  value is determined to be  $H^* = 0.0776$ . Only the data points presented by red circles are considered valid data, and these points confirm the accuracy of the model. The analysis of the outliers of the MLP 3 model using the *trainbr* and *log-sig* is shown in Fig. 7(a). The plot indicates that 513/541 data points (94.82 %) fall within the valid domain, while only 28/541 data points (5.18 %) are outside the applicable range of the optimised MLP 3 model. Fig. 7(b) shows the William’s plot for the optimal RBF 3 model. The plot reveals that 518/541 data points (95.75 %) are considered valid, while 23/541 data points (4.25 %) are classified as suspected. Fig. 7(c) presents the William’s plot for the optimised SVM model. The plot shows that out of the 541 data

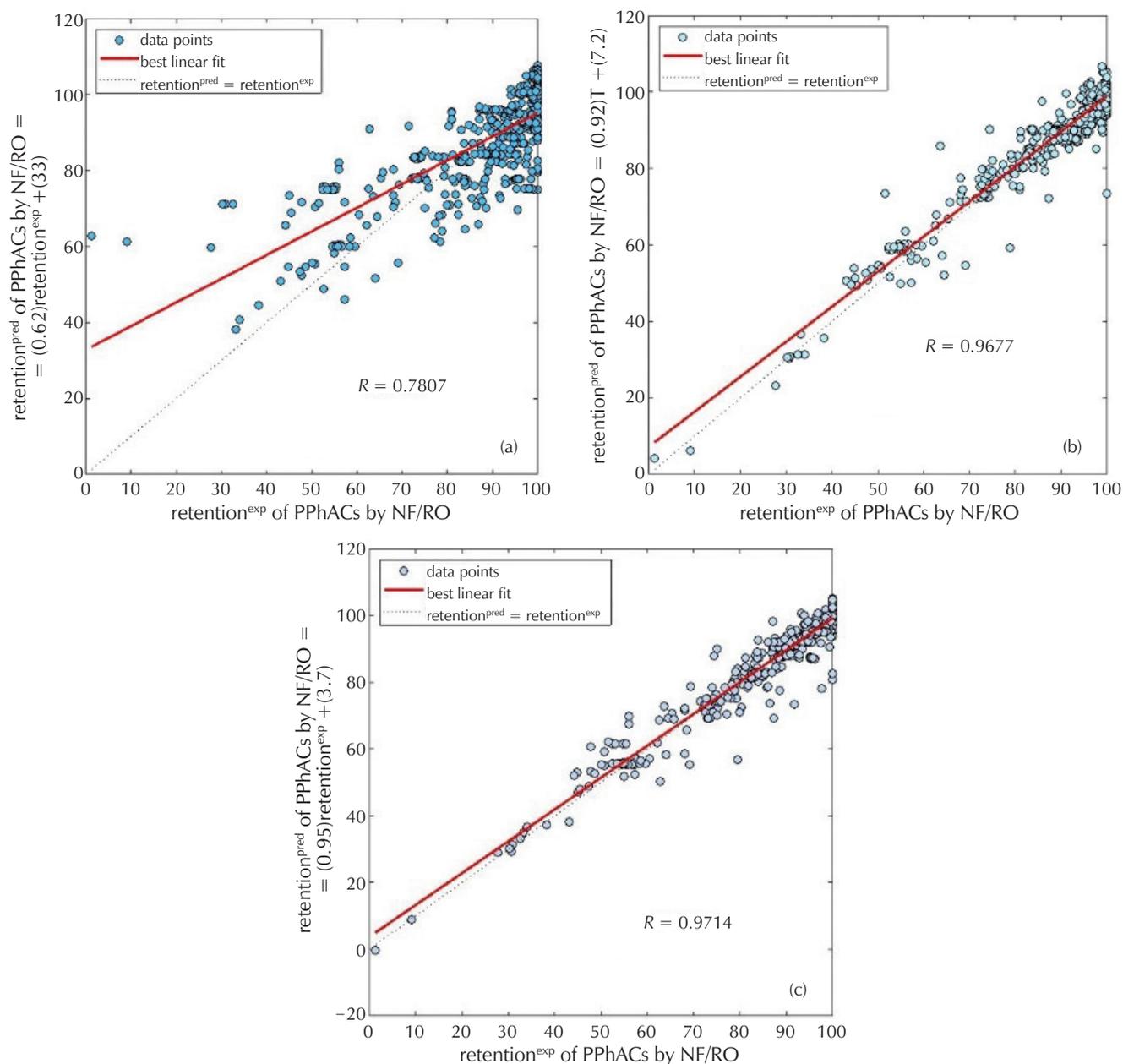


Fig. 6 – Predicted vs. experimental data for: (a) RBF 3, (b) SVM 3, and (c) MLP 3

points, 518 (95.75 %) are considered valid, while only 23 (4.25 %) are classified as outside the applicability domain.

## 7 Conclusions

The prediction of the retention of PPhACs by NF/RO membranes was performed using three types of algorithms: MLP, RBF, and SVM. MLP and RBF models were optimised using the Quasi-Newton BFGS algorithm. This algorithm provided better results in terms of speed, convergence, and performance generation for the MLP models. The results demonstrated a high training and predictive capacity for the retention of PPhACs by NF/RO membranes with high correlation coefficient ( $R = 0.9714$ ), and a low root mean squared error (RMSE = 3.9139 %) for the entire data set.

Furthermore, the statistical indicators of robustness helped in selecting a better training algorithm, activation function, and network architecture [13 15 1] (thirteen neurons in the input, fifteen in the hidden layer, and one in the output layer). The prediction by the MLP also demonstrated a strong correlation between the experimental and predicted values of PPhACs retention. This suggested that the MLP model had superior predictive power. The sensitivity analysis was conducted, which highlighted that the retention of PPhACs is governed by three interactions arranged in descending order of importance: the polarity interactions (hydrophobicity/hydrophilicity), electrostatic repulsion (charge effect), and steric hindrance (sieving effect). This research suggests that the retention of PPhACs on the NF/RO membranes strongly depends on the topological polar surface area.

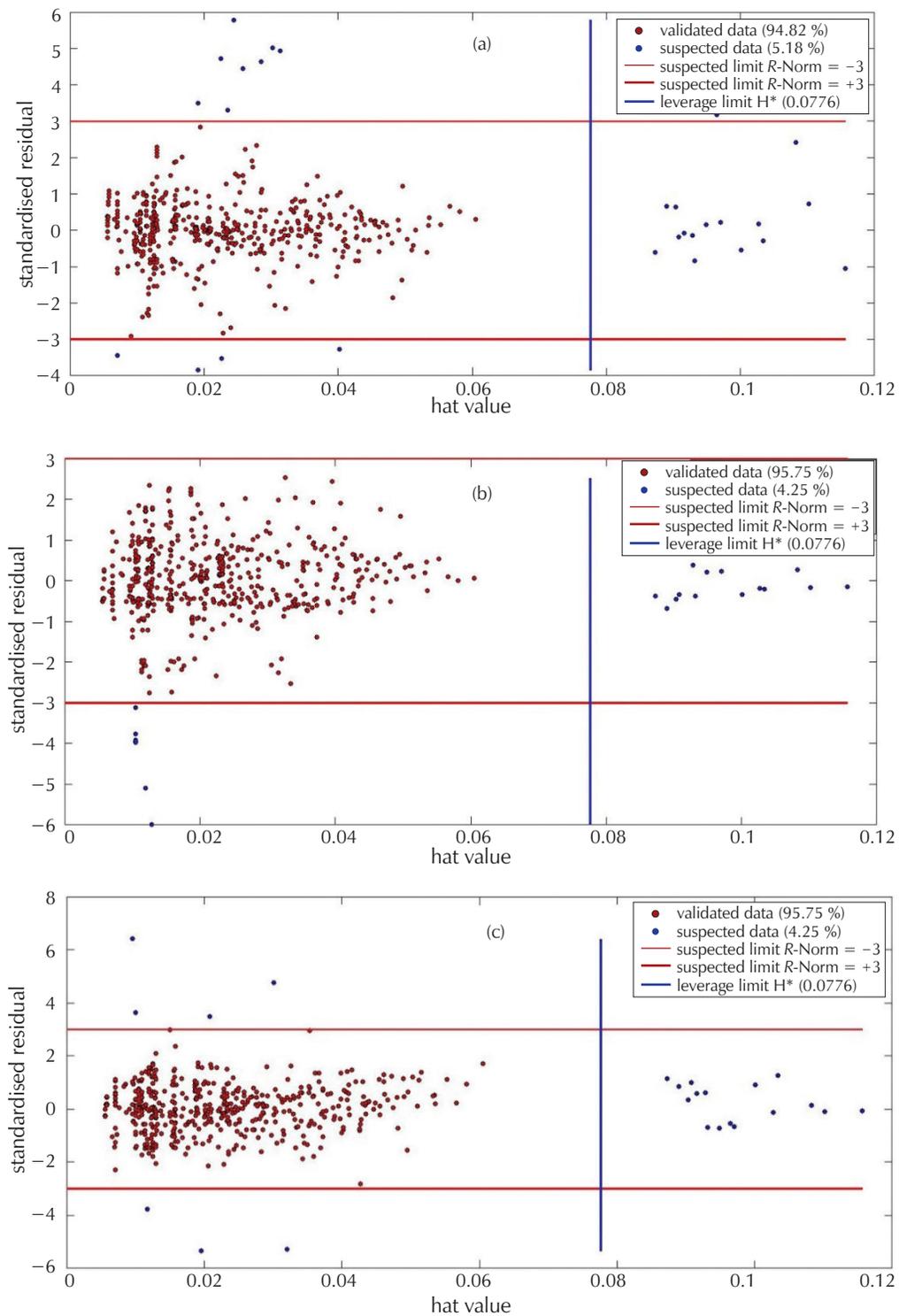


Fig. 7 – William's plots for detection of outliers for: a) MLP 3, b) RBF 3, and c) SVM models

## ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge the Ministry of Higher Education of Algeria for funding this research PRFU Projects N° A16N01UN260120220004. The authors also express their appreciation to the Laboratory of Biomaterials and Transport Phenomena at the University of Médéa.

## List of abbreviations and symbols

ANN	– artificial neural network
BANN	– bootstrap aggregated neural network
BFGS	– Broyden–Fletcher–Goldfarb–Shanno
$b^o$	– bias of the output layer
$b^H$	– bias of the hidden layer
$d_c$	– effective diameter of the organic compound in water
MLP	– multi-layer perceptron
RBF	– feedforward neural networks with radial basis function
$j$	– number of neurons in hidden layer
$i$	– number of neurons in input layer
$\log D_{ow}$	– distribution coefficient
MAE	– mean absolute error
MLR	– multiple linear regression
NF	– nanofiltration
NF/RO	– nanofiltration and reverse osmosis
PLS	– partial least squares
PPhACs	– polar pharmaceutical active compounds
QSAR	– quantitative structure-activity relationship
$R$	– correlation coefficient
RMSE	– root mean squared error
RO	– reverse osmosis
SVM	– support vector machines
$w^H$	– hidden layer output connection weight
$w^I$	– connection weight between input and hidden layer

## References

## Literatura

1. V. Albergamo, B. Blankert, E. R. Cornelissen, B. Hofs, W.-J. Knibbe, W. van der Meer, P. de Voogt, Removal of polar organic micropollutants by pilot-scale reverse osmosis drinking water treatment, *Water Res.* **148** (2019) 535–545, doi: <https://doi.org/10.1016/j.watres.2018.09.029>.
2. B. Teychene, F. Chi, J. Chokki, G. Darracq, J. Baronn, M. Joyeux, H. Gallard, Investigation of polar mobile organic compounds (PMOC) removal by reverse osmosis and nanofiltration: rejection mechanism modelling using decision tree, *Water Supply* **20** (2020) 975–983, doi: <https://doi.org/10.2166/ws.2020.020>.
3. K. Kim, K. H. Chu, Y. A. J. Al-Hamadani, C. M. Park, M. Jang, D. Kim, M. Yu, J. Heo, Y. Yoon, Removal of contaminants of emerging concern by membranes in water and wastewater: a review, *Chem. Eng. J.* **335** (2018) 896–914, doi: <https://doi.org/10.1016/j.cej.2017.11.044>.
4. W. Shichong, L. Lei, Y. Shuili, D. Bingzhi, G. Naiyun, W. Xianyun, A review of advances in EDCs and PhACs removal by nanofiltration: Mechanisms, impact factors and the influence of organic matter, *Chem. Eng. J.* **406** (2021) 126722, doi: <https://doi.org/10.1016/j.cej.2020.126722>.
5. D. Libotean, J. Giral, R. Rallo, Y. Cohen, F. Giral, H. F. Ridgway, G. Rodriguez, D. Phipps, Organic compounds passage through RO membranes, *J. Membr. Sci.* **313** (2008) 23–43, doi: <https://doi.org/10.1016/j.memsci.2007.11.052>.
6. V. Yangali-Quintanilla, M. Kennedy, G. Amy, T. U. Kim, Modeling of RO/NF membrane rejections of PhACs and organic compounds: A statistical analysis, *Drink. Water Eng. Sci.* **1** (2008) 7–15, doi: <https://doi.org/10.5194/dwes-1-7-2008>.
7. V. Yangali-Quintanilla, A. Verliède, T. U. Kim, A. Sadmani, M. Kennedy, G. Amy, Artificial neural network models based on QSAR for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis membranes, *J. Membr. Sci.* **342** (2009) 251–262, <https://doi.org/10.1016/j.memsci.2009.06.048>.
8. V. Yangali-Quintanilla, A. Sadmani, M. McConville, M. Kennedy, G. A. Amy, QSAR model for predicting rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by nanofiltration membranes, *Water Res.* **44** (2010) 373–384, doi: <https://doi.org/10.1016/j.watres.2009.06.054>.
9. L. Flyborg, B. Björleinius, M. Ullner, K. M. Persson, A PLS model for predicting rejection of trace organic compounds by nanofiltration using treated wastewater as feed, *Sep. Purif. Technol.* **174** (2017) 212–221, doi: <https://doi.org/10.1016/j.seppur.2016.10.029>.
10. S. G. Meshram, V. P. Singh, O. Kisi, V. Karimi, C. Meshram, Application of Artificial Neural Networks, Support Vector Machine and Multiple Model-ANN to Sediment Yield Prediction, *Water Resour. Manag.* **34** (2020) 4561–4575, doi: <https://doi.org/10.1007/s11269-020-02672-8>.
11. H. Benimam, C. Si-Moussa, M. Laidi, S. Hanini, Modeling the activity coefficient at infinite dilution of water in ionic liquids using artificial neural networks and support vector machines, *Neural Comput. Appl.* **32** (2020) 8635–8653, doi: <https://doi.org/10.1007/s00521-019-04356-w>.
12. J. García-Alba, J. F. Bárcena, C. Ugarteburu, A. García, Artificial neural networks as emulators of process-based models to analyse bathing water quality in estuaries, *Water Res.* **150** (2019) 283–295, doi: <https://doi.org/10.1016/j.watres.2018.11.063>.
13. C. Bellona, J. E. Drewes, The role of membrane surface charge and solute physico-chemical properties in the rejection of organic acids by NF membranes, *J. Membr. Sci.* **249** (2005) 227–234, doi: <https://doi.org/10.1016/j.memsci.2004.09.041>.
14. H. Huang, H. Cho, K. Schwab, J. G. Jacangelo, Effects of feed-water pretreatment on the removal of organic microconstituents by a low fouling reverse osmosis membrane, *Desalination* **281** (2011) 446–454, doi: <https://doi.org/10.1016/j.desal.2011.08.018>.
15. L. D. Nghiem, S. Hawkes, Effects of membrane fouling on the nanofiltration of pharmaceutically active compounds (PhACs): Mechanisms and role of membrane pore size, *Sep. Purif. Technol.* **57** (2007) 176–184, doi: <https://doi.org/10.1016/j.seppur.2007.04.002>.
16. C. Y. Tang, Y.-N. Kwon, J. O. Leckie, Effect of membrane chemistry and coating layer on physiochemical properties of thin film composite polyamide RO and NF membranes: II. Membrane physiochemical properties and their dependence on polyamide and coating layers, *Desalination* **242** (2009) 168–182, doi: <https://doi.org/10.1016/j.desal.2008.04.004>.
17. D. Dolar, A. Vuković, D. Ašperger, K. Košutić, Effect of water matrices on removal of veterinary pharmaceuticals by na-

- nofiltration and reverse osmosis membranes, *J. Environ. Sci.* **23** (2011) 1299–1307, doi: [https://doi.org/10.1016/S1001-0742\(10\)60545-1](https://doi.org/10.1016/S1001-0742(10)60545-1).
18. D. Dolar, T. I. Zokic, K. Kosutic, D. Asperger, D. M. Pavlovic, RO/NF membrane treatment of veterinary pharmaceutical wastewater: comparison of results obtained on a laboratory and a pilot scale, *Environ. Sci. Pollut. Res.* **19** (2012) 1033–1042, doi: <https://doi.org/10.1007/s11356-012-0782-7>.
  19. D. Dolar, K. Kosutic, D. Asperger, Influence of Adsorption of Pharmaceuticals onto RO/NF Membranes on Their Removal from Water, *Water Air Soil Pollut.* **224** (2013a) 1–13, doi: <https://doi.org/10.1007/s11270-012-1377-0>.
  20. D. Dolar, K. Košutić, M. Periša, S. Babić, Photolysis of enrofloxacin and removal of its photodegradation products from water by reverse osmosis and nanofiltration membranes, *Sep. Purif. Technol.* **115** (2013b) 1–8, doi: <https://doi.org/10.1016/j.seppur.2013.04.042>.
  21. J. Garcia-Ivars, L. Martella, M. Massella, C. Carbonell-Alcaina, M.-I. Alcaina-Miranda, M.-I. Iborra-Clar, Nanofiltration as tertiary treatment method for removing trace pharmaceutically active compounds in wastewater from wastewater treatment plants, *Water Res.* **15** (2017) 360–373, doi: <https://doi.org/10.1016/j.watres.2017.08.070>.
  22. K. P. M. Licon, L. R. O. Geaquinto, J. V. Nicolini, N. G. Figueiredo, S. C. Chiapetta, A. C. Habert, L. Yokoyama, Assessing potential of nanofiltration and reverse osmosis for removal of toxic pharmaceuticals from water, *J. Water Process Eng.* **25** (2018) 195–204, doi: <https://doi.org/10.1016/j.jwpe.2018.08.002>.
  23. R. Xu, P. Zhang, Q. Wang, X. Wang, K. Yu, T. Xue, X. Wen, Influences of multi-influent matrices on the retention of PPCPs by nanofiltration membranes, *Sep. Purif. Technol.* **212** (2019) 299–306, doi: <https://doi.org/10.1016/j.seppur.2018.11.040>.
  24. Y.-y. Zhao, X.-m. Wang, H.-w. Yang, Y.-f. F. Xie, Effects of organic fouling and cleaning on the retention of pharmaceutically active compounds by ceramic nanofiltration membranes, *J. Membr. Sci.* **1** (2018) 734–742, doi: <https://doi.org/10.1016/j.memsci.2018.06.047>.
  25. C. F. Couto, A. V. Santos, M. C. Santos Amara, L. C. Lange, L. H. de Andrade, A. F. S. Foureaux, B. S. Fernandes, Assessing potential of nanofiltration, reverse osmosis and membrane distillation drinking water treatment for pharmaceutically active compounds (PhACs) removal, *J. Water Process Eng.* **33** (2020) 101029, doi: <https://doi.org/10.1016/j.jwpe.2019.101029>.
  26. URL: <http://www.chemspider.com> (15. 3. 2020).

## SAŽETAK

### Strojno učenje i neuronske mreže u modeliranju zadržavanja polarnih farmaceutski aktivnih tvari nanofiltracijom i reverznom osmozom

Yamina Ammi,\* Cherif Si-Moussa i Salah Hanini

Zadržavanje polarnih farmaceutski aktivnih tvari (PPhAC) tijekom nanofiltracije i reverzne osmoze (NF/RO) od iznimne je važnosti u membranskim separacijskim procesima. Membransko zadržavanje 21 PPhAC-a korelirano je sa svojstvima PPhAC-a, karakteristikama membrane i uvjetima provedbe procesa filtracije. Pri tome su primijenjene tehnike umjetne inteligencije: višeslojni perceptron (MLP), neuronska mreža s radijalnom baznom funkcijom (RBF) i metoda potpornih vektora (SVM). Iz literature je prikupljena 541 vrijednost zadržavanja. Rezultati su pokazali visok kapacitet predviđanja MLP modela za cijeli skup podataka, s vrlo visokom vrijednošću koeficijenta korelacije ( $R = 0,9714$ ) i vrlo niskom vrijednošću korijenja srednje kvadratne pogreške (RMSE = 3,9139 %). Usporedba s preostala dva modela (RBF i SVM) pokazala je superiornost MLP modela. Analiza osjetljivosti ukazala je na to da zadržavanjem PPhAC-a upravljaju tri interakcije i to (padajućim redoslijedom): polarne interakcije (hidrofobnost/hidrofilnost), elektrostatsko odbijanje i steričke smetnje. Provedeno istraživanje sugerira da zadržavanje PPhACs na NF/RO membrani snažno ovisi o topologiji polarne površine.

#### Ključne riječi

Strojno učenje, neuronske mreže, modeliranje, zadržavanje, PPhACs, nanofiltracija, reverzna osmoza

Laboratory of Biomaterials and Transport Phenomena (LBMP), University of Médéa, 26 000 Médéa, Alžir

Izvorni znanstveni rad  
Prispjelo 30. prosinca 2022.  
Prihvaćeno 29. svibnja 2023.