https://doi.org/10.15255/KUI.2022.081
KUI-37/2023
Original scientific paper
Received December 9, 2022
Accepted March 17, 2023

Support Vector Machines for Evaluating the Impact of the Forward Osmosis Membrane Characteristics on the Rejection of the Organic Molecules

F. Kratbi,* *Y. Ammi*, and *S. Hanini*

Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, Algeria

Abstract

The forward osmosis (FO) process is currently being studied more despite other energy-consuming processes. In addition, several works show the performance of FO membranes as its major challenges, the study of the rejection of different molecules, energy consumption, and modelling of different objectives related to this process. The main purpose of our study was to evaluate the impact of the FO membranes characteristics on the rejection of organic molecules (neutral) by modelling of the latter. However, the current work deals with the application of Support Vector Machines (SVM) for predicting the rejection of organic molecules (53) by the FO membranes. In addition, the SVM model was compared with two other models: Artificial Neural Network (ANN) and Multiple Linear Regression (MLR). The coefficient of correlation (*R*) for the testing data was applied to display the best SVM model. The SVM model generated with Radial Basis Function (RBF) as the kernel function showed the best *R* value equal to 0.8526. MLR and ANN models had *R* values of 0.7630 and 0.8723, respectively.

Keywords

Support vector machines, forward osmosis, membranes, rejection, organic molecules

1 Introduction

Membrane separation processes include a large number of techniques for performing separation in different phases (liquid, gas, and mixed-phase) under the action of different driving forces, such as potential difference, electrical force, pressure, or chemical potential.[1,2] Microfiltration (MF), ultrafiltration (UF), nanofiltration (NF), and reverse osmosis (RO) constitute the main separation membranes process using the gradient of pressure as the driving force.[1–4] Recently, many works have studied the forward osmosis (FO) process as the replacement of those techniques in order to reduce the amount of energy used to generate the pressure difference upstream and downstream of the membrane.[5–7] However, several technical barriers impede FO industrial applications, the high performance membrane is one of the most important challenges of the FO process.[8] For this purpose, the literature highlights several points. Firstly, it is essential to explain correctly the process of FO with the existence of many models of transfer, namely, irreversible thermodynamics, solubilisation, diffusion, and pores.[9,10] Indeed, FO presents many interactions between solutes, water, and membranes.[11–16] In addition, the desired characteristics of the FO membrane would be higher water permeability and low solute permeability.[8,17] Artificial intelligence has been proposed to describe nonlinear systems due to its numerous be nefits (minimisation of the number of experiments, cost reduction, time saving, and studying the systems without profound knowledge).[18] Support Vector Machines (SVM) have been commonly used to model nonlinear systems with different components.[19] Moreover, for the FO process, there are few numbers of studies on the modelling of the FO system compared with other separation membrane processes, such as nanofiltration and reverse osmosis. Nevertheless, few studies have used artificial neural networks (ANNs) to predict the rejection of organic molecules (neutral) by the FO membranes. *Pardeshi et al.*[20] successfully applied ANN for determination of the optimal conditions for FO desalination of groundwater. In 2020, the modelling of FO process was studied by *Jawad et al.* using an ANN to predict the permeate flux based on nine inputs effect: membrane (type and orientation), feed solution (concentration, velocity, and temperature), and draw solution (concentration, molecular weight, velocity, and temperature), with different parameters of the ANN based on the number of neurons and hidden layers.[21] The results obtained proved the capability of ANN to predict the relationships between inputs and outputs in a way better than Multiple Linear Regression (MLR). The machine learning was applied to predict the internal concentration polarisation in FO.[22] In another study, the machine learning was used to optimise the performance of FO for the treatment of textile industry wastewater.[23] Furthermore, literature offers no study on using SVM to predict the FO process compared with the other membranes processes.

To the best of our knowledge, our work is the first to use SVM to study the FO process by studying the impact of membrane characteristics on the rejection of organic molecules. In summary, several models will be generated to predict the rejection of organic molecules by FO membranes. Firstly, many SVM models will be developed using different inputs, training algorithms, kernel functions, SVM parameters, and different database subdivision.

* Corresponding author: Fouad Kratbi, PhD
Email: kratbi.fouad@univ-medea.dz

Secondly, the SVM models developed will be compared to obtain the best model. Thirdly, other models, namely ANN and MLR models, will be generated using the database and the subdivision that gives the best SVM model accuracy. The last one is the comparison between all the optimal models obtained with different methods (SVM, ANN, and MLR). In the end, a conclusion recaps the results obtained.

2 Materials and methods

2.1 Support Vector Machines

Support Vector Machines (SVM) is an algorithm developed by the machine learning community (*Vapnick* 1982; *Burges* 1998). It is a superior form of machine learning that uses a classification algorithm. Its approach controls automatically the flexibility of the resulting classifier on the training data.[19,24,25]

Since its development, the SVM has attracted attention and extensive applications because of its remarkable performance. The SVM method is based on statistical learning theory, with the structural risk minimisation theory.[19]

The regression function makes the relationship between the inputs and outputs in the SVM method; its result can be depicted with Eq. (1).[19,26]

$$f\left(x_i\right) = w^{\mathrm{T}}\varphi\left(x_i\right) + b,\ i = 1,2,\ldots,N \tag{1}$$

$f(x_i)$ represents the predicted value of the SVM model, $\varphi(x_i)$ is the nonlinear function that draws input finite-dimensional into a higher-dimensional space, while w^{T} and b are the weight vector and the bias, respectively.[19]

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

The SVM optimisation model is given by Eq. (2) (for the training database):

$$\begin{cases} \min \mathrm{R}\left(w,\xi,\xi^*,\varepsilon\right) = \dfrac{1}{2}w^2 + C\left[v\varepsilon + \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)\right] \\ \text{subject to: } y_i - w^{\mathrm{T}}\varphi\left(x_i\right) - b \le \varepsilon + \xi_i \\ w^{\mathrm{T}}\varphi\left(x_i\right) + b - y_i \le \varepsilon + \xi_i \\ \xi^*,\varepsilon \ge 0 \end{cases} \tag{2}$$

C is the parameter used to balance the empirical risk and model complexity term w^2, ξ_i is the error between the real and estimated value, ξ_i^* represents the slack variable to denote the distance of the sample outside of the ε-tube, ε is equal to 0.1000, and v is the vector of independent variables.

As a standard nonlinear constrained optimisation problem, the previous problem can be resolved using the Lagrange multipliers techniques:[19]

$$\begin{cases} \max \mathrm{R}\left(a_i,a_i^*\right) = \displaystyle\sum_{i=1}^{N} y_i\left(a_i,a_i^*\right) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(a_i,a_i^*\right)\left(a_j,a_j^*\right)K(x_i,x_j) \\ \text{subject to: } \displaystyle\sum_{i=1}^{N} y_i\left(a_i,a_i^*\right) = 0 \\ 0 \le a_i,a_i^* \le C/N \\ \displaystyle\sum_{i=1}^{N}\left(a_i + a_i^*\right) \le C \cdot v \end{cases} \tag{3}$$

$K(x_i,x_j)$ is the kernel function satisfying Mercer's condition, and a_i and a_i^* are the non-negative Lagrange multipliers, respectively. Eq. (1) becomes Eq. (4).[19]

$$f\left(x_i\right) = \sum_{i=1}^{N}\left(a_i - a_i^*\right)K\left(x - x_i\right) + b,\ i = 1,2,\ldots,N \tag{4}$$

2.2 Modelling procedure

The modelling procedure contains six chapters. The first consists of the collection of the database as complete as possible from the literature. This database makes three databases used to model the rejection of organic molecules by the FO membranes using the SVM method. The second is about the pretreatment and statistical analysis of the data set; it contains the minimum, maximum, standard deviation, and mean of each input. The third chapter contains the modelling of all databases using the SVM method with different splitting of the database, different kernel functions, and algorithm parameters. The next chapter is the comparison of all SVM models generated for obtaining the optimal SVM model. The database and subdivision given, this optimal model was used to generate other models with different methods (ANN and MLR); it is the chapter on the development of ANN and MLR models. The last one is a comparison between the results obtained (the best performance SVM model obtained, the optimal ANN model, and the best efficient MLR model developed). Fig. 1 demonstrates the cited chapters.

2.2.1 Database; collection, pretreatment, and analysis

We collected a database from the accessible literature,[27–43] intending to group all the characteristics of the studied system. The software "Get data Graph" was used to take the values of the rejection from the representative diagrams. The inputs are represented in the properties of organic molecules, membrane characteristics, and purification operating conditions for the rejection of the organic molecules by FO membranes. The inputs considered in this work are the molecule properties (effective diameter of an organic molecule in water d_c, molecular length "length", molecular width "width", molecular depth "depth", dipole moment, logarithm of the octanol-water partition coefficient "log-K_{ow}"), and functional conditions (pH, crossflow velocity, and water flux). For each group of inputs, we used one of the membrane characteristics (surface membrane charge as "zeta potential", hydrophobicity as "contact angle", and porosity as "pore diameter"). The values of d_c, width, and depth were calculated with Eqs. (5)–(7), respectively.[44–46]

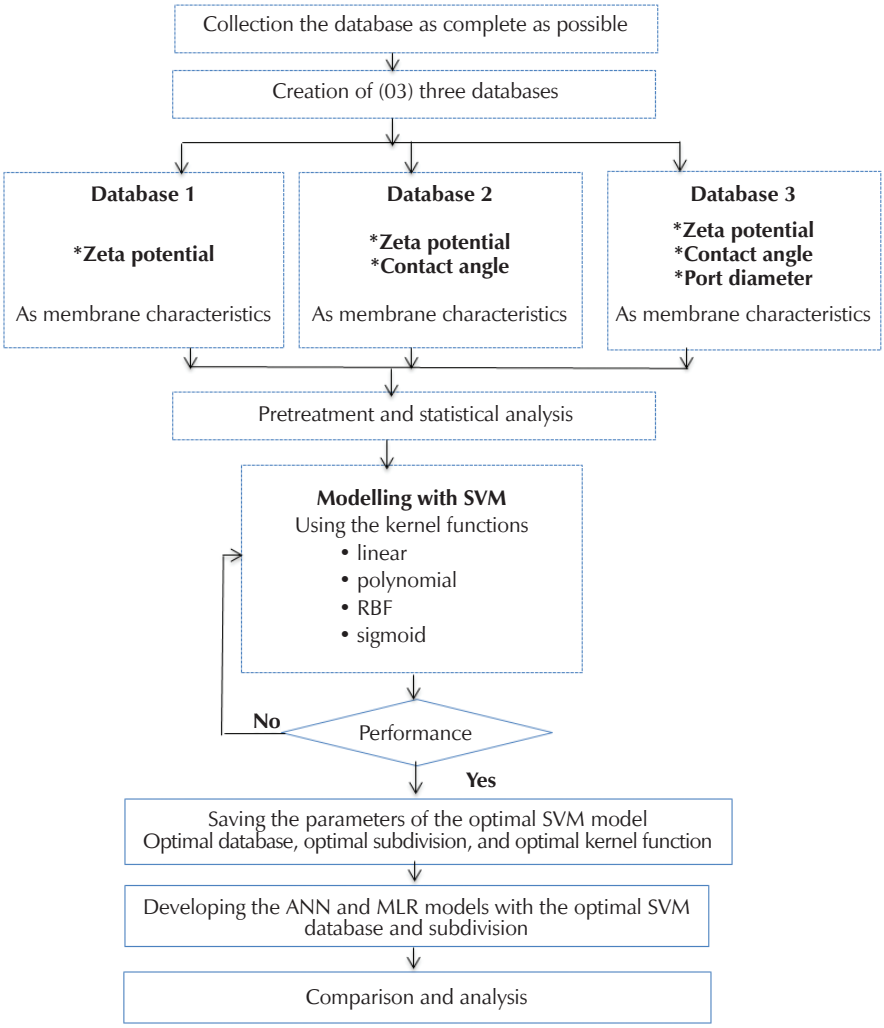

Fig. 1 – Modelling procedure

$$d_c = 0.065 \cdot M_w^{0.0438} \quad (5)$$

$$\text{width} = \frac{1}{2}\sqrt{S_{min}} \quad (6)$$

$$\text{depth} = \frac{1}{2}\sqrt{S_{max}} \quad (7)$$

M_w is molecular weight. S_{min} and S_{max} represent the minimum and maximum surface area, respectively. Table 1 presents the inputs and outputs of each database.

The properties (descriptors) of the organic molecules (available in supplementary data Table S1) used in this work were applied with the software *Hyperchem*, 2008, for the minimum and maximum surface area calculation, as well as with the software *Chembio*, 2014, for determination of the organic molecules length.

The statistical analysis is preliminary (standard deviations (STD), minimum, maximum, and mean), and is shown in supplementary data (Table S2).

For each database, a matrix correlation was used to edit the interactions between the variables (organic molecule properties, membrane characteristics, and operating conditions).

2.2.2 Development of the SVM models

For each database, several SVM models were developed using different kernel functions (linear, polynomial, radial basis function (RBF), and sigmoid). For each kernel function, several parameters (Coefficient, Gamma, Nu, and Degree) values were tested to obtain the best results. Many models were established by separating each database into two parts (training and testing phase), and using different kernel functions to obtain the optimal SVM model. The correlation coefficient R and the mean absolute error (MAE; Eq. (8)) were used to compare the performance of SVM models obtained for the organic molecules rejection by the FO membranes developed.[47]

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_{i,exp} - y_{i,cal}\right| \quad (8)$$

where $y_{i,exp}$ and $y_{i,cal}$ represent the experimental and the predicted values, respectively, and N is the number of data.

Table 1 – Components of each database

	Database 1	Database 2	Database 3
Inputs	d_c / g mol^{-1}	d_c / g mol^{-1}	d_c / g mol^{-1}
	logK_{ow}	logK_{ow}	logK_{ow}
	dipole moment (Debye)	dipole moment (Debye)	dipole moment (Debye)
	length / nm	length / nm	length / nm
	width / nm	width / nm	width / nm
	depth / nm	depth / nm	depth / nm
	zeta potential / mV	zeta potential / mV	zeta potential / mV
	–	contact angle / °	contact angle / °
	–	–	pore diameter / nm
	pH	pH	pH
	crossflow velocity / m h^{-1}	crossflow velocity / m h^{-1}	crossflow velocity / m h^{-1}
	water flux / l m^{-2} h^{-1}	water flux / l m^{-2} h^{-1}	water flux / l m^{-2} h^{-1}
Outputs	rejection / %	rejection / %	rejection / %
Data size	**333**	**193**	**115**

STATISTICA was used for the modelling of the rejection of organic molecules by the FO membranes using the SVM method to study the influence of the membrane characteristics on the rejection of the OM.

3 Results and discussion

3.1 Developing the SVM models

Each database cited previously was divided into two subsets: training and testing. Many models were developed using the different subdivisions, different kernel functions, and different parameters (training constants). Table 2 recaps the results of the first database with the zeta potential as the only membrane characteristic.

According to Table 2, the subdivision of 70 % for training set and 30 % for testing gave the best results for this database compared to the other results obtained by other subdivisions.

The linear kernel function gave a correlation coefficient of 0.5453 and a value of the MAE above 8.3000 %. The second function (polynomial) offered a model with a coefficient of correlation above 0.6900 for the testing phase, and MAE less than 8.8000 %. RBF developed the best performance model for this first database with *R* equal to 0.7525 and MAE less than 5.8750 %. The last model with the sigmoid function was less efficient than the previous model, and with values of *R* and MAE for the testing phase equal to 0.5917 and 7.9172 %, respectively.

Table 2 – Results of the first database DB1

Kernel function	Parameters	Phases	*R*	MAE / %
Linear	SVM N° = 93 C = 18.000 Nu = 0.3	training phase (70 %) testing phase (30 %) total phase	0.4433 0.5453 0.4707	10.1288 8.3192 9.5854
Polynomial	SVM N° = 73 C = 10.000 Nu = 0.2 Degree = 4 Gamma = 0.2 Coefficient = 2	training phase (70 %) testing phase (30 %) total phase	0.6961 0.6907 0.6939	9.1503 8.7934 9.0431
RBF	SVM N° = 122 C = 8.000 Nu = 0.2 Gamma = 0.3	training phase (70 %) testing phase (30 %) total phase	0.7454 0.7525 0.7451	6.5714 5.8747 6.3622
Sigmoid	SVM N° = 96 C = 15.000 Nu = 0.4 Gamma = 0.1 Coefficient = 2	training phase (70 %) testing phase (30 %) total phase	0.4145 0.5917 0.4636	9.3139 7.9172 8.8983

For this first database, the SVM model developed for studying the effect of the FO membrane characteristics on the OM rejection using the RBF kernel function had the best performance among the other models developed with the other kernel functions.

Table 3 summarises the results obtained for the second database that contains the zeta potential and contact angle as the membrane characteristics. The table demonstrates that the subdivision of 80 % for training data and 20 % for testing gave the best results in comparison with the other database subdivisions for the modelling of the rejection of organic molecules by the FO membranes.

The linear kernel function gave *R* value equal to 0.7341 and a value of MAE greater than 9.92 %. The polynomial function generated a model with values 0.7825 for *R* and 8.9716 % for MAE. RBF developed the best performance model for this database with *R* value equal to 0.8526 and MAE less than 8.07 %. The last model using the sigmoid function as kernel function is the lowest performance model compared to the previous models. The values of *R* and MAE for the testing phase were 0.7138 and 9.3985 %, respectively.

For the second database, the SVM model developed with the RBF kernel function is the best performing model among all the other models generated.

The third database using the following membrane characteristics (zeta potential, contact angle, and pore diameter) was modelled with different parameters. Table 4 summarises its results.

Table 3 – Results obtained by the second database DB2

Kernel function	Parameters	Phases	*R*	MAE/%
Linear	SVM N° = 39 C = 3.000 Nu = 0.2	training phase (80 %) testing phase (20 %) total phase	0.5527 0.7341 0.6947	6.1685 9.9283 6.9191
Polynomial	SVM N° = 64 C = 8.000 Nu = 0.2 Degree =3 Gamma = 0.3 Coefficient = 2	training phase (80 %) testing phase (20 %) total phase	0.7478 0.7825 0.7591	5.4302 8.9716 6.1458
RBF	SVM N° = 87 C = 3.000 Nu = 0.4 Gamma = 0.5	training phase (80 %) testing phase (20 %) total phase	0.7231 0.8526 0.7682	4.4298 8.0672 5.1649
Sigmoid	SVM N° = 34 C = 9.000 Nu = 0.2 Gamma = 0.1 Coefficient = 1	training phase (80 %) testing phase (20 %) total phase	0.5432 0.7138 0.5898	6.2312 9.3985 6.8712

Table 4 – Results obtained by the third database DB3

Kernel function	Parameters	Phases	*R*	MAE/%
Linear	SVM N° = 13 C = 10.000 Nu = 0.1	training phase (80 %) testing phase (20 %) total phase	0.4734 0.6649 0.5234	17.4101 15.5788 17.0373
Polynomial	SVM N° = 44 C = 10.000 Nu = 0.5 Gamma = 0.3 Coefficient = 3 Degree = 3	training phase (80 %) testing phase (20 %) total phase	0.7693 0.7880 0.7688	8.5211 10.3110 8.8854
RBF	SVM N° = 58 C = 10.000 Nu = 0.5 Gamma = 0.3	training phase (80 %) testing phase (20 %) total phase	0.7398 0.8124 0.7535	6.7877 7.4933 6.9313
Sigmoid	SVM N° = 34 C = 6.000 Nu = 0.3 Gamma = 0.3 Coefficient = 3	training phase (80 %) testing phase (20 %) total phase	0.3492 0.4328 0.3638	11.6417 12.3188 11.7795

The subdivision that gave the best results among all subsets employed was the subset with 80 % for training and 20 % for testing.

Following the preceding Table, the values of *R* and MAE for this database were less than the previous results. For the linear kernel function, the values were sequential; 0.6649 and 15.5788 % for the *R* and MAE. The polynomial function gave *R* value equal to 0.7880 and MAE greater than 10.3100 %. The SVM model using the RBF function generated *R* value equal to 0.8124 and MAE equal to 7.4933 %. The sigmoid function provided *R* value equal to 0.4328 and MAE above 12.3100 %. Thus, the model developed with the RBF is the best performance model among the others.

According to the previous results, the best SVM model to study the effect of the FO membrane characteristics on the OM rejection by its prediction is the model with the second database, RBF is the kernel function. The second database, formed with the zeta potential and contact angle as the membrane characteristics, is the optimal database. This result suggests that the modelling of the rejection of organic molecules is more efficient with the coexistence of surface membrane charge (zeta potential) and the hydrophobicity (contact angle) as the membrane characteristics. This advantage was compared to the surface membrane charge as the only membrane characteristic; also that the porosity is an additional characteristic to the surface membrane charge and hydrophobicity. Table 5 illustrates the structure of the best performing SVM model.

Table 5 – Structure of the best performance model obtained

Database	Subdivision	SVM parameters	Kernel function	Kernel function parameter
DB2 11 Inputs 1 Output	80 % for training 20 % for testing	N° = 87 C = 3.000 Nu = 0.4	RBF	Gamma = 0.5

The scheme and the parameters of the linear regression were directly generated with the MATLAB function "Postreg". Fig. 2 demonstrates a comparison between the experimental values of the rejection of the organic molecules and the predicted values using the best performance SVM model obtained. The parameters of the ideal agreement vectors are [$\alpha = 1$ (slope), $\beta = 0$ (intercept), $R = 1.0000$]. For this model, the training sample was depicted with [α, β, and R] equal to [0.35, 61, and 0.7231]. The testing sample gave the agreement vectors with [0.34, 63, and 0.8526], and the total with [0.34, 62, and 0.7682].

3.2 Developing the ANN and MLR models

The structure of the best-performance SVM model (database and subdivision) was used for the generation of two other models: ANN and MLR, to predict the rejection of the organic molecules by the FO membranes.

3.2.1 Artificial neural network

ANN is one of the most popular machine learning tools; recently, it has been employed for the modelling of different complex systems because of its performance. ANN contains three layers (units): input, hidden, and output layers. The input variables (obtained from the external systems) are converted to instantaneous values, which are transformed towards the output. This operation can be carried out using neurons by exploiting the weights and biases of each neuron.[47–49]

The second database (DB2) with the optimal subdivision (80 % for training phase and 20 % for testing phase) was used to model the rejection of the organic molecules by FO.

The transfer functions (activation functions) used in the hidden layer (HL) were Logistic, Tangh, Exponential, and Sine. The activation function in the output layer (OL) was the Identity function (Purelin). The number of neurons in the hidden layer varied between three and twenty-five, the output layer with one neuron, and eleven neurons in the input layer. Table 6 presents the results of the different models.

The preferred performance model among the models created was based on the value of *R* for the testing phase. The table suggests, firstly, that all the models developed gave values of the coefficient correlation greater than 0.8000, except the 4th model using the Sine as the activation func-

Table 6 – Results of the ANN models developed

Training algorithm	Transfer function (HL)	Transfer function (OL)	Number of neurons (HL)	Phase	*R*	MAE/%
BFGS 42	Logistic	Purelin	3	Training Testing Total	0.8710 0.8520 0.8684	5.2224 3.7986 4.9420
BFGS 35	Tangh	Purelin	3	Training Testing Total	0.8704 0.8723 0.8684	4.8853 4.6708 4.8431
BFGS 10	Exponential	Purelin	13	Training Testing Total	0.8178 0.8366 0.8156	5.8350 5.2648 5.2727
BFGS 25	Sine	Purelin	4	Training Testing Total	0.6171 0.6682 0.6148	7.2929 7.1627 7.7732

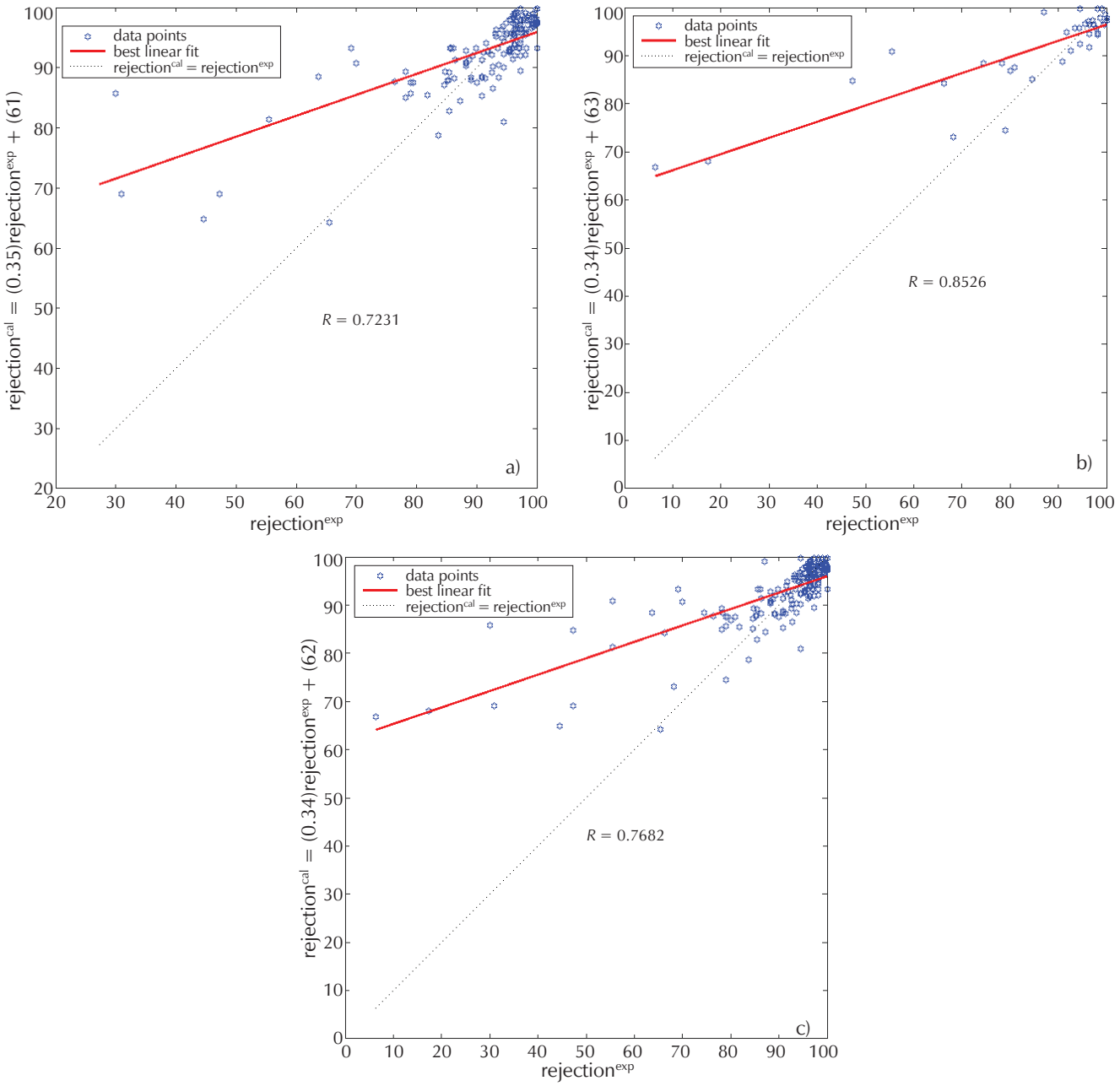

Fig. 2 – Agreement vectors with the SVM model obtained: (a) training phase, (b) testing phase, and (c) total phase

tion in the hidden layer, which generated a value of *R* less than 0.7000. These results explain the aptitude of the ANN models developed to predict the rejection of organic molecules by FO membranes.

Secondly, the best ANN model was generated by the BFGS as the training algorithm; the Tangh and the Purelin are the activation functions in the hidden and output layer, respectively. The number of neurons was eleven, three, and one neuron in the input, hidden, and output layer, respectively. This configuration represents the most efficient model among the four (4) ANN models created to predict the rejection of the organic molecules by FO membranes.

Fig. 3 depicts the scheme of the linear vectors obtained by the MATLAB "Postreg" function. This figure is no more than a comparison between the experimental values of rejection of organic molecules by the FO membranes and the predicted values established with the best ANN model developed. As cited previously, the ideal agreement vectors are usually represented by the [$\alpha = 1$ (slope), $\beta = 0$ (intercept), $R = 1.0000$]. For this ANN model, the training sample was generated with [α, β, and R] equal to [0.77, 20, and 0.8704], the testing sample offered agreement vectors of [0.97, 42.5, and 0.8723], and the total was [0.79, 18, and 0.8683].

3.2.2 Multiple linear regression

Multiple linear regression (MLR) is a statistical tool used to model linear systems. It consists of modelling an objective

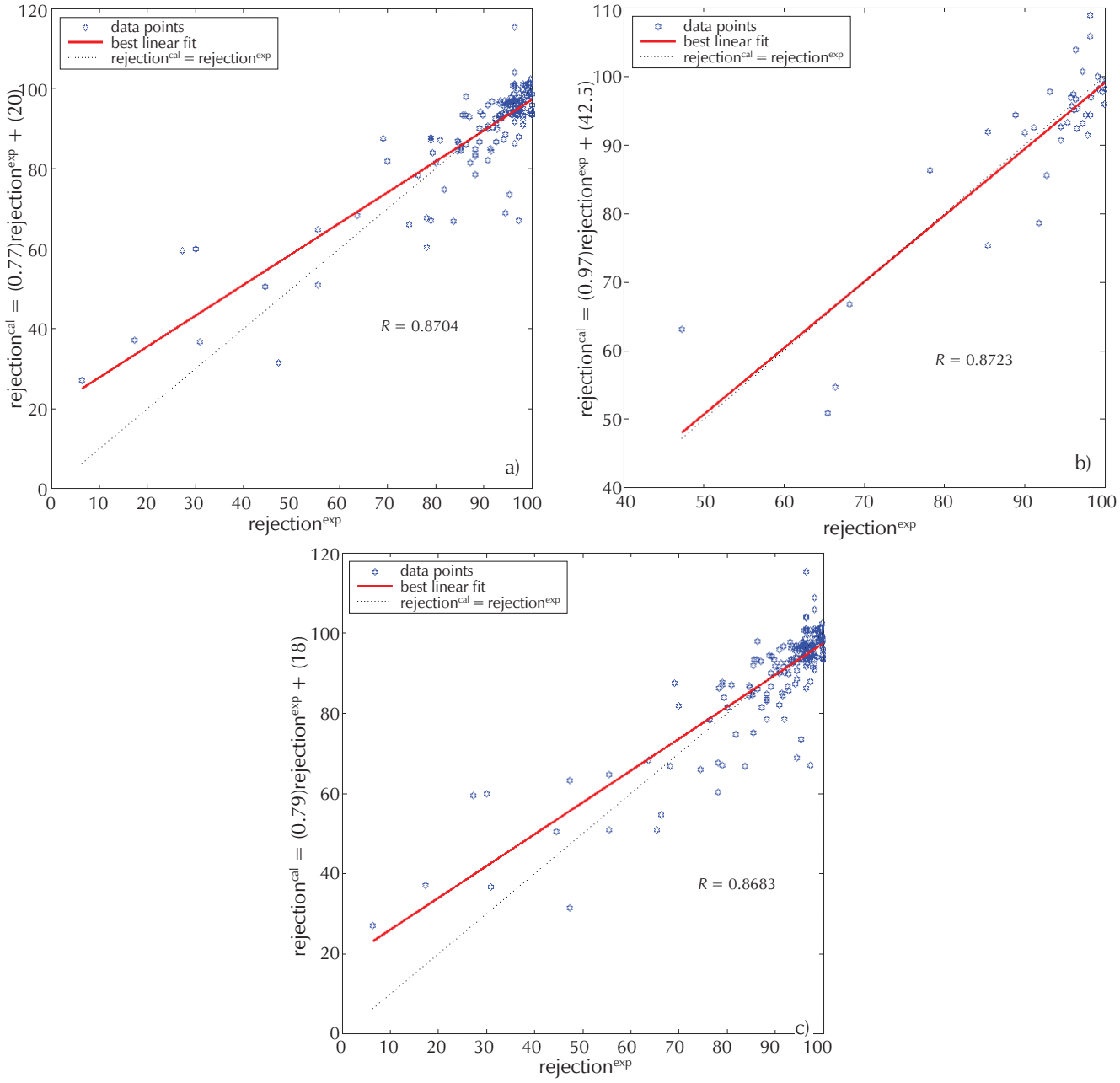

Fig. 3 – Agreement vectors with the ANN model obtained: (a) training phase, (b) testing phase, and (c) total phase

(output) according to several inputs using the quantitative relationships as cited in Eq. (9).[47,49]

$$y_{i,\,\mathrm{cal}} = B + \sum_{i=1}^{N}\sum_{j=1}^{M} A_j X_{i,j} \tag{9}$$

where $y_{i,\mathrm{cal}}$ is the predicted value of the rejection, B represents the intercept, A_j is the constant of each input, $X_{i,j}$ are the inputs, and N and M are the sizes of data sets.

The optimal database with the optimal subdivision (80 % for training sample) of the best-performance SVM model was used to create the MLR model of the rejection of OM by the FO membranes. The testing data was used to examine the performance of the model obtained. The MLR predicted values developed in this work are given by Eq. (10).

$$\mathrm{rejection}_{\mathrm{OM}} = B + \sum_{i=1}^{193}\sum_{j=1}^{11} A_j X_{i,j} \tag{10}$$

The organic molecules rejection by the FO membranes obtained with the MLR model is given by Eq. (11).

$$\begin{aligned}\mathrm{rejection}_{\mathrm{OM/FO}} = {} & 17.6436 \cdot d_c - 0.1964 \cdot \log K_{\mathrm{ow}} - \\ & - 1.3516 \cdot \text{dipole moment} + 19.1273 \cdot \text{length} + \\ & + 29.1402 \cdot \text{width} + 15.7089 \cdot \text{depth} + \\ & + 0.0024 \cdot \text{zeta potential} + 0.0981 \cdot \text{contact angle} + \\ & + 0.2791 \cdot \mathrm{pH} - 0.0116 \cdot \mathrm{CFV} + 0.4267 \cdot \text{water flux}\end{aligned} \tag{11}$$

The effective diameter, length, width, and depth are the inputs having the most influenced factors on the rejection

of organic molecules. Zeta potential, contact angle, pH, and water flux with less importance, and crossflow velocity (CFV), logK_{ow}, and dipole moment with negative factors.

The impact of the membrane characteristics on the rejection of organic molecules by the FO membranes was remarkable, because the increase in zeta potential (represents the membrane surface charge) led to the increase in organic molecules rejection (following the existence of a positive factor). The contact angle (identifies the hydrophobicity of membranes) had the same effect on the rejection of the organic molecules by the FO membranes, which was reflected, also, by the presence of a positive factor as cited in the rejection equation. In summary, according to the MLR model of the rejection of the organic molecules by the FO membrane developed, the rejection of OM was proportional to the FO membrane characteristics. Table 7 summarises the results relating to the performance of MLR model obtained.

Table 7 – Results of the MLR model

Phase	*R*	MAE/%
Training	0.5620	6.5094
Testing	0.7630	9.2871
Total	0.6155	7.0707

Fig. 4 shows the pattern of the linear vector achieved by the MATLAB "Postreg" function. This figure is an assess-

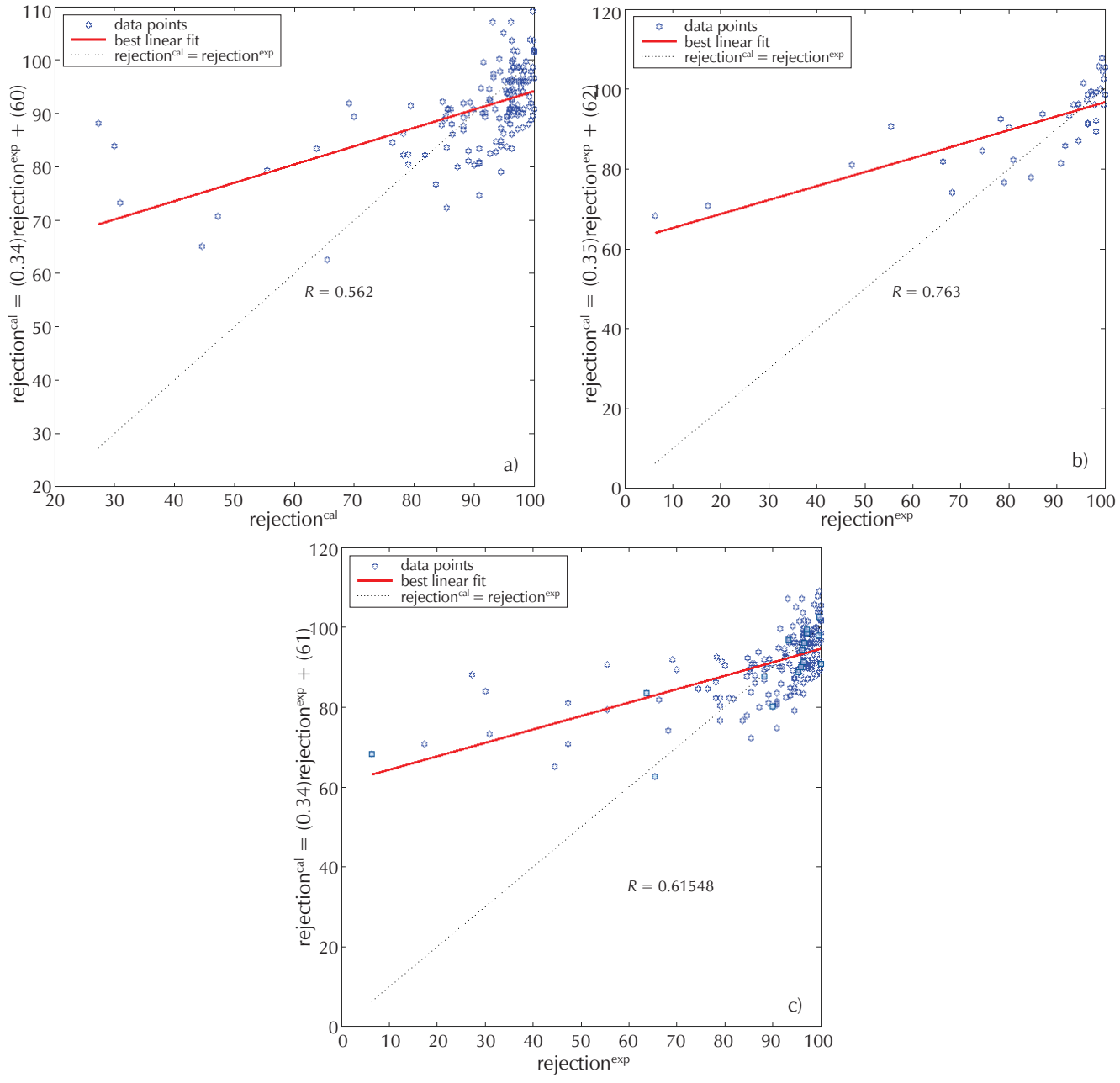

Fig. 4 – Agreement vectors with the MLR model obtained: (a) training phase, (b) testing phase, and (c) total phase

ment of the experimental values of the organic molecules rejection and the predicted values, which were obtained by the MLR model developed. For this MLR model, the training sample was produced with [α, β, and *R*] equal to [0.34, 60, and 0.5620], the testing sample had an agreement vector [0.35, 62, and 0.7630], and the total was [0.34, 61, and 0.6154].

3.3 Comparison between SVM, ANN, and MLR models

The comparison of the SVM, ANN, and MLR models served to extract the best-performance model among them. For this purpose, in addition to the correlation coefficient and the MAE, other statistical errors, root mean square error (RMSE), model predictive error (MPE), and standard error of prediction (SEP), were used to compare the testing phases of each model. The errors values were obtained with the following equations:[47,48–50]

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}\left(y_{i,\mathrm{exp}}-y_{i,\,\mathrm{cal}}\right)^2}{N}} \tag{12}$$

$$\mathrm{MPE}(\%)=\frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_{i,\mathrm{exp}}-y_{i,\mathrm{cal}}}{y_{i,\mathrm{exp}}}\right| \tag{13}$$

$$\mathrm{SEP}(\%)=\frac{\mathrm{RMSE}}{Y_e}\cdot 100 \tag{14}$$

where $y_{i,\mathrm{exp}}$, $y_{i,\mathrm{cal}}$ represent the experimental and the calculated value, respectively, N is the number of data, and Y_e is the mean value of experimental data.

Table 8 – Statistical comparison of models obtained

Model	*R*	MAE/%	RMSE/%	MPE/%	SPE/%
SVM	0.8526	8.0672	15.9551	39.5809	18.6216
ANN	0.8723	4.6708	6.2630	5.7765	6.8415
MLR	0.7630	9.2871	16.3703	41.5978	19.1061

Table 8 shows the comparison of the performance between the three models obtained. The *R* value of the SVM model was above the *R* of the MLR model and near the value of the ANN model. This result demonstrates the superiority of our model (SVM) in predicting the rejection of the organic molecules using the FO membrane in comparison with the MLR model, and the closeness of the SVM model to the ANN model.

The histogram of the MAE (presented by Fig. 5) shows that the SVM model had a value (8.0672 %) below the value of the MLR model (9.2871 %) and above that of the ANN model (4.6708 %). The SVM model had a RMSE value above the ANN model and lower than the value of the MLR model. Its value was equal to 15.9500 %; the MLR model with a value greater than 16.3700 %, and the ANN model with a value less than 6.2600 %.

For the MPE, the SVM and the MLR models had values above the value of the ANN model. The SVM model offered a value of 39.5809 %, while MLR model generated a value of 41.5978 %. The best model (ANN) gave a value equal to 5.7765 %.

The ANN model gave SEP value less than the two other models values. Their values were equal to 18.6216, 6.8415, and 19.1061 % for the SVM, ANN, and MLR models, respectively.

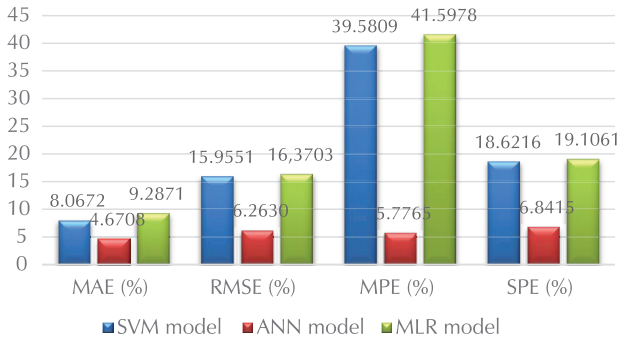

Fig. 5 – Comparison of the models obtained

4 Conclusion

This work investigates the use of the SVM in the FO membrane process with the aim of reviewing the weight of membrane characteristics on the rejection of organic molecules. Many models were created with diverse inputs, including the properties of OM, membrane characteristics, and operating conditions.

The generated SVM models demonstrated the existence of a direct relationship between the rejection of organic molecules and the FO membrane characteristics. It is translated by the variation of the rejection models obtained according to the membranes characteristics.

The comparison between the SVM models developed with different parameters and different kernel functions described that the model created with the RBF as the kernel function was the most efficient. Moreover, the database contained the zeta potential and contact angle as the membrane characteristics, and the subdivision 80 % for the training phase and 20 % for the testing phase was the best structure that gave the best performance model. Many ANN and MLR models were developed using this structure with different training algorithms, different activation functions in the HL, "Identity" as the activation function in the OL, and a varied number of neurons in the HL. The ANN model created with BFGS as the training algorithm, "Tangh" and Identity as the activation functions in the HL and OL, respectively, was the best performing model among the ANN models generated. Among all these models, the SVM was more efficient than the MLR model, and was close to the ANN model. The comparison between the SVM model, the ANN model and MLR model based on the correlation coefficient (*R*) and the MAE

showed R = 0.8526 and MAE = 8.0672 % for the SVM model, R = 0.8723 and MAE = 4.6708 % for the ANN model, and R = 0.7630 and MAE = 9.2781 % for the MLR model.

This work studied the precision of each model generated using RMSE, MPE, and SEP values with the outstanding superiority of the ANN model in comparison with SVM and MLR models.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Ministry of Higher Education of Algeria (PRFU Projects N° A16N01UN260120220004), and the group of Laboratory of Biomaterials and Transport Phenomena of University of Médéa.

Compliance with ethical standards

Conflict of interest

Authors declare that there is no conflict of interest.

SUPPLEMENTARY DATA

Table S1 – Names and properties of the organic molecules

N°	Name	Molecular formula	d_c / g mol^{-1}	Dipole (Debye)	logK_{OW}	Length / nm	Width / nm	Depth / nm
1	17a-Estradiol	$C_{18}H_{24}O_2$	0.757752	1.674	2.19	1.572	0.9332	1.0793
2	17a-Ethynylestradiol	$C_{20}H_{24}O_2$	0.7863256	1.031	1.43	1.57	1.0265	1.1286
3	17b-Estradiol	$C_{18}H_{24}O_2$	0.757752	2.536	2.19	1.523	0.9377	1.0822
4	Amitriptyline	$C_{20}H_{23}N$	0.7638373	0.4773	2.36	1.497	1.0492	1.1473
5	Androstenedione	$C_{19}H_{26}O_2$	0.7745941	0.5433	4.36	1.727	0.9632	1.0891
6	Androsterone	$C_{19}H_{30}O_2$	0.7793609	3.879	4.25	1.584	0.9846	1.1060
7	Atenolol	$C_{14}H_{22}N_2O_3$	0.7503338	1.464	−0.94	1.543	1.1312	1.1693
8	Atrazine	$C_8H_{14}ClN_5$	0.6841178	2.553	3.29	1.537	1.0181	1.0517
9	Bisphenol A	$C_{15}H_{16}O_2$	0.7013432	1.657	1.31	1.548	0.9731	1.0414
10	Caffeine	$C_8H_{10}N_4O_2$	0.6533663	3.822	−1.06	1.471	0.9253	0.9584
11	Carbamazepine	$C_{15}H_{12}N_2O$	0.7119776	3.286	−0.28	1.53	0.8428	1.0379
12	Clozapine	$C_{18}H_{19}ClN_4$	0.8207036	2.2283	−0.73	1.5661	1.0266	1.1524
13	DEET	$C_{12}H_{17}NO$	0.6490448	3.089	1.45	1.538	1.0160	1.0150
14	Diclofenac	$C_{14}H_{11}Cl_2NO_2$	0.7860234	3.026	−0.21	1.826	0.9572	1.0687
15	Dilantin	$C_{15}H_{12}N_2O_2$	0.7327073	1.684	0.73	1.55	0.8955	1.0405
16	Estriol	$C_{18}H_{24}O_3$	0.776935	1.183	1.42	1.554	0.9443	1.0909
17	Estrone	$C_{18}H_{22}O_2$	0.7552856	1.79	2.72	1.564	0.9445	1.0839
18	Etiocholanolone	$C_{19}H_{30}O_2$	0.7793609	3.879	4.25	1.565	0.9795	1.1060
19	Fluoxetine	$C_{17}H_{18}F_3NO$	0.8011581	1.709	1.93	1.548	1.1027	1.1534
20	Gemfibrozil	$C_{15}H_{22}O_3$	0.7302467	0.9124	2.51	1.55	1.1488	1.1263
21	Hydroxyzine	$C_{21}H_{27}ClN_2O_2$	0.8715522	2.248	1.23	1.535	1.2027	1.2751
22	Ibuprofen	$C_{13}H_{18}O_2$	0.6708811	1.764	2.75	1.523	1.0388	1.0459
23	Ketoprofen	$C_{16}H_{14}O_3$	0.7352712	2.721	2.56	1.535	0.9674	1.0687
24	Linuron	$C_9H_{10}Cl_2N_2O_2$	0.7286602	2.502	−0.35	1.803	1.0443	1.0348
25	Meprobamate	$C_9H_{18}N_2O_4$	0.0914904	1.472	3.03	1.546	1.0186	1.0529
26	Naproxen	$C_{14}H_{14}O_3$	0.7039877	1.556	0.56	1.525	0.9921	1.0453
27	Nonylphenol	$C_{15}H_{24}O$	0.6905528	1.518	3.9	1.539	1.1316	1.1370
28	Omeprazole	$C_{17}H_{19}N_3O_3S$	0.8408326	2.521	−1.44	1.873	1.1234	0.3795
29	Paracetamol	$C_8H_9NO_2$	0.5854728	4.309	−1.32	1.509	0.8741	0.9045
30	PFBS	$C_4HF_9O_3S$	0.7905983	1.539	3.3500	1.869	0.9548	0.9263
31	PFDA	$C_{10}HF_{19}O_2$	1.0008028	2.515	7.62	1.53	1.0938	1.1295

N°	Name	Molecular formula	d_c / g mol^{-1}	Dipole (Debye)	logK_{OW}	Length / nm	Width / nm	Depth / nm
32	PFDoA	$C_{12}HF_{23}O_2$	1.0818376	2.624	9.29	1.521	1.2157	1.1926
33	PFHpA	$C_7HF_{13}O_2$	0.8604133	2.378	5.12	1.528	1.0239	1.0007
34	PFHxA	$C_6HF_{11}O_2$	0.807108	2.337	4.28	1.527	0.9250	0.9326
35	PFHxS	$C_6HF_{13}O_3S$	0.8967429	1.523	5.02	1.521	0.9974	0.9900
36	PFNA	$C_9HF_{17}O_2$	0.9569319	2.685	6.78	1.537	1.0328	1.0441
37	PFOA	$C_8HF_{15}O_2$	0.9103149	2.372	5.95	1.526	0.9965	1.0104
38	PFOS	$C_8HF_{17}O_3S$	0.9888073	1.394	6.68	1.528	1.1240	1.0973
39	PFPeA	$C_5HF_9O_2$	0.7475012	2.674	3.45	1.533	0.8846	0.8813
40	Polyparaben (propylparaben)	$C_{10}H_{12}O_3$	0.6323157	3.81	0.8	1.543	0.5000	0.9701
41	Primidone	$C_{12}H_{14}N_2O_2$	0.6876625	3.175	0.72	1.554	0.8960	0.9973
42	Risperidone	$C_{23}H_{27}FN_4O_2$	0.9049504	1.578	1.03	1.527	1.0442	1.2339
43	Salicylic acid	$C_7H_6O_3$	0.5627892	2.186	−0.04	1.491	0.7637	0.8299
44	Simvastatin	$C_{25}H_{38}O_5$	0.9146349	3.998	4.43	1.54	1.1967	1.2718
45	Simvastatin Hydroxy Acid	$C_{25}H_{40}O_6$	0.9316776	2.664	4.14	1.54	1.2904	1.3179
46	Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	0.7339907	6.033	−1.54	1.578	0.9786	1.0529
47	Testosterone	$C_{19}H_{28}O_2$	0.7769822	3.672	3.84	1.569	0.9691	1.1032
48	t-Octylphenol	$C_{14}H_{22}O$	0.6709523	0.5321	3.44	1.572	1.0230	1.0250
49	Triamterene	$C_{12}H_{11}N_7$	0.733978	0.1612	1.19	1.468	0.8105	1.0296
50	Triclocarban	$C_{13}H_9Cl_3N_2O$	0.8082195	1.495	−0.5	1.814	0.9984	1.0736
51	Triclosan	$C_{12}H_7Cl_3O_2$	0.7783022	1.532	−0.6	1.811	1.0006	1.0446
52	Trimethoprim	$C_{14}H_{18}N_4O_3$	0.7792081	0.8107	−2.22	1.516	1.0235	1.1398
53	Verapamil	$C_{27}H_{38}N_2O_4$	0.9483294	4.344	0.57	1.6	1.3362	1.3626

Table S2 – Statistical analysis of inputs and outputs variables for all databases

	Min	Max	STD	Mean
d_c / g mol^{-1}	0.0914	1.1000	0.1291	0.7661
log P	0.1612	6.0330	1.3551	2.5554
dipole moment (Debye)	−2.2200	9.3000	2.7008	1.3628
length / nm	1.4680	1.8730	0.1258	1.6036
width / nm	0.5000	1.3362	0.0953	0.9871
depth / nm	0.3795	1.4000	0.0926	1.0519
zeta potential / mV	−39.9100	19.7000	11.9568	−7.6705
contact angle / °	42.7000	90.3000	6.2294	64.6435
Port diameter / nm	0.5400	0.7400	0.0571	0.7223
pH	3.0000	9.0000	1.1940	6.2432
crossflow velocity / m h^{-1}	288.0000	1094.4000	332.6317	504.0000
water flux / l m^{-2} h^{-1}	5.2579	20.0000	4.6249	9.3460
rejection / %	6.3636	100.0000	15.0353	88. 0400

List of abbreviations

MF – microfiltration
UF – ultrafiltration
NF – nanofiltration
RO – reverse osmosis
FO – forward osmosis
SVM – support vector machines
OM – organic molecules
ANN – artificial neural networks
MLR – multiple linear regression
d_c – effective diameter of an organic molecule in water
M_w – molecular weight
min – minimum
max – maximum
STD – standard deviation
RBF – radial basis function
MAE – mean absolute error
DB – database
R – correlation coefficient
Tangh – tangent hyperbolic function
RMSE – root mean square error
MPE – model predictive error
SEP – standard error of prediction
BFGS – Broyden–Fletcher–Goldfarb–Shanno
Purelin – pure linear function
HL – hidden layer
OL – output layer

References
Literatura

1. *Y. Ammi, L. Khaouane, S. Hanini*, Stacked neural networks for predicting the membranes performance by treating the pharmaceutical active compounds, Neural. Comput. Appl. **33** (19) (2021) 12429–12444, doi: https://doi.org/10.1007/s00521-021-05876-0.
2. *C. Cordier, K. Guyomard, C. Stavrakakis, P. Sauvade, F. Coelho, P. Moulin,* Culture of microalgae with ultrafiltered seawater: a feasibility study, SciMed. J. **2** (2) (2020) 56–62, doi: https://doi.org/10.28991/SciMedJ-2020-0202-2.
3. *S. Zhao, L. Zou, C. Y. Tang, D. Mulcahy*, Recent developments in forward osmosis: Opportunities and challenges, J. Membr. Sci. **396** (2012) 1–21, doi: https://doi.org/10.1016/j.memsci.2011.12.023.
4. *K. Lutchmiah, A. R. D. Verliefde, K. Roest, L. C. Rietveld, E. R. Cornelissen,* Forward osmosis for application in wastewater treatment: A review, Water Res. **58** (2014) 179197, doi: https://doi.org/10.1016/j.watres.2014.03.045.
5. *B. Mi, M. Elimelech*, Organic fouling of forward osmosis membranes: Fouling reversibility and cleaning without chemical reagents, J. Membr. Sci. **348** (1-2) (2010) 337–345, doi: https://doi.org/10.1016/j.memsci.2009.11.021.
6. *T. Y. Cath, A. E. Childress, M. Elimelech*, Forward osmosis: Principles, applications, and recent developments, J. Membr. Sci. **281** (1-2) (2006) 70–87, doi: https://doi.org/10.1016/j.memsci.2006.05.048.
7. *D. L. Shaffer, J. R. Werber, H. Jaramillo, S. Lin, M. Elimelech*, Forward osmosis: where are we now? Desalination **356** (2015) 271–284, doi: https://doi.org/10.1016/j.desal.2014.10.031.
8. *R. Wang, L. Shi, C. Y Tang, S. Chou, C. Qiu, A. G. Fane*, Characterization of novel forward osmosis hollow fiber membranes, J. Membr. Sci. **355** (1-2) (2010) 158–167, doi: https://doi.org/10.1016/j.memsci.2010.03.017.
9. *S. Gur-Reznik, I. Koren-Menashe, L. Heller-Grossman, O. Rufel, C. G. Dosoretz*, Influence of seasonal and operating conditions on the rejection of pharmaceutical active compounds by RO and NF membranes, Desalination **277** (1-3) (2011) 250–256, doi: https://doi.org/10.1016/j.desal.2011.04.029.
10. *Y. Lanteri, A. Szymczyk, P. Fievet*, Influence of steric, electric, and dielectric effects on membrane potential, Langmuir **24** (15) (2008) 7955–7962, doi: https://doi.org/10.1021/la800677q.
11. *T. S. Chung, S. Zhang, K. Y. Wang, J. Su, M. M. Ling*, Forward osmosis processes: yesterday, today and tomorrow, Desalination **287** (2012) 78–81, doi: https://doi.org/10.1016/j.desal.2010.12.019.
12. *C. R. Martinetti, A. E. Childress, T. Y. Cath*, High recovery of concentrated RO brines using forward osmosis and membrane distillation, J. Membr. Sci. **331** (1-2) (2009) 31–39, doi: https://doi.org/10.1016/j.memsci.2009.01.003.
13. *R. L. McGinnis, M. Elimelech,* Energy requirements of ammonia–carbon dioxide forward osmosis desalination, Desalination **207** (1-3) (2007) 370–382, doi: https://doi.org/10.1016/j.desal.2006.08.012.
14. *S. Zhao, L. Zou,* Effects of working temperature on separation performance, membrane scaling and cleaning in forward osmosis desalination, Desalination **278** (1-3) (2011) 157–164, doi: https://doi.org/10.1016/j.desal.2011.05.018.
15. *W. R. Bowen, M. G. Jones, J. S. Welfoot, H. N. Yousef,* Predicting salt rejections at nanofiltration membranes using artificial neural networks, Desalination **129** (2) (2000) 147–162, doi: https://doi.org/10.1016/S0011-9164(00)00057-6.
16. *G. R. Shetty, H. Malki, S. Chellam*, Predicting contaminant removal during municipal drinking water nanofiltration using artificial neural networks, J. Membr. Sci. **212** (1-2) (2003) 99–112, doi: https://doi.org/10.1016/S0376-7388(02)00473-8.
17. *S. S. Manickam, J. R. McCutcheon*, Understanding mass transfer through asymmetric membranes during forward osmosis: A historical perspective and critical review on measuring structural parameter with semi-empirical models and characterization approaches, Desalination **421** (2017) 110–126, doi: https://doi.org/10.1016/j.desal.2016.12.016.
18. *M. Dornier, M. Decloux, G. Trystram, A. Lebert*, Dynamic modeling of crossflow microfiltration using neural networks, J. Membr. Sci. **98** (3) (1995) 263–273, doi: https://doi.org/10.1016/0376-7388(94)00195-5.
19. *Y. Ammi, S. Hanini, L. Khaouane*, An artificial intelligence approach for modeling the rejection of anti-inflammatory drugs by nanofiltration and reverse osmosis membranes using kernel support vector machine and Neural Networks, Compt-es Rendus. Chimie **24** (2) (2021) 243–254, doi: https://doi.org/10.5802/crchim.76.
20. *P. M. Pardeshi, A. A. Mungray, A. K. Mungray*, Determination of optimum conditions in forward osmosis using a combined Taguchi–neural approach, Chem. Eng. Res. Des. **109** (2016) 215–225, doi: https://doi.org/10.1016/j.cherd.2016.01.030.
21. *J. Jawad, A. H. Hawari, S. Zaidi*, Modeling of forward osmo-

sis process using artificial neural networks (ANN) to predict the permeate flux, Desalination **484** (2020) 114427, doi: https://doi.org/10.1016/j.desal.2020.114427.

22. *I. Ibrar, S. Yadav, A. Braytee, A. Altaee, A. Hossein Zadeh, A. K. Samal, F. Fantozzi,* Evaluation of machine learning algorithms to predict internal concentration polarization in forward osmosis. J. Membr. Sci. **646** (2022) 120257, doi: https://doi.org/10.1016/j.memsci.2022.120257.
23. *K. Aghilesh, A. Mungray, S. Agarwal, J. Ali, M. C. Garg,* Performance optimisation of forward-osmosis membrane system using machine learning for the treatment of textile industry wastewater, J. Clean. Prod. **289** (2021) 125690, doi: https://doi.org/10.1016/j.jclepro.2020.125690.
24. *P. C. Deka*, Support vector machine applications in the field of hydrology: a review, Appl. Soft Comput. **19** (2014) 372–386, doi: https://doi.org/10.1016/j.asoc.2014.02.002.
25. *W. C. Leong, R. O. Kelani, Z. Ahmad*, Prediction of air pollution index (API) using support vector machine (SVM), J. Environ. Chem. Eng. **8** (2020) (3) 103208, doi: https://doi.org/10.1016/j.jece.2019.103208.
26. *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.
27. *A. A. Alturki, J. A. McDonald, S. J. Khan, W. E. Price, L. D. Nghiem, M. Elimelech,* Removal of trace organic contaminants by the forward osmosis process, Sep. Purif. Technol. **103** (2013) 258–266, doi: https://doi.org/10.1016/j.seppur.2012.10.036.
28. *G. S. Arcanjo, F. C. Costa, B. C. Ricci, A. H. Mounteer, E. N. DE Melo, B. F. Cavalcante, M. C. Amaral*, Draw solution solute selection for a hybrid forward osmosis-membrane distillation module: effects on trace organic compound rejection, water flux and polarization, Chem. Eng. J. **400** (2020) 125857, doi: https://doi.org/10.1016/j.cej.2020.125857.
29. *Y. Cui, X. Liu, Y. T. S. Chung, M. Weber, C. Staudt, C. Maletzko*, Removal of organic micro-pollutants (phenol, aniline and nitrobenzene)via forward osmosis (FO) process: evaluation of FO as an alternative method to reverse osmosis (RO), Water Res. **91** (2016) 104–114, doi: https://doi.org/10.1016/j.watres.2016.01.001.
30. *J. Heo, S. Kim, N. Her, C. M. Park, M. Yu, Y. Yoon*, Removal of contaminants of emerging concern by FO, RO, and UF membranes in water and wastewater, Contam. Emerg. Concern Water Wastewater **139** (2020) 139–176, doi: https://doi.org/10.1016/B978-0-12-813561-7.00005-5.
31. *S. J. Im, H. Lee, A. Jang*, Effects of co-existence of organic matter and microplastics on the rejection of PFCs by forward osmosis membrane, Environ. Res. **194** (2021) 110597, doi: https://doi.org/10.1016/j.envres.2020.110597.
32. *S. Jamil, P. Loganathan, C. Kazner, S. Vigneswaran,* Forward osmosis treatment for volume minimisation of reverse osmosis concentrate from a water reclamation plant and removal of organic micropollutants, Desalination **372** (2015) 32–38, doi: https://doi.org/10.1016/j.desal.2015.06.013.
33. *Y. Kim, S. Li, L. Chekli, Y. C. Woo, C. H. Wei, S. Phuntsho, H. K Shon*, Assessing the removal of organic micro-pollutants from anaerobic membrane bioreactor effluent by fertilizer-drawn forward osmosis, J. Membr. Sci. **533** (2017) 84–95, doi: https://doi.org/10.1016/j.memsci.2017.03.027.
34. *F. X. Kong, H. W. Yang, X. M. Wang, Y. F. Xie, R*ejection of nine haloacetic acids and coupled reverse draw solute permeation in forward osmosis, Desalination **341** (2014) 1–9, doi: https://doi.org/10.1016/j.desal.2014.02.019.
35. *H. Lee, S. J. Im, J. H. Park, A. Jang,* Removal and transport behavior of trace organic compounds and degradation byproducts in forward osmosis process: Effects of operation conditions and membrane properties, Chem. Eng. J. **375** (2019) 122030, doi: https://doi.org/10.1016/j.cej.2019.122030.
36. *C. Li, H. Li, Y. Yang, L. A. Hou,* Removal of pharmaceuticals by fouled forward osmosis membranes: Impact of DOM fractions, Ca^{2+} and real water, Sci. Total Environ. **738** (2020) 139757, doi: https://doi.org/10.1016/j.scitotenv.2020.139757.
37. *R. V. Linares, V. Yangali-Quintanilla, Z. Li, G. Amy,* Rejection of micropollutants by clean and fouled forward osmosis membrane, Water Res. **45** (20) (2011) 6737–6744, doi: https://doi.org/10.1016/j.watres.2011.10.037.
38. *H. T. Madsen, N. Bajraktari, C. Hélix-Nielsen, B. Van der Bruggen, E. G. Søgaard,* Use of biomimetic forward osmosis membrane for trace organics removal, J. Membr. Sci. **476** (2015) 469–474, doi: https://doi.org/10.1016/j.memsci.2014.11.055.
39. *M. Rastgar, A. Shakeri, A. Karkooti, A. Asad, R. Razavi, M. Sadrzadeh,* Removal of trace organic contaminants by melamine-tuned highly cross-linked polyamide TFC membranes, Chemosphere **238** (2020) 124691, doi: https://doi.org/10.1016/j.chemosphere.2019.124691.
40. *M. Xie, L. D. Nghiem, W. E. Price, M. Elimelech*, Comparison of the removal of hydrophobic trace organic contaminants by forward osmosis and reverse osmosis, Water Res. **46** (2012) (8) 2683–2692, doi: https://doi.org/10.1016/j.watres.2012.02.023.
41. *M. Xie, L. D. Nghiem, W. E. Price, M. Elimelech,* Impact of organic and colloidal fouling on trace organic contaminant rejection by forward osmosis: Role of initial permeate flux, Desalination **336** (2014) 146–152, doi: https://doi.org/10.1016/j.desal.2013.12.037.
42. *M. Xie, W. E. Price, L. D. Nghiem, M. Elimelech*, Effects of feed and draw solution temperature and transmembrane temperature difference on the rejection of trace organic contaminants by forward osmosis, J. Membr. Sci. **438** (2013) 57–64, doi: https://doi.org/10.1016/j.memsci.2013.03.031.
43. *L. Zheng, W. E. Price, J. McDonald, S. J. Khan, T. Fujioka, L. D. Nghiem,* New insights into the relationship between draw solution chemistry and trace organic rejection by forward osmosis, J. Membr. Sci. **587** (2019) 117184, doi: https://doi.org/10.1016/j.memsci.2019.117184.
44. *Y. Ammi, L. Khaouane, S. Hanini,* A model based on bootstrapped neural networks for modeling the removal of organic compounds by nanofiltration and reverse osmosis membranes. Arabian J. Sci. Eng. **43** (11) (2018) 6271–6284, doi: https://doi.org/10.1007/s13369-018-3484-8.
45. *D. Dolar, T. Ignjatić Zokić, K. Košutić, D. Ašperger, D. Mutavdžić Pavlović,* RO/NF membrane treatment of veterinary pharmaceutical wastewater: comparison of results obtained on a laboratory and a pilot scale, Environ. Sci. Pollut. Res. **19** (4) (2012) 1033–1042, doi: https://doi.org/10.1007/s11356-012-0782-7.
46. *J. L. Santos, P. de Beukelaar, I. F. Vankelecom, S. Velizarov, J. G. Crespo*, Effect of solute geometry and orientation on the rejection of uncharged compounds by nanofiltration, Sep. Purif. Technol. **50** (1) (2006) 122–131, doi: https://doi.org/10.1016/j.seppur.2005.11.015.
47. *Y. Ammi, L. Khaouane, S. Hanini*, A comparison of "neural networks and multiple linear regressions" models to describe the rejection of micropollutants by membranes, Kem. Ind. **69** (3-4) (2020) 111–127, doi: https://doi.org/10.15255/KUI.2019.024.
48. *N. A. Darwish, N. Hilal, H. Al-Zoubi, A. W. Mohammad,* Neural networks simulation of the filtration of sodium chlo-

ride and magnesium chloride solutions using nanofiltration membranes, Chem. Eng. Res. Design **85** (2007) 417–430, doi: https://doi.org/10.1205/cherd06037.

49. *N. Melzi, L. Khaouane, Y. Ammi, S. Hanini, M. Laidi, H. Zentou,* Comparative study of predicting the molecular diffusion coefficient for polar and non-polar binary gas using neural networks and multiple linear regressions, Kem. Ind. **68** (11-12) (2019) 573–582, doi: https://doi.org/10.15255/KUI.2019.010.
50. *M. R. Fissa, Y. Lahiouel, L. Khaouane, S. Hanini,* QSPR estimation models of normal boiling point and relative liquid density of pure hydrocarbons using MLR and MLP-ANN methods, J. Mol. Graph. Model. **87** (2019) 109–120, doi: https://doi.org/10.1016/j.jmgm.2018.11.013.

SAŽETAK

Metoda potpornih vektora u procjeni utjecaja karakteristika unaprijednih osmotskih membrana na zadržavanje organskih molekula

Fouad Kratbi, Yamina Ammi* i *Salah Hanini*

Proces unaprijedne osmoze (FO) trenutačno se učestalo proučava, a glavne su tematike zadržavanje različitih molekula, potrošnja energije i modeliranje samog procesa. Glavna svrha ovog istraživanja bila je, primjenom modeliranja, procijeniti utjecaj karakteristika FO membrana na zadržavanje neutralnih organskih molekula. Rad je fokusiran na primjenu metode potpornih vektora (engl. *Support Vector Machines*, SVM) za predviđanje zadržavanja organskih molekula (53) FO membranama. Razvijeni SVM model uspoređen je s dva druga modela: modelom umjetne neuronske mreže i modelom višestruke linearne regresije. SVM model generiran uz radijalnu baznu funkciju pokazao je najbolju vrijednost koeficijenta korelacije u iznosu 0,8526. Vrijednosti koeficijenta korelacije kod modela umjetne neuronske mreže i modela višestruke linearne regresije iznosile su 0,7630, odnosno 0,8723.

Ključne riječi
Metoda potpornih vektora, unaprijedna osmoza, membrane, zadržavanje, organske molekule

Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, Alžir

Izvorni znanstveni rad
Prispjelo 9. prosinca 2022.
Prihvaćeno 17. ožujka 2023.