Chemical Engineering Journal 419 (2021) 129540

Contents lists available at ScienceDirect

Chemical Engineering Journal

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

Artificial neural network modeling of wastewater treatment and desalination using membrane processes: A review

Jasir Jawad [a], Alaa H. Hawari [b,*], Syed Javaid Zaidi [a]

[a] Centre for Advanced Materials, Qatar University, P.O. Box 2713, Doha, Qatar
[b] Department of Civil and Architectural Engineering, Qatar University, P.O. Box 2713, Doha, Qatar

ARTICLE INFO

Keywords:
Neural network model
Desalination
Membrane separation
Wastewater treatment

ABSTRACT

The freshwater scarcity is causing a major challenge due to the growing global population. The brackish water and seawater are the biggest sources of water on the planet. Therefore, using desalination and water treatment techniques, household and industrial demands can be met. Microfiltration (MF), ultrafiltration (UF), nanofiltration (NF), reverse osmosis (RO), membrane bioreactor (MBR), and membrane distillation (MD) are some of the membrane processes used in water and wastewater treatment. Artificial intelligence models, such as artificial neural networks (ANN), have recently become a popular alternative to modeling these processes due to several advantages over the conventional model. Therefore, this paper presents a review of ANN models from the last two and a half decades developed for the membrane processes used in wastewater treatment and desalination. Moreover, a complete procedure for the development of two types of ANN models is provided in the paper. The study also discusses the development strategies and comparison of different sorts of ANN models. These models have been applied to several lab-scale, pilot and commercial plants for simulation, optimization, and process control. This work may aid in the development of new ANN models for membrane processes by considering the recent improvements in the field.

1. Introduction

Potable water is essential for sustaining human life. Freshwater scarcity is a significant problem faced by the world nowadays [1]. This is due to the limited available natural resources along with global challenges such as population growth, industrialization, and climate change [2]. About three-quarters of the Earth's surface is water, out of which 99% of water is available as seawater and brackish water, whereas freshwater only accounts for 1%. The demand for freshwater is fulfilled by producing freshwater using membrane and thermal processes. Membrane processes account for almost 65% of global water treatment efforts, whereas thermal processes and other treatment methods account for 35% [3]. The membrane processes for desalination include reverse osmosis, forward osmosis, and membrane distillation. Microfiltration, ultrafiltration, and nanofiltration are also used for water treatment. UF and NF processes are often used as pre-treatment for the desalination processes such as RO and MD. Membrane bioreactor uses activated sludge technology with MF and UF to treat wastewater and sewage water.

Mathematical models are necessary for the simulation and optimization of the process. Permeate flux, salt rejection, and membrane fouling are some of the most studied parameters in the membrane filtration processes. The theoretical or transport-based models are not able to fully predict these characteristics, possibly due to the assumptions required to solve them [4-6]. Moreover, extensive knowledge of the process is required, along with the calculation of thermodynamic, transport, and membrane properties. In the last two decades and a half, artificial intelligence has stepped in to provide an alternative way of modeling these membrane processes with accuracy, among several other advantages. These AI techniques include ANN [7-10], fuzzy logic [11-15], Adaptive Neuro-Fuzzy Inference System (ANFIS) [16-19], genetic programming [20-23] and support vector machine [24-26].

ANN is one of the most popular machine learning techniques, which is a subset of AI. Neural networks are from a class of 'black box' models as the information about the physical parameters of the process is not required [27]. It creates a relationship between input and output variables. Fig. 1 shows the number of publications on the modeling of the membrane processes mentioned above, using the ANN technique. The trend seems to be shifting more towards ANN models, as seen by the increasing number of studies over the years. In 2019, the maximum

* Corresponding author.
E-mail address: a.hawari@qu.edu.qa (A.H. Hawari).

https://doi.org/10.1016/j.cej.2021.129540
Received 23 June 2020; Received in revised form 10 March 2021; Accepted 22 March 2021
Available online 29 March 2021

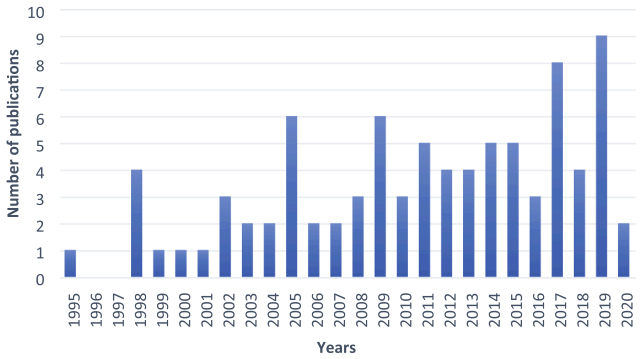

Fig. 1. Number of publications on the modeling of membrane processes in the last 2.5 decades.

number of studies were conducted on water treatment and desalination using the membrane process. This is possible due to certain advantages of ANN over the conventional models. These advantages are given as follows:

- model complicated non-linear functions with high accuracy
- capable of multiple inputs and multiple outputs (MIMO) modeling
- work with noisy and incomplete data
- less computational time
- the ability to update/train the model with new data.

With the advantages, there are also some limitations associated with the ANN modeling, which are given as follows:

- the model parameter (no. of neurons, hidden layers) holds no physical significance
- no standard way to determine the network architecture; trial and error may lead to overfitting or underfitting
- does not provide a unique solution
- an improperly trained network may converge to a local minimum
- computer dependency due to a large amount of data associated with updating the weights.

Over the years, many studies have been conducted on the development of neural network models that can predict the performance of desalination and wastewater treatment using membranes. Several researchers have improved the methodology and combined the ANN model with other techniques such as genetic algorithm or particle swarm optimization to optimize the membrane filtration processes. In this paper, a complete methodology of the development of two types of neural networks (MLP and RBFNN) is presented. Moreover, the paper presents the review of several studies conducted in the past two and a half-decade on the MF, UF, NF, RO, FO, MBR, and MD processes. The conclusion and outcomes of the reviewed publications may assist the development of new ANN models by researchers to simulate and optimize water treatment and desalination plants.

2. Artificial neural network

An artificial neural network is a machine learning technique based on the functioning of biological neurons. The computational model uses a learning process that is similar to the working of a human brain to solve different problems. ANN is one of the most popular tools in artificial intelligence and machine learning. The neural network is a black-box type model as it is not governed by the physical laws, and the model parameters may not hold any physical significance [28]. The main objective of ANN is to develop a relationship between input (independent) and output (dependent) variables. This relationship is established from a learning process in which a data set is provided to the network; hence, it is a data-driven model. Recently, ANNs have become a popular alternative due to their decreased computational time, high accuracy, and ability to non-linearly map relationships between the inputs and outputs of a system. The application of ANNs are not just limited to function approximation but also includes classification, clustering, forecasting, pattern recognition, and image processing.

There are several types of ANNs based on their topology or architecture and model parameters. The most commonly used ANN is a Back Propagation (BP) feedforward neural network, also known as Multilayer Perceptron (MLP). Other types of ANN, such as Radial Basis Function Neural Network (RBFNN), Recurrent Neural Network (RNN), Elman Neural Network (ENN), and Deep Neural Network (DNN), are some of the variations of MLP-ANN with changes in model parameters, structure, and training algorithm.

2.1. Multi-layer Perceptron

An MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer has a certain number of neurons or nodes such that the number of input parameters and output parameters is equal to the number of input and output layer neurons, respectively. One or more hidden layers may consist of several neurons, which is determined based on the optimal architecture of the ANN model. The value of each neuron in the hidden and output layer depends on the weight, which is calculated and assigned during the training phase. In addition, a constant value called threshold value or bias is a weighted unit input to each neuron in the hidden and output layer. The MLP architecture is shown in Fig. 2. The MLP is a feedforward network; therefore, the calculations proceed from the input layer to the output layer without any feedback loops.

In general, there are three phases of building an ANN model: training, validation, and prediction. For this purpose, the dataset must be divided into three different categories that are training dataset, validation dataset, and prediction dataset. The training data consists of the largest number of data points, followed by the validation and prediction dataset. Typically, 70–80% of the data is used for training, whereas the remaining data is used equally for validation and prediction.

The training phase is the first stage of building an ANN model. It is a learning phase where the calculations take place to develop the relationship between the input and output variables. During the training, the weights associated with each neuron are updated after each epoch using one of the training algorithms until a highly accurate network is obtained. The stopping criterion of the training may be specified, such as

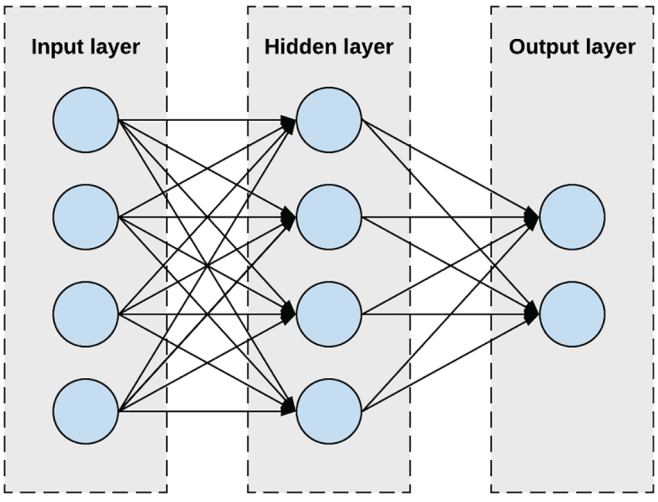

Fig. 2. A typical single hidden layer MLP architecture (4–4-2) having 4 input, 4 hidden and 2 output neurons.

the number of epochs or iterations, minimum error value, and validation checks. Once the trained network is achieved, the validation dataset (unseen data) is used to check the prediction capability of the ANN model. Having several validation checks would also ensure not falling in the local minima during the validation phase. Finally, in the prediction phase, a set of unseen data is used as an input to predict the output parameters using the trained and validated network.

The normalization process before training the network ensures that each input contributes equally to the prediction of the output and minimizes the redundancy [29,30]. The model input and outputs can be normalized using Eq. (1).

$$y = y_{\min} + \frac{(x - x_{\min})(y_{max} - y_{\min})}{(x_{max} - x_{\min})} \tag{1}$$

where x_{max} and $x_{\min}$ are the maximum and minimum value of the data set, y_{max} and $y_{\min}$ are the range for normalization, and y is the normalized value of x. Typically, the range for normalization is either (0,1) or (-1,1). The output of each neuron (Eq. (2)) is based on the previous layer's neuron and its associated weights, which is shown in Fig. 3.

$$a_{ij} = f_j\left(\sum_{k=1}^{n_{(j-1)}}\left(a_{k(j-1)}w_{ki(j-1)}\right) + b_{ij}\right) \tag{2}$$

where a_{ij} and b_{ij} are the output and bias of the i-th neuron in the j-th layer, $a_{k(j-1)}$ and $w_{ki(j-1)}$ are the output and the weight of neuron from the previous layer, respectively, $n_{(j-1)}$ is the number of neurons in the layer $(j-1)$ and f_j is the activation or transfer function of the j-th layer.

The activation function transforms neuron activation by introducing non-linearity to the network [31]. The commonly used activation functions are logistic sigmoid (log-sigmoid), hyperbolic tangent sigmoid (tan-sigmoid), and linear transfer functions (purelin), whose output ranges and equations are given in Table 1.

The Levenberg-Marquardt (LM) method is the most popular BP training algorithm used in the literature due to its convergence speed and performance [9,32,33]. It is a curve-fitting method used to improve the solution by adjusting the learning rate [34]. Few other training algorithms are quasi-Newton, scaled conjugate gradient, gradient descent, Broydon - Fletcher - Shanno (BFGS), and Bayesian regularization. The overall process algorithm for ANN development is given in Fig. 4. The most common indicators used to evaluate the performance of the ANN are Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Sum of Squared Error (SSE). The coefficient of determination (R^2) given by Eq. (3) can be used to evaluate whether the fitted model is able to explain the variability in the response.

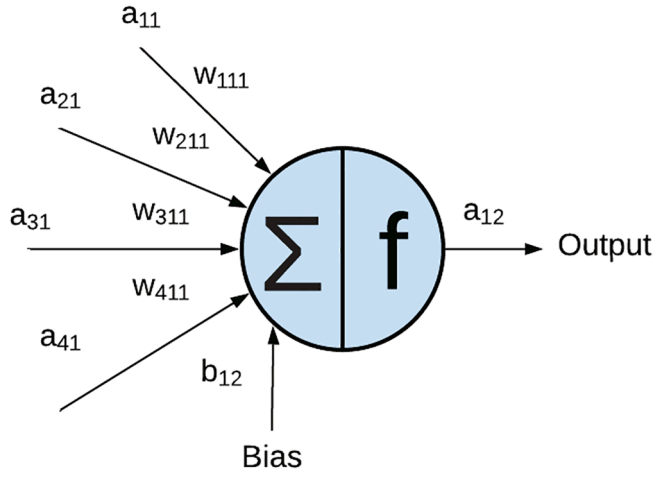

Fig. 3. A close-up of the 1st neuron in the hidden layer (2nd layer).

Table 1
Activation functions used in the MLP network.

Activation function	Equation	Range
Logistic sigmoid	$f(z) = \frac{1}{1+e^{-z}}$	0 to 1
Hyperbolic tangent sigmoid	$f(z) = \frac{e^{2z}-1}{e^{2z}+1}$	−1 to 1
Pure linear	$f(z) = z$	-∞ to ∞

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^t - y_i^p\right)^2}{\sum_{i=1}^{n}\left(y_i^t - \bar{y}\right)^2} \tag{3}$$

where y_i^t is the target output, y_i^p is the predicted output and $\bar{y}$ is the average of target outputs over n number of data points.

The number of hidden layers and neurons is heavily dependent on the system being approximated. Therefore, it is usually a trial and error process of finding the optimum architecture of the neural network model. PSO and GA techniques have also been used to find the optimal number of hidden layers, neurons, and epochs to minimize the error in search of the best architecture.

2.2. Radial basis function neural network

Radial basis function neural network is another type of ANN modeling technique. It is a feedforward network consisting of input, output, and hidden layers similar to an MLP. The output of each neuron is given by a radially symmetric basis function. The RBF network has universal approximation ability without any local minima issues [35]. Furthermore, it has a simpler structure and faster convergence as compared to other neural networks. Instead of using inputs and their associated weights, the distance between input and the center is used as the learning process for the RBFNN calculated using standard Euclidean distance. In this way, the input neurons are mapped to a hidden layer showing a non-linear property. The value of the nodes in the hidden layer is calculated using an activation function. The output of the network is given the same as MLP using the weighted sum of the hidden neuron's output. The number of hidden neurons is determined by the training. At each iteration, one RBF neuron is added to the network, which is followed by the continuous addition of neurons until the error goal or the maximum number of neurons is reached [36]. An RBF neural network is presented in Fig. 5.

The Gaussian radial basis function is most commonly used in the literature [4,35-39]. This activation function is given by Eq. (4).

$$\varphi_i = exp\left(-\frac{\|x_i - c_i\|}{\sigma_i^2}\right) \tag{4}$$

where φ_i is the output of hidden neuron i, x_i is the input parameter, c_i is the center and σ_i is the spread of the neuron. The output value y is given using Eq. (5).

$$y = \sum_{i=0}^{m}\varphi_i w_i + w_0 \tag{5}$$

where m is the total number of neurons in the hidden layer, w_i is the weight associated with neuron i and w_0 is the bias.

The training process for an RBF network consists of two main steps. At first, the centers are selected from the data for training, or the training data is clustered to construct the centers. The next step is the weight calculation between the hidden and the output layer. The spread or radial distance from the center of the RBF neuron controls the width of an area in the input space where the neuron responds. The optimal selection of spread is necessary as it should be large enough so that the neurons can strongly react to the whole range of input data [36]. Using this activation function, each hidden neuron produces a value between

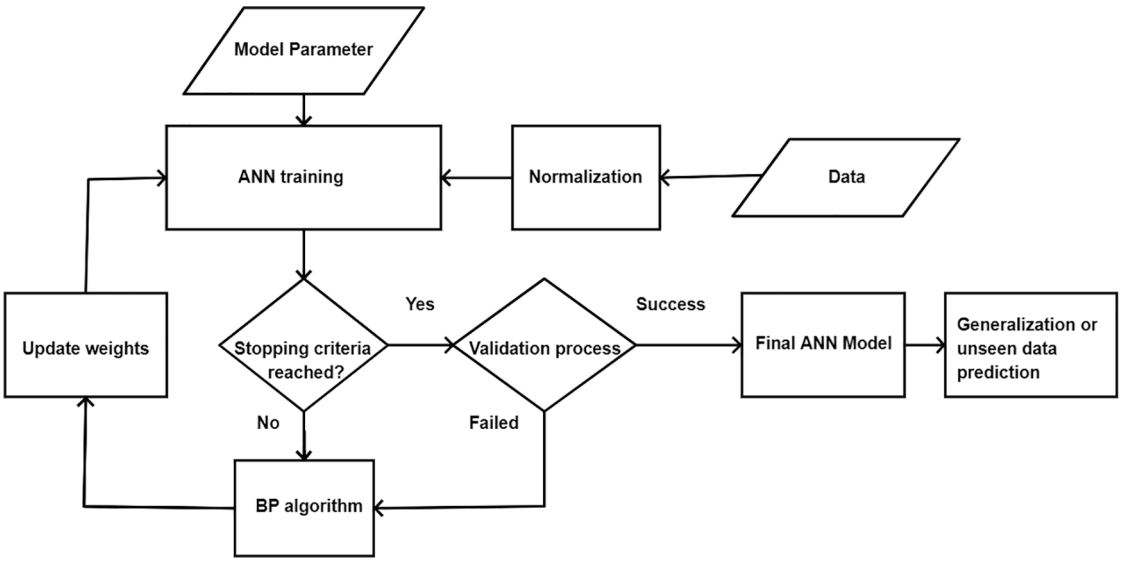

Fig. 4. Neural network development algorithm.

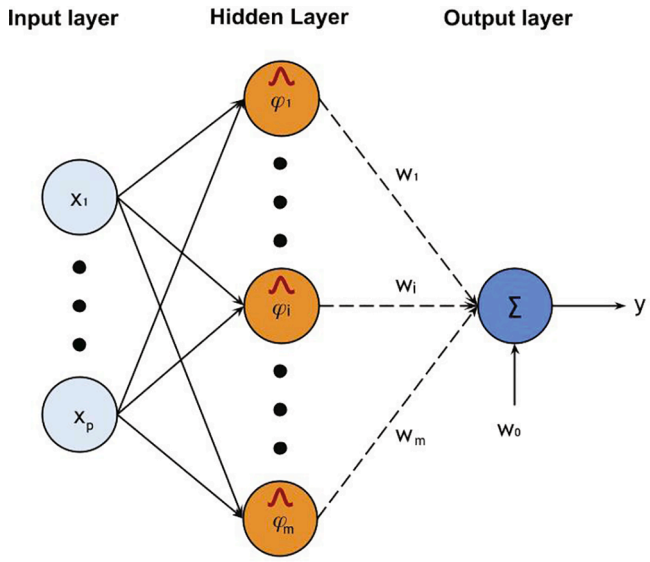

Fig. 5. Typical architecture of RBF neural network model.

0 and 1 depending on the proximity of input to the center. The inputs closer to the center contribute more towards the output. The selection of centers can be made using random selection, K-mean technique, max–min algorithms, or using weight vectors as center parameters [4]. The most common learning technique is self-organized center selection using Orthogonal Least Square (OLS) [36,40]. This method uses the gram-Schmidt algorithm for the selection and update of the center, whereas adaptive gradient descent is used to select the weights [41]. Each RBF center must include at least one input; therefore, the total number of centers should be less than or equal to the number of input parameters. Setting a large value of centers may not improve the performance of the network as the global error may be low, but the network might be overtrained [36].

3. Review of ANN models in the membrane processes

3.1. Microfiltration

Microfiltration is a pressure-driven separation process used to separate fine particles from a liquid. It has various applications in chemical, biochemical, and environmental processes, including water and wastewater treatment. The MF has been effective in the removal of colloidal organic, inorganic solids, and various ions with the aid of macromolecules and surfactants [42]. The nominal pore size for an MF is the range of 0.1 to 1.0 μm [43]. The ceramic-based membranes are often used due to their advantages over polymer membranes, which has a maximum operating pressure of 10 bar [44]. The MF can be operated in two different modes: dead-end filtration and cross-flow filtration. Table 2 presents the characteristics of the ANN model developed for the microfiltration process.

Hamachi et al. modeled crossflow microfiltration of clay solution representing solid particles in water treatment using recurrent neural networks [28]. The permeate flux, along with the evolution of deposit thickness, was calculated based on certain process variables, including the transmembrane pressure (TMP), in addition to the starting values of permeate flux and deposit thickness (state of the process). The recurrent network was set up by first developing a static network to obtain the weight vector, which is later incorporated into the adopted recurrent network shown in Fig. 6.

Aydiner et al. used Koltuniewicz's surface renewal model and neural networks to model phosphate removal from aqueous solution with fly ash [42]. The research presented two neural network models, namely, NN1 (4 inputs) and NN2 (5 inputs), to study flux decline in microfiltration. Based on different network architecture that was tested, the one with the highest R^2 was selected to present the experimental data. It was found that the neural network model was able to outperform Koltuniewicz's surface renewal model based on R^2. Aydiner et al. investigated membrane fouling for the removal of nickel ions by evaluation of fouling parameters and prediction of permeate flux using neural networks [45]. The study focused on the hybrid process, i.e., Powdered Activated Carbon/Microfiltration (PAC/MF) experiments for different membrane types, pore sizes, and surfactant adsorbent. The Taguchi method was used to plan the experiments with the standard L8 orthogonal array matrix [46]. The optimum network R^2 was calculated to be 98.6%. Furthermore, the relative influences from the weights were calculated using a package Neuro Shell 2, which showed that the importance of the input variables was different during various phases in the process. Overall, the surfactant type was the most important, whereas the membrane pore size was relatively the least important variable.

Chellam quantitatively interpreted the weights associated with the input variables in the ANN model to predict transient permeate flux

Table 2
Features of ANN models developed for microfiltration process.

Process	Method	Input	Output	Network Architecture	Activation	Training algorithm	Performance	References
MF of bentonite solution	RNN	TMP, crossflow velocity, concentration of the suspension	deposit thickness, permeate flux	5–7-2	log-sigmoid	Quasi-Newton method	Error < 10%	[28]
Removal of phosphate from aqueous solution with fly ash	MLP	time, concentration, TMP, membrane type	transient flux	5–5–5–1	sigmoid	Qnet's training algorithms	0.991	[42]
MF of polydispersed suspension	MLP	time, feed concentration, TMP, entrance shear rate, initial permeate flux	normalized permeate flux	5–5-1	log-sigmoid	Levenberg–Marquardt	Error < 10%	[47]
brackish water treatment using ceramic membranes	MLP	temperature, turbidity, filtration time, feed pressure, pH	TMP	–	–	conjugate gradients	Error < 10%	[51]
MF of water to study membrane fouling	MLP	permeate flux, raw water turbidity, UV254, operating time, backwash types	TMP	5–5-8–1	log-sigmoid	Levenberg–Marquardt	$R^2 = 0.85$	[50]
Hybrid PAC/MF to study membrane fouling	MLP	time, adsorbent type, membrane type, pore size, surfactant type, concentration	transient flux	6–6-3–1	tan-sigmoid	–	$R^2 = 0.986$	[45]
Treatment of oily wastewater using ceramic membrane	MLP	pressure, feed oil concentration, time	permeate flux	3–5-7–1	–	–	Error < 7%	[52]
Separation of oil from wastewater using mullite ceramic membrane	MLP	oil concentration, feed temperature, TMP, crossflow velocity, filtration time	permeate flux	5–4-15–1	tan-sigmoid	Levenberg–Marquardt	$R^2 = 0.997$	[53]
Separation of starch from starch industry wastewater	MLPRBF	Flow, temperature, pH, concentration	permeate flux, rejection factor	4–1000-700–500-100–2	–	Gradient descent with adaptive learning rate	$R^2 = 0.82$RBF Error: 0.0027	[37]
Separation of particulate suspensions using MF with turbulence promote	MLP	Inlet velocity, TMP, concentration	flux improvement efficiency	3–12-1	log-sigmoid	Gradient descent	$R^2 = 0.9891$	[54]

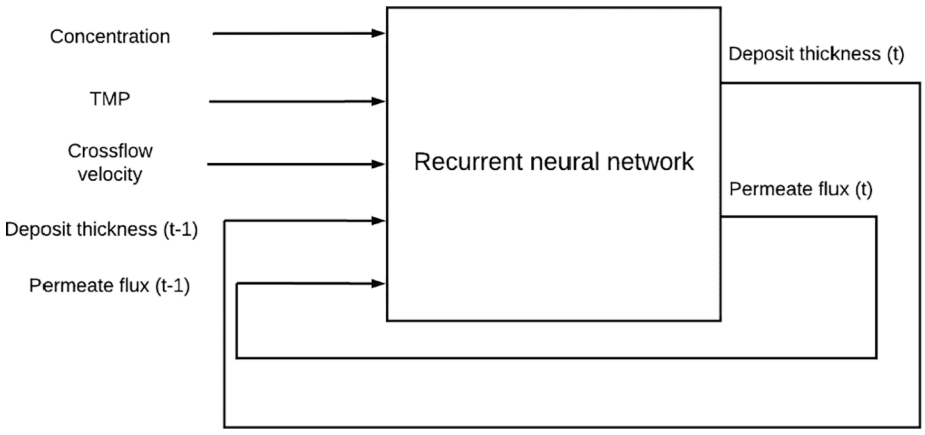

Fig. 6. Recurrent neural network structure to predict deposit thickness and permeate flux [28].

decline caused by polydispersed colloids [47]. The relative importance of the variable in regards to membrane fouling was calculated using the weight partitioning methodology [48,49]. The study found that time and initial flux had a negative influence on the flux, where the importance of each was 20–40% and 23–60%, respectively. A Positive influence was seen from TMP and initial shear rate with 10–22% and 7–15% importance, respectively. Liu et al. presented an ANN model built on experimental data from the hollow fiber membrane module under constant flux mode to treat synthetic water simulating surface water [50]. The study evaluates membrane performance and fouling behavior. The research group also assessed the relative effect or importance of the input variables over the trained network using the arithmetic mean. Time was found to be the most important variable, while the other parameters had a similar impact. The author also concluded that the higher number of training data results in better fitting performance.

Ceramic membrane pilot plant was modeled by Strugholtz et al. using neural networks to predict TMP at the start and the end of the filtration cycle [51]. The model included the influence of coagulation, pretreatment, and Chemically Enhanced Backwash (CEB). A patented method called SecurityNet was used to evaluate the prediction accuracy of the ANN. For this purpose, 5 different neural networks were developed with different weight initializations, and the accuracy was

determined. The predictions from the ANN model were coupled with GA to minimize the operational cost by finding optimum filtration time, flux, and aluminum dosage. In 2010, Nandi et al. evaluated the performance of a low-cost ceramic membrane for filtration of oily wastewater emulsions using different pore blocking and neural network models [52]. Analysis from various pore blocking models suggested membrane fouling due to cake filtration. The ANN architecture consisted of two hidden layers with the optimum neurons in each layer decided by performing simulations at different learning rates. The comparison between the models showed that ANN predicted permeate flux to a higher degree than the cake filtration model. Similar to Nandi and the co-author's work, Shokrkar et al. also modeled a mullite ceramic membrane for the filtration of oily industrial wastewater, adding more input parameters to the network such as TMP and crossflow velocity [53]. The study presented a comparison with different sizes of data for training, and the optimal network's prediction efficiency (R^2) of permeate flux was 99.8%.

Sargolzaei et al. used backpropagation and RBF neural networks to predict the performance of the plate and frame membrane module for the treatment of starchy wastewater [37]. For MLP, the authors tested the Gradient descent algorithm with and without momentum factor along with various network architectures to find the optimal model. The best network used the gradient descent algorithm without momentum and a high number of neurons in the hidden layers. The authors also implemented an RBF neural network and found that network performance was better at a spread constant value of 1. Liu et al. developed an ANN model for the microfiltration process to predict Flux Improvement Efficiency (FIE) by turbulence promoter [54]. Moreover, the operating conditions were optimized to generate maximum FIE while using a turbulence promoter. To determine the best architecture for the network, the researchers conducted various studies to find the optimum transfer function and neurons in hidden layers based on trial and error. The feed concentration shows a positive effect on the FIE, whereas other inputs had a mixed influence on efficiency.

For microfiltration, MLP-ANN has been used the most and shows a lack of research with other types of ANN models such as RNN and RBF, with only one study each. Therefore, this is one area to be explored to check whether other types of ANN can perform better than an MLP-ANN. Weights associated with neurons can be interpreted for meaningful insights using the weight partitioning method or arithmetic mean. The comparison research on different weight analysis techniques may enhance the understanding of the developed model itself and help evaluate the relative importance of the variables more clearly. Although ANN has evidently shown that it is superior to transport-based models in terms of prediction accuracy, however, the comparison of computational cost and resources (such as experimental setup, size of the data for training, and generalization capabilities) should be able to highlight the true advantage of ANN over conventional modeling techniques.

3.2. Ultrafiltration

Ultrafiltration is also a membrane-based separation process to remove heavy metals, suspended solids, and macromolecules. Removal of organic and inorganic contaminants from the wastewater is one of many applications of UF [55]. Typically, the pore size for a UF membrane varies 1 nm to 0.05 μm [44]. The UF is mostly operated tangentially as the feed sweeps away the upper layer of the membrane, thereby increasing the flux and filter life. It is also used as pre-treatment to other desalination processes such as reverse osmosis. Table 3 presents the characteristics of the ANN model developed for the ultrafiltration process.

In 1995, Niemi and the group developed the MLP model for the ultrafiltration of the bleach plant effluent, where the experimental data was obtained from a tubular module [56]. The research showed 12% mean relative deviation for the non-trained data, while training to all data was at 2% deviation. Moreover, the computational time was decreased by 50% for the neural network model compared to mass transfer models.

Delgrange et al. predicted the total hydraulic resistance at the end of the filtration cycle and after backwash to study the effect of membrane fouling on-resistance [57]. The three different networks (R1, R2, and R3) were developed with various input parameters, consisting of water quality parameters and operating conditions. The best network was selected among each category by varying the number of neurons to produce the minimum error. A phenomenological comparison was conducted to find the optimum model by predicting resistance for two cases of variating and stable TMP, which results in R3 as a superior model. In another study, Delgrange et al. predicted the TMP for the same UF pilot setup using a different neural network to be used for process control [58]. The resistances to study membrane fouling were switched with the TMP and followed a similar methodology for the development of the neural network.

ANN may be used along with the conventional models as a hybrid. Cabassud et al. conducted a study that presented a hybrid model consisting of Darcy's law combined with neural networks to predict hydraulic resistances at the end of the ultrafiltration cycle and backwashing [59]. Darcy's law was used to compute the initial resistances from the initial TMP and permeate flow from plant data, which becomes an input to the ANN model for the cycle. Afterward, the resistances at the end of filtration and backwashing (ANN outputs) are used to compute TMP at the end of each process for that cycle using Darcy's law. The algorithm of the hybrid model is presented in Fig. 7. A hybrid model was also presented by Chew et al., which included a cake filtration model based on Darcy's law combined with an ANN model for the water treatment plant [60]. The output-specific cake resistance is fed back to the model as input to calculate the next value. The sensitivity analysis showed a linear correlation between turbidity and resistance, whereas the correlation between the outputs, i.e., resistance and TSS, was minimal.

Teodosiu et al. developed two neural networks that could predict flux at any time instant during ultrafiltration and after backwashing as opposed to the start and end of the filtration cycle [61]. The two models were trained separately in which the 1st model describes flux decline during the ultrafiltration cycle as a function of time and initial flux. The 2nd model takes in the final flux from ultrafiltration and backwashing time to calculate the initial flux for the new filtration cycle.

Bowen et al. used the neural networks to predict permeate flux w.r.t filtration time for the ultrafiltration of silica particles [62]. Ionic strength and zeta potential represent interparticle interaction, whereas the pressure difference is the driving force, which had a strong influence on the flux; therefore, these were selected as the input variables to the network. They were able to predict time-dependent flux profiles using ANNs, while the mechanistic models could only be used for steady-state flux calculations.

Bhattacharya et al. developed ANNs for modeling continuous stirred ultrafiltration to predict permeate flux and concentration [63]. They presented a study for the selection of optimum architecture where learning rate, momentum coefficient, neurons, and the number of hidden layers were manipulated to find a network that generates minimum root mean square error. The network with two hidden layers and a high number of neurons was found to be optimal for the prediction of the two outputs with the prediction accuracy of ± 5%. Oh et al. applied the Kalman neuro training network (KNT) to model hollow fiber UF of groundwater to predict permeate flux [64]. Various KNT architectures were trained, which resulted in the finding of an optimal network with 3 hidden layers and 5 neurons in each layer with an R^2 of 0.9970. Moreover, the mean absolute deviation was reported to be 0.011 and 0.004 in the learning and testing phase, respectively. Kabsch-Korbutowicz and Kutyłowska applied ANN modeling to forecast the turbidity retention coefficient for the coagulation/ultrafiltration of reservoir water [65]. The authors developed 20 different models by changing internal parameters of the neural network such as number of neurons, training

Table 3
Features of the ANN model developed for the ultrafiltration process.

Process	Method	Input	Output	Network Architecture	Activation	Training algorithm	Performance	References
UF of bleach plant effluent	MLP	pressure, tube flow velocity, the concentration ratio of the effluent	rejection of COD, permeate flux	3–8-2	log-sigmoid	Levenberg-Marquardt	Relative deviation = 12%	[56]
UF pilot plant for water treatment	MLP	turbidity, temperature, permeate flow, resistance backwash, resistance end	hydraulic resistance	3–5-2	log-sigmoid	Quasi-Newton learning	Avg. error = 3.6×10^{-5}	[57]
Drinking water production using UF	MLP	permeate flow rate, turbidity, temperature	TMP	3–4-2	log-sigmoid	Quasi-Newton learning	Avg. error = 1.22×10^{-5}	[58]
Separation of silica suspensions	MLP	zeta potential, pressure difference, ionic strength, time	permeate flux	4–10-1	log-sigmoid	–	Avg. error = 3.6%	[62]
Secondary refinery effluent treatment	MLP	time instants, initial flux value	current flux values	2–15-5–1	tan-sigmoid	backpropagation with adaptive learning rate and momentum	Relative error = 3.8%	[61]
Surface raw water filtration using pilot plant	RNN	water quality parameters, operating conditions during filtration time and during the backwash	hydraulic resistance	–	sigmoid	–	–	[59]
Continuous stirred filtration of PEG-6000 solute	MLP	bulk concentration, stirrer speed, pressure, time	permeate concentration, flux	4–30-20–2	log-sigmoid	gradient descent algorithm	Error < 10%	[63]
Pilot plant filtration of groundwater to produce drinking water	MLP	inlet pressure, filtration duration, turbidity, temperature, UV254	permeate flux, TMP	5–5–5–5-2	log-sigmoid	KNT	0.9970	[64]
Filtration of suspended colloids in laboratory scale experiment	MLP/ RBF	particle size, ionic strength, pH, TMP, elapsed time	permeate flux	5–4-2–1	tan-sigmoid Gaussian radial basis	Levenberg-Marquardt	R^2 = 0.958R^2 = 0.988	[4]
Filtration of suspended colloids in laboratory-scale experiment	MLP/ RBF	particle size, ionic strength, pH, TMP, elapsed time	permeate flux	5–7-4–1	tan-sigmoidGaussian radial basis	Levenberg-Marquardt	R^2 = 0.9965R^2 = 0.9935	[36]
Lab-scale filtration of synthetic raw water	MLP	permeate flow rate, turbidity, operating time	TMP	3–5-1	log-sigmoid	Levenberg-Marquardt	$R^2 = 1$	[67]
Removal of zinc ions from wastewater	MLP	TMP, pH, feed concentration, surfactant to metal molar ratio, ligand–zinc ratios, electrolyte concentration, brij35/SDS molar ratio	permeate flux, rejection rate	7–5-2	tan-sigmoid	Levenberg-Marquardt	R^2 = 0.929, 0.983	[8]
Reservoir water filtration using pilot plant	MLP	turbidity raw water, turbidity tank, TMP, temperature, pH, permeate flux	turbidity retention coefficient	6–12-1	tan-sigmoid exponential	Quasi-Newton learning	$R^2 = 0.9216$	[65]
micellar-enhanced UF of synthetic wastewater containing lead ions	MLP	pH, feed concentration, surfactant to metal molar ratio	permeate flux, rejection rate	3–5-2	log-sigmoid	Levenberg-Marquardt	R^2 = 0.9254, 0.9813	[16]
Treatment of refinery wastewater	MLP	temperature, TMP, crossflow velocity, pH	permeate flux, fouling resistance	4–8-2	tan-sigmoid	Levenberg-Marquardt	$R^2 = 1$	[70]
UF of refinery wastewaters in laboratory	MLP	temperature, TMP, crossflow velocity, pH, elapsed time	permeation flux	5–7-7–1	tan-sigmoid	Bayesian regularization	$R^2 = 0.99997$	[6]
Pilot plant filtration of PEG	MLP	TMP, crossflow velocity, time	permeation flux	3–5-1	tan-sigmoid	Levenberg-Marquardt	$R^2 = 0.9977$	[68]
Industrial UF membrane water treatment plant	MLP	turbidity, TMP, filtration time, specific cake resistance	specific cake resistance, TSS	4–5-1	tan-sigmoid	Levenberg-Marquardt	MSE = 9.87×10^{-4}	[60]
Removal of heavy metals from mining wastewater	MLP	S/M ratio, pH, cumulative sampling volume	permeate flux, rejection rate	3–12-2	log-sigmoid	–	$R^2 = 0.99745$	[69]

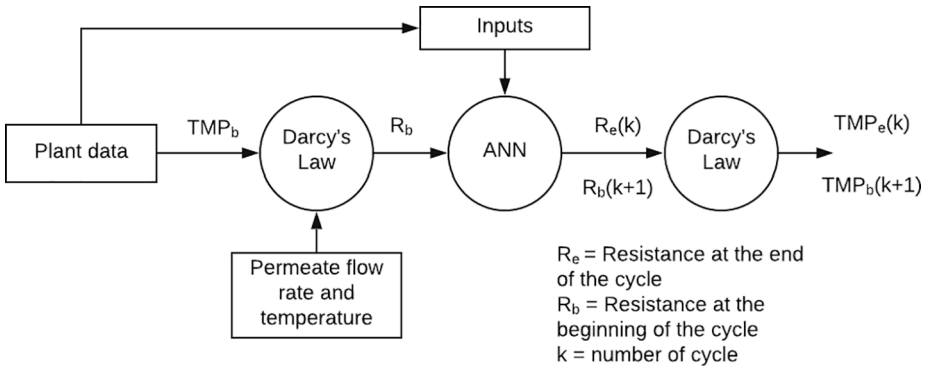

Fig. 7. Algorithm of a hybrid model (Darcy's law + ANN) as presented in [59].

method, and activation functions. The optimum network based on low relative mean square error failed to forecast the output ($R^2 = 0.8295$) with the same efficiency as in the training phase ($R^2 = 0.9216$).

Chen modeled the UF of suspended particles [66] using multiple regression, backpropagation MLP and radial basis neural network to predict permeate flux as a function of ionic strength, TMP, particle size, and pH [4]. The optimal spread value was determined to be 80 through trial and error as spread affects the prediction quality in the RBF network while the maximum neurons were set to 80 as well. A comparison between the three modeling techniques showed that RBF outperformed MLP and regression with an R^2 of 0.988 compared to the latter with the R^2 values of 0.958 and 0.755, respectively. The author also observed relatively less computational time for the RBF network, with 97% of predictions falling in ± 10% of relative error. The experimental data [66], along with similar input/output parameters, were also used by Sahoo et al. to develop MLP and RBF network where the architecture and internal parameters of the networks were optimized using GA [36]. GA-ANN hybrid was used to minimize the error by optimizing spread and maximum neurons for RBFNN, whereas the epoch size and neurons in the two hidden layers were optimized for MLP. The flowchart for the process is given in Fig. 8. This technique proves to be much superior in comparison with trial and error in generating internal parameters for training. Furthermore, the scaling of data before ANN was also investigated, which revealed that scaling helps RBFNN to achieve solution domain for spread. The authors presented that even small datasets can achieve high predictive performance when optimized and trained well as the optimal solution lies in a narrow region compared to training large datasets.

Liu and Kim evaluated the performance blocking laws (cake, intermediate, complete, and standard model), combined blocking laws and ANN models to identify the dominant fouling mechanism for the filtration of synthetic raw water [67]. The study found that blocking laws were unable to predict TMP for the whole operating period. Therefore, the process was divided into three different phases with an individual application of blocking models to each phase, which significantly improved each model's performance. Moreover, the combined blocking laws performed better than single blocking laws. However, the ANN model predicted TMP based on turbidity, flow rate, and operating time, with an excellent agreement to the experimental data and outperforms the other models with an R^2 value of 1. Corbatón-Báguena et al. compared the ANN model to Hermia's fouling model for the filtration of polyethylene glycol (PEG) [68]. The research also presented a comparison on the pretreatment of the data, which showed better ANN prediction accuracy when data were normalized, and a fouling indicator was used as an additional input. Furthermore, the initialization of random weights improved the model's efficiency compared to null initialization. Compared with the classical fouling models, ANN was found to be more powerful with the predictions.

Rahmanian et al. applied ANN modeling to micellar-enhanced ultrafiltration for the removal of zinc ion from synthetic wastewater [8]. A fractional factorial design was used to screen the experiments to include only the significant subset of full factorial experiments. ANOVA analysis of the output variables of the model showed that the variability in the data was explained by the model. However, it also pointed out significant curvature in the design, which could not be predicted using linear regression models. The ANN model consisted of 7 input variables and 2 output variables, namely, permeate flux and rejection rate, for which the R^2 was calculated to be 0.929 and 0.983, respectively. In 2012, a similar model with fewer input variables was constructed by Rahmanian et al. to predict permeate flux and rejection rate for the removal of lead ions [16]. The process was first analyzed using response surface methodology to design the experiments, which were modeled and compared using the ANN and ANFIS techniques. Lin et al. also modeled the micellar enhanced ultrafiltration process for the removal of copper, nickel, and cobalt ions [69]. The authors presented a novel method combining ANN with Monte Carlo simulation to train 1000 ANN models for each metal. Relative importance for each input was also calculated, which showed pH and S/M ratio had more importance than sampling volume for

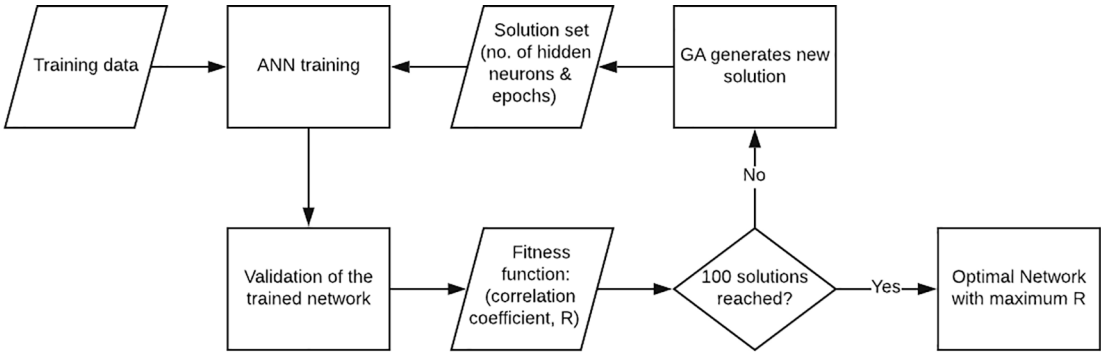

Fig. 8. Flow chart to describe the use of a genetic algorithm to find the optimal network parameters for MLP [36].

rejection rate and permeate flux predictions.

Soleimani et al. created two ANN models for the prediction of permeate flux and membrane resistance during the ultrafiltration of oily wastewater [70]. They studied the effect of TMP, crossflow velocity, pH, and temperature on the two outputs. The model was coupled with the genetic algorithm to predict optimized operating conditions for minimum resistance and maximum flux. A similar study was conducted by Badrnezhad and mirza, who developed a hybrid ANN-GA for modeling the filtration of refinery wastewater and optimization of the process [6]. The ANN architecture (7 neurons in 2 hidden layers) was selected based on trial and error, having the least mean squared error. The model, coupled with GA, was able to predict optimum operating parameters and conditions. Furthermore, the research presented sensitivity analysis using Spearman Rank Correlation, Gamma Correlation, Kendall's Tau, and Pearson Correlation techniques, which all pointed towards filtration time being the most important, followed by temperature, crossflow velocity, pH and TMP.

In general, ANN can take a wide variety of inputs and produce multiple outputs. In a special type of ANN such as RNN, the output from the previous calculation may be used as input. Therefore, solving the problem of developing multiple ANN models. GA incorporation allows for a formal search of the optimum network. Using trial and error is time-consuming, and even then, there is a possibility that the optimum is a local minimum in terms of error. A genetic algorithm eases the search for the optimum and reduces time, especially in the case of multiple hidden layers, by widening the search range and selecting random networks to be tested. Most studies used weights of the network to calculate the influence of the input conditions. Some other techniques to analyze sensitivity are Spearman Rank Correlation, Gamma Correlation, Kendall's Tau, and Pearson Correlation. Incorporating one of these

Table 4
Features of the ANN model developed for the nanofiltration process.

Process	Method	Input	Output	Network Architecture	Activation	Training algorithm	Performance	References
Membrane separation of inorganic salts using a pilot plant	MLP	applied pressure, pH, Ion diffusivity, Ion valence, concentration	ion rejection	14–12-4	log-sigmoid	Scaled conjugate gradients	Avg. error = 1.751×10^{-5}	[72]
Drinking water production by filtration of ground and surface water	MLP	permeate flux, feed water recovery, element recovery, feed water TDS, pH, target contaminant concentration, diffusivity coefficient	ratio of permeate concentration to feed concentration	7–8-4–1	log-sigmoid	Levenberg–Marquardt	Relative error within 10%	[74]
Pilot and full-scale filtration of municipal drinking water	MLP	influent flow rate, feedwater flow rate, permeate flux, operation time, pH, TDS, UV254, temperature	membrane resistance	8–8-1	log-sigmoid	Levenberg–Marquardt	Relative error = 5%	[75]
Full-scale membrane filtration of groundwater	RBF MLP	feed water quality, system flux, recovery	TDS	3–2-1	Gaussian radial basis tan-sigmoid	–	$R^2 = 0.99$ $R^2 = 0.99$	[38]
Water softening using NF	MLP	TMP, concentrations of Ca, Cl, Na, polyethyleneimine, polyacrylic acid	volume flux, rejection	6–8-4	log-sigmoid	–	Error < 3.61%Avg. error = 1.06%	[76]
Filtration of sodium and magnesium chloride solutions	MLP	feed pressure, permeate flux, concentration	rejection	3–4-1	log-sigmoid	Bayesian Regularization	Absolute deviation < 5%	[77]
Potassium and sulfate salts rejection by NF membrane	MLP	feed pressure, permeate flux, concentration	rejection	3–4-1	log-sigmoid	Conjugate gradient backpropagation with Polak-Ribiére	Absolute deviation < 8%	[73]
Treatment of leather plant effluent	MLP	molecular weight cut off, TMP, time, total dissolved solids, cross-flow rate	permeate flux	5–6-5–1	log-sigmoid	gradient descent algorithm	Absolute error < 1%	[78]
Removal of organic micropollutants by NF	MLP	membrane salt rejection, molecular length, equivalent width, hydrophobicity	rejection	4–2-1	tan-sigmoid	Levenberg–Marquardt	$R^2 = 0.97$	[80]
Removal of organic neutral and ionic compounds	MLP	MW, log D, dipole, length, eqwidth, MWCO, salt rejection, Zeta potential, contact angle, pH, pressure, recovery	rejection	12–24-1	tan-sigmoid	BFGS quasi-Newton backpropagation	$R^2 = 0.9527$	[79]
Separation of organic compounds	BANN	MW, log D, dipole, length, eqwidth, MWCO, salt rejection, Zeta potential, contact angle, pH, pressure, recovery	rejection	–	log-sigmoid, tan-sigmoid, exponential function	BFGS quasi-Newton backpropagation	$R^2 = 0.98622$	[81]
Filtration of sodium and calcium chloride	DNN	Initial fouling with corresponding × coordinates, membrane type, time, initial flux, fouling layer axial images	fouling thickness, flux	–	concentrated rectified linear unit (CReLu)	–	$R^2 = 0.9833$	[82]

techniques in the development of the ANN models will help figure out the role of the input variables on the model. Furthermore, developing several ANN models with a different set of input neurons may also help indicate the factors that influence output the most. Literature shows multiple studies in which ANN was combined with conventional models. These hybrid models show the capability of ANN to integrate with Darcy's law to study membrane fouling where resistances were calculated by the ANN. An in-depth comparison of hybrid models with completely mass transfer models, in terms of accuracy and computational cost, will further support the use of such models for control processes.

3.3. Nanofiltration

Nanofiltration is a pressure-driven separation process that lies between ultrafiltration and reverse osmosis. In general, the pore size range for NF is from 1 and 10 nm [71]. The features of NF include high rejection of multivalent ions, moderate rejection of monovalent ions, and high rejection of the organic compounds. It has great potential due to its application in fine chemicals, pharmaceutical, and water treatment industries. The NF systems generally operate at a lower pressure than the RO process producing a high permeate flow rate. It is also used in combination with RO as hybrid or pre-treatment applications. Table 4 presents the characteristics of the ANN model developed for the nanofiltration process.

Bowen et al. developed ANN to predict the rejection of ions (single salts and salt mixtures) from a nanofiltration process [72]. Apart from operating parameters, the authors also considered the concentration, ion valence, and diffusivity of each ion as input variables in their study. A unique ANN methodology is presented in the paper where the model can switch on and off a certain number of input neurons depending on the availability of the data. The study also found that the mixture of monovalent and divalent ions was represented better through the ANN model as compared to physics-based models. Zhao et al. predicted Total Dissolved Solids (TDS) and compared using solution diffusion, modified solution diffusion, MLP, and RBFNN models for an RO/NF process [38]. The diffusion models over or under-predicted the TDS while the ANN models were superior in performance. Al-Zoubi et al. developed an ANN model to investigate the performance of nanofiltration, which is compared with predictions from the Spiegler-Kedem model [73]. The study involved the separation of single salts using 3 different NF membranes. While the Spiegler-Kedem was able to able to accurately predict for all except for the case of KCl, the ANN model showed acceptable performance with a deviation under 8%.

Shetty et al. applied ANN modeling to a set of nanofiltration experiments, which could be universally applicable to municipal water filtration [74]. The model was used to predict the ratio of permeate concentration to feed concentration using operating and water quality parameters for deterministic and pseudo-stochastic simulations. This work was later extended by Shetty and Chellam to study membrane fouling in nanofiltration [75]. With the development of a similar ANN model, the researchers were able to calculate total hydraulic resistance for the process. The filtration data was gathered from bench-scale experiments consisting of flat sheet and single spiral wound membranes, municipal water treatment full scale, and pilot plants. The study showed that accurate prediction of membrane fouling required only 10% training data as compared to 40% for permeate quality. The author argued that ANN is a powerful alternate where techniques such as linear regression may be insufficient for the calculation of fouling rates for a changing feed water quality.

Mousavi et al. presented an ANN model to predict permeate flux and rejection for a nanofiltration process in the presence of polyelectrolytes [76]. The data used for modeling had been divided between studying the effect of pressure drop, the concentration of calcium ion, concentration of polyethyleneimine, and polyacrylic acid for the water softening process. The performance of the model was evaluated using normalized bias value, i.e., the mean difference between predicted and experimental data divided by the experimental data. The results showed that the model under-predicted the outputs in all cases; however, the error was insignificant.

Darwish et al. used ANN to compute rejections from pressure and permeate flux for three different nanofiltration membranes [77]. They found that the conjugate gradient, quasi-Newton principles, and regularization training algorithm performed better than the other training algorithms. The study used the Bayesian regularization training algorithm because of its faster computational time and stability towards random initial guesses for weights and biases. The results showed good prediction capability of the model to predict flux and rejection except for a few cases where the efficiency was unaffected even after increasing the training data.

Purkait et al. proposed the NF/RO plant for the treatment of leather plan effluent with an ANN model to investigate the dynamic performance of the membrane processes [78]. The authors used molecular weight cut-off as one of their input variables to distinguish between different NF membranes. Other input variables included the effect of TMP, TDS, and flow rate on the permeate flux. The developed ANN predicts the output with a mean absolute error of<1%. Ammi et al. applied 3 ANN models to predict the rejection of neutral and ionic organic compounds to study the effect of the molecular weight cut-off, salt rejection ($MgSO_4$), and both on the model performance [79]. The model with a molecular weight cut-off alone was unable to explain the variability efficiently, whereas the other models with $MgSO_4$ salt rejections were effective in predicting the experimental trend. The research also identified that recovery, molecular weight cut-off, equivalent width, and salt rejection ($MgSO_4$) carried the most importance based on input weights analysis.

Yangali-Quintanilla et al. developed an ANN model based on the information from Quantitative Structure-Activity Relationship (QSAR) to predict organic salt rejection [80]. With the help of the correlation matrix and principal component analysis, the number of original molecular descriptors is reduced. The final model inputs consist of dipole moment, molecular length, and equivalent width and a hydrophobicity descriptor. Moreover, salt rejection of magnesium sulfate is used as a parameter instead of the molecular weight cut off to incorporate size hindrance and electrostatic repulsions effects. Multiple models were trained with a different number of inputs where the model without salt rejection ($MgSO_4$) parameter showed the worst performance. Using different transfer functions in the output layer, the number of neurons and training algorithms had an insignificant effect on network performance. Furthermore, the authors determined that the equivalent width was relatively the most influential parameter, while hydrophobicity presented the least important factor.

Khaouane et al. introduced a novel bootstrap aggregated neural networks (BANN) to predict the rejection of ionic and neutral organic compounds for NF and RO membranes [81]. The author used NaCl salt rejection as the porosity indicator. The method involved stacking 30 different neural networks to predict the rejection by averaging the result from each model. Fig. 9 represents the schematic of BANN. The advantage of this technique was seen when the model was applied to unseen data. Whereas single ANN, multiple linear regression (MLR), and

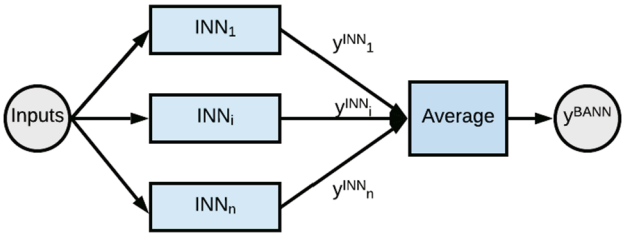

Fig. 9. Schematic of bootstrap aggregated neural network (BANN) [81].

BAMLR failed to forecast and predict the unseen data with high accuracy, BANN outperformed the other techniques and turned out to be superior. Park et al. presented a study to predict flux decline and organic membrane fouling growth by applying DNN using 13,708 high-resolution images of the fouling layer [82]. The pictures were taken using optical coherence tomography (OCT), and the resulting DNN model was compared with the Faridirad model and pore and cake model to estimate fouling thickness and flux declines. The DNN model showed higher accuracy in the prediction of flux and the buildup of the fouling layer.

Many times, when modeling using AI techniques, the user is faced with incomplete or missing data. One solution to this problem is using ANN with a unique ability to switch between input variables depending upon their availability. This methodology enables the user to model a process when there is missing data with one of the input variables. Such ANN models are very scarce and need further investigation of their capabilities. In general, the diffusion models are the most widely used for RO/NF processes. Therefore, comparison with ANN in terms of computational cost and forecasting ability would provide more insights to the consumer for its superiority against diffusional models. The model performance and training efficiency are different for every case in the development of an ANN model. Depending on the quality and type of data to be predicted and the input variables, the percentage of data used for training may differ. Levenberg–Marquardt is the most commonly used training algorithm in the literature. However, studies have indicated that conjugate gradient, quasi-Newton principles, and regularization algorithm outperformed other training algorithms. The discrepancy encourages more studies on the selection of training algorithms based on mathematical stability and computation time. ANN is a powerful tool such that for modeling different membranes for predicting single or multiple outputs, only one ANN model is enough. This is possible if molecular weight cutoff is used as one of the input variables to develop a single ANN model.

3.4. Reverse and forward osmosis

The natural flow of water from dilute feed to the concentrated side due to osmotic pressure difference is called direct or forward osmosis (FO). Reverse osmosis is a pressure-driven membrane separation process used to remove solutes such as salts and low molecular weight organic compounds while allowing the water to pass through the membrane. A pressure greater than the osmotic pressure is applied on the concentrated side for the water from permeating to the dilute side of the membrane. The separation range of RO is from 0.0001 to 0.001 μm [83]. RO has many applications, from wastewater treatment and desalination to food processing. Spiral wound and hollow fiber are the two main module configurations used in the desalination process. The RO and FO process are shown in Fig. 10. Table 5 presents the characteristics of ANN

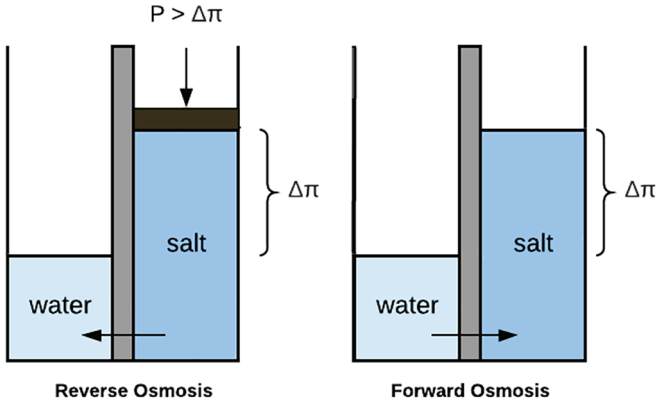

Fig. 10. Concept of reverse and forward osmosis process.

models developed for the reverse and forward osmosis process.

Al-Shayji presented ANN modeling of a reverse osmosis desalination plant [84,85]. The study compared two multiple-input single-output (MISO) and a multiple-input multiple-output (MIMO) network architectures. Therefore, using one model to predict both outputs is more reasonable. Furthermore, input variables were selected based on engineering know-how, whereas a factor analysis reduced the number of variables based on the most significant dependent variable. The comparison in the study showed that engineering know-how (4 variables) showed a more accurate prediction than a smaller number of inputs.

Jafar et al. modeled two RO desalination plants using MLP and RBFNN using 8 input variables to predict the permeate flow rate and TDS [86]. Data clustering and histogram equalization techniques were used to determine and select the centers and width of the RBFNN. A redistributed RBFNN presented more accurate results, high performance with a smaller number of neurons and weights associated with the network than the MLP and fixed RBFNN. Aish et al. predicted permeate TDS and flow rate of the Gaza strip RO desalination plants using MLR, MLP, and RBFNN models [39]. Different training algorithm was tested for MLP, among which Levenberg-Marquardt showed the best performance. The RBF network is trained using orthogonal least squares and Gaussian radial basis functions. All three techniques showed excellent agreement with the experimental data for predicting TDS; however, the least error was observed for MLP for predicting permeate flow. The statistical analysis from the simulation revealed a strong correlation between outputs and the electrical conductivity and a negative correlation with the feed pressure. Furthermore, to evaluate the predictive capacity models, they were used to forecast the data one week ahead, which was successfully verified by observations.

Murthy and Vora applied ANN modeling to a separation of NaCl and water system using RO to predict permeate flux and solute rejection [87]. The flow rate was predicted within ± 1% except for initial and final value at low pressure, while the rejection was within ± 1.5%. Abbas and Al-Bastaki proposed a neural network model for seawater and brackish water RO, considering a range of operating conditions to generalize the model [88]. To check the interpolation capability, the authors predicted the permeate flux for the untrained data using the optimum model with an R^2 of 0.989. However, when it came to extrapolation, the model's predictions proved to be inaccurate.

Lee et al. developed an ANN model for a seawater RO desalination plant using its operational data for one year and optimized operation using temperature control [89]. The model was able to predict permeate TDS with higher accuracy than the flow rate due to the variance difference between the outputs. The effect of temperature studied through the model enabled the researchers to define an optimization strategy that involved linear increase control for feed temperature to achieve minimum permeate TDS. Madaeni et al. investigated the use of a neural network to model the performance degradation of three RO plant operations for process control and optimization [33]. An MLP was used to predict permeate flow and conductivity. The GA optimization was used with ANN simulation to minimize the permeate conductivity of all three plants. Based on the technique, the study presented control strategies using the optimum path for TMP and feed flow rate to maximize the performance of the plants. Cabrera et al. developed an ANN model for a wind turbine-operated RO desalination plant to investigate the impact of varying electrical power supplied to control feed flow and operating pressure setpoint [90]. 10-fold cross-validation and GA optimization were used to select the optimal network architecture, including the number of hidden layers and neurons. The ANN is trained using the data divided into 16 power blocks and 4 levels of temperature and conductivity of feed water. The ANN predicted flow and pressure setpoint based on the varying power supplied from the wind turbine keeping the recovery rate within a particular range to control the desalination process.

Righton developed a neural network model for the RO treatment of groundwater and wastewater [91]. In addition, two RO desalination plants were also modeled using the ANN technique, which accurate

Table 5
Features of ANN models developed for reverse osmosis process.

Process	Method	Input	Output	Network Architecture	Activation	Training algorithm	Performance	References
RO desalination plant	MLP	feed temperature, feed pressure, feed flow rate, feed pH	product flux, product conductivity	4–30-15–2	tan-sigmoid	Gradient-descent algorithm	$R^2 = 0.916704$, $R^2 = 0.95122$	[84]
Groundwater desalination plant	RBFN	–	permeate flow rate, permeate TDS	–		–	Error = 1.73Error = 2.32	[86]
Lab-scale separation of sodium chloride - water system	MLP	concentration, pressure, flow rate	rejection, flux	3–10-10–2	log-sigmoid	Levenberg-Marquardt	Absolute error < 1%	[87]
RO water desalination unit	MLP	feed pressure, temperature and salt concentration	permeate rate	3–5-1	log-sigmoid	Levenberg-Marquardt	$R^2 = 0.998$	[88]
Seawater desalination plant	MLP	feed temperature, feed total dissolved solids, TMP, feed flow rate, time	permeate TDS, flow rate	5–12-2	log-sigmoid	Back propagation algorithm	$R^2 = 0.96$, 0.75	[89]
Separation of sodium chloride and calcium carbonate	MLP	pressure, various feed concentrations, flow rates, concentration of the permeate and reject streams	permeate flux, solute rejection percentages	–	log-sigmoid	Levenberg-Marquardt	$R^2 = 0.9989$	[91]
Brackish water desalination plant	MLP	flow rate, conductivity, feed pressure, pH, temperature	permeate flux, salt passage	5–3-1	tan-sigmoid	Levenberg-Marquardt	Mean absolute error = 0.9%	[92]
RO desalination pilot plant	MLP	Feed concentration, temperature, flow rate, pressure	permeate flux, rejection	4–5-3–1	log-sigmoid	Levenberg-Marquardt	$R^2 = 1$	[93]
RO desalination	MLP	pressure, temperature, concentration, pore radius, friction constants, potential parameter, ratio of average pore length to fractional pore area	separation factor, solvent flux, total flux	9–20-3	tan-sigmoid	Levenberg-Marquardt	$R^2 = 0.9966$, 0.998, 0.9982	[95]
RO desalination using artificial groundwater	MLP	pH, feed temperature, pressure, concentration	water recovery, TDS rejection, specific energy consumption	4–5-3	log-sigmoid	Back propagation algorithm	$R^2 = 0.9611$	[94]
RO desalination process	MLP	time, salinity, operating pressure, membrane type	water permeability	10 hidden layers with 4 neurons	tan-sigmoid	Levenberg-Marquardt	$R^2 = 0.9964$	[9]
RO wastewater pilot plant	MLP	inlet concentration, TDS inlet concentration, time	permeate flow	3–4-3–1	tan-sigmoid	Levenberg-Marquardt	$R^2 = 0.998$	[96]
Large and small-scale brackish water desalination	RBFNMLP	water temperature, pH, conductivity, pressure	TDS, permeate flow	4–6-1	Gaussian kernel/tan-sigmoid	Levenberg-Marquardt	$R^2 = 1$, $0.9873R^2 = 1$, 0.9904	[39]
RO water treatment in a power plant	MLP	time, TMP, conductivity, flow rate	permeate flow, conductivity	4–11-5–2	log-sigmoid	Levenberg-Marquardt	$R^2 = 0.94$, 0.99	[33]
RO desalination plant	MLP	production capacity, water flux recovery, energy consumption, price of electrical tariff	operating and maintenance cost	5–4-1	tan-sigmoid	Levenberg-Marquardt	$R^2 = 1$	[98]
Small scale pilot plant seawater desalination plant	MLP	power, temperature, conductivity	Pressure, flow	3–71-17–13–69-13–1	sigmoid	Resilient backpropagation algorithm	Mean absolute error = 0.405% Mean absolute error = 0.867%	[90]
Brackish water desalination plant		temperature, TMP, time, concentration	water flux	4–16-1	–	Levenberg-Marquardt	$R^2 = 0.99475$	[97]
FO of ground water using 2 M NaCl as draw solution	MLP	feed CFV and temperature, draw solution CFV and temperature	reverse solute flux selectivity (RSFS)	4–8-1 4–7-1	exponential	BFGS quasi-Newton backpropagation	$R^2 = 0.9943$ $R^2 = 0.9988$	[29]
Modeling of Lab-scale FO desalination	MLP	membrane type, membrane orientation, molarity feed solution (FS), molarity draw solution (DS), MW, FS velocity, DS velocity, FS temperature, DS temperature	membrane flux	9–25-25–40-1	log-sigmoid, tan-sigmoid, log-sigmoid	Levenberg-Marquardt	$R^2 = 0.973$	[99]
Bench-scale FO unit using different concentrations of NaCl solutions	MLP	osmotic pressure difference, FS velocity, DS velocity, FS temperature, DS temperature	membrane flux	5–10-1	tan-sigmoid	Levenberg-Marquardt	$R^2 = 0.98036$	[101]

predictions for permeate flux and salt rejection. Libotean et al. demonstrated the ability of the ANN model to forecast the performance of an RO plant based on temporary variations in the permeate flux and salt passage [92]. Three forecasting models, i.e., standard time-series correlation (STSC) method, sequential forecast, and marching forecast, were evaluated. The support vector regression (SVR) model was compared to an MLP model. The study showed that it was possible to use the models to forecast with reasonable accuracy using short-term memory intervals up to a 24-hour period.

Khayet et al. built a response surface methodology (RSM) and an ANN model to predict and maximize the performance index of an RO process, which were later compared [93]. The RSM was divided into two models (for low and high salt concentration) as one model was unable to provide generalized but accurate results. Although both types of models showed accuracy in predicting the salt rejection and permeate flux, neural networks proved to be superior as one model was enough to explain the variability as compared to RSM. The optimum values for the input parameters to maximize performance index were determined using ANN and RSM, both of which showed similar results. RSM and ANN modeling were also compared by Garg and Joshi, who predicted water recovery, TDS rejection, and specific energy consumption (SEC) [94]. The optimum input parameter values were found for both models by minimizing SEC and maximizing TDS rejection and water recovery. The results showed ANN predictions were much closer to the validation experiments for the optimum conditions. Li et al. predicted separation factor, pure solvent flux, and total flux for an RO process using neural networks based input parameters of modified surface force-pore flow model [95]. The input data set consisted of membrane properties, model parameters, and friction constants. The methodology resulted in highly accurate results and less computational time.

Barello et al. modeled an RO desalination process to predict water permeability considering the effects of membrane type, salinity, pressure, and fouling [9]. To train the neural network, the authors conducted a study to determine the effect of the number of hidden layers, number of neurons, and transfer function for the selection of the optimum network. Model using tan-sigmoid transfer function along with the least number of hidden layers showed improved performance. The study of the number of neurons in the layer presented oscillatory behavior; however, the least error is found at 20 neurons in the hidden layer. Furthermore, the ANN calculated water permeability was validated against literature correlations. Salgado-Reyna et al. applied ANN modeling to predict the permeate flow rate of a proposed RO pilot plant at 9 different operating and feed conditions [96]. Each case was separately trained to build optimum ANN models. Case 9 showed the highest permeate flow and lowest TDS in the outlet stream, therefore, corresponding to the optimum operating conditions. The case was also trained using different network types as well as network configurations to find the optimum architecture.

Farahbakhsh et al. applied the ANN modeling on an RO process to study membrane's antifouling properties prepared by blending polypyrrole (PPy) coated multiwalled carbon nanotubes (MWCNTs) [97]. Two neural networks were developed for both raw and oxidized MWCNTs-PPy membranes. The model was able to identify that the oxidized MWCNTs-PPy membrane generated higher water flux with a constant trend. Moreover, the agglomeration of molecules on the surface of the membrane was also predicted as the decrease in salt rejection and water flux.

The application of neural networks had been extended by Ruiz-García and Feo-García to model seawater RO desalination plants to evaluate operating and maintenance costs (O&M) [98]. The O&M cost of 12 desalination plants in Spain was evaluated based on production capacity, recovery, and SEC. The model is used to predict and study cost profiles by varying recovery at different plant capacity.

Compared to other membrane processes, studies on the modeling of forward osmosis using ANN are scarce. Pardeshi et al. presented a Taguchi-neural network model to investigate the performance of an FO groundwater desalination process in both active layer facing draw side (AL-DS) and active layer facing feed side orientations (AL-FS) [29]. The researchers predicted reverse solute flux selectivity (RSFS) for flat sheet membrane laboratory-scale experiments where the draw solution was 2 M NaCl. The Taguchi method was used to conduct the design of experiments, which resulted in a set of 16 experiments, including 4 factors at 4 levels for each orientation. ANN was used to find the optimum conditions to minimize RSFS and validated against experimental data. The ANOVA analysis on the data revealed that the draw solution temperature was the dominant parameter in FO performance. Jawad et al. developed an ANN model for a generalized prediction of membrane flux for lab-scale FO desalination [99]. The model was compared to multiple linear regression and published mathematical models, which showed satisfactory performance. The ANN model consisted of multiple hidden layers with a high number of neurons and took 9 input variables into consideration that affected the membrane flux. The study combined several existing studies in the literature on FO lab-scale experiments to train and validate the ANN model. Later, the authors presented a novel combination of the ANN-RSM model developed using the data published by Hawari et al. [100] to study the impact of concentration, temperature, and velocity on the membrane flux [101]. The trained ANN model was used to complete the Box-Behnken Design to create an RSM model. The study showed that in the absence of experimental data in the form of experimental design, an ANN could be used to build an RSM model for optimization of the parameters. This is possible since ANN does not specifically require an experimental design for its development.

ANN modeling has been used extensively to model the RO process in the past few years. The comparison studies showed that MLR is a weak modeling technique for representing the RO process. However, the accuracy and superiority of MLP-ANN and RBFANN are both supported by studies. This means that the efficiency of each model probably depends on the type of data used for training and training algorithm, which in most cases differ from study to study. ANN models showed superiority over RSM models, possibly because of the lack of dependency on experimental design and rigorous training algorithm. A single ANN model is enough for all sorts of data, whereas multiple RSM models may be needed to explain the variability in the data. RSM and GA can be used alongside ANN to minimize or maximize the outputs of the ANN by coupling it with the developed ANN model. GA optimization can also be employed to optimize the network architecture, i.e., the number of hidden layers and neurons, instead of using the trial and error method. Factor analysis can be used to reduce the high number of input variables to the ANN-based on their significance which can simplify the model allowing for more accuracy and less computational time for training. Another important aspect investigated for ANN is its interpolation and extrapolation capabilities. For interpolation, the results have shown great promise; however, extrapolation does not yield reasonable results, which could be due to the lack of enough data for training. Developing ANN using different sizes of the training data and testing for extrapolation efficiency would be a good way to confirm this phenomenon.

3.5. Membrane bioreactor

Membrane bioreactors combine the activated sludge process with an immersed microfiltration or ultrafiltration membranes for wastewater treatment and reclamation. This process replaces gravity sedimentation to produce high-quality effluent for irrigation and industrial use [102]. Due to the efficient membrane performance, the biosolids are retained within the reactor without the concerns for the cleaning and washing out process [103]. MBR is a simplified process that combines aeration, clarification, and filtration into a single compact unit. Hollow fiber is the most commonly used membrane module for the MBR. Table 6 presents the characteristics of the ANN model developed for the MBR process.

Geissler et al. conducted ENN modeling for an MBR filtration process with submerged capillary modules [104]. An ENN consists of a feedback connection from the output layer to the hidden layer that enables

Table 6
Features of ANN models developed for membrane bioreactor process.

Process	Method	Input	Output	Network Architecture	Activation	Training algorithm	Performance	References
Pilot plant MBR with submerged capillary modules	ENN	TMP, rate of TMP change, TMP backwash, filtration time, backwash time, solid retention time, TSS, temperature, Oxygen decay rate	permeate flux	9–55-1	–	–	Avg. deviation < 2.7%	[104]
Pilot treatment of pharmaceutical wastewater using TPAD-MBR	MLP	oxidation–reduction potential, hydraulic retention time, pH, COD loading rate, DO concentration	effluent COD of MBR	5–4-35–4-25–4-1	tan-sigmoid	Gradient descent with momentum and adaptive learning rate	MSE = 0.008546	[106]
Treatment of submarine wastewater using a combined membrane process	MLP	hydraulic retention time, pH, COD loading rate, DO concentration, temperature, pressure, turbidity, TSS	COD, ammonia nitrogen, turbidity, anionic surfactants		log-sigmoid	–	Avg. error COD = 5.14%Avg. error NH_4+-N = 6.20%RMSE turbidity = 2.76%RMSE LAS = 1.41%	[107]
UF of printing and dyeing wastewater	MLP	MLSS, DO, pH, temperature, suction power	permeate flux	5–12-1	–	Gradient descent method/Particle swarm optimization	Avg. error = 3.59%	[108]
Hypersaline oily wastewater treatment using MBR	MLP	TDS, OLR, reaction time, time	COD, TOC, MLSS, oil in sludge	4–9-4	tan-sigmoid	Batch Back Propagation	R^2 = 0.97339	[111]
MBR Sewage treatment plant	MLP	MLSS, total resistance, operating pressure	permeate flux	–	log-sigmoid	Levenberg-MarquardtGenetic algorithm	Avg. error = 0.0740Avg. error = 0.0331	[110]
MBR pilot plant with combined aeration and filtration	MLPRBF	Time, TSS, COD, sludge retention time (SRT), MLSS	TMP, Permeate flux	5–10-2	tan-sigmoidGaussian	Levenberg-Marquardt	R^2 = 0.98, 0.93	[112]
Treatment of combined municipal and industrial wastewater	RBFNN	influent concentration, HRT, MLSS, TDS, pH	effluent concentration	5–5-1	Gaussian function	–	R^2 = 0.99882	[113]
Treatment of petrochemical wastewater	MLP	MLSS, HRT, time	TMP, COD concentration	3–17-2	tan-sigmoid	Levenberg-Marquardt	R^2 = 0.9999	[114]
Sewage treatment	ENN	MLSS, operating pressure, total resistance, pH, COD, temperature	permeate flux	6–10-1	log-sigmoid	Gradient descent method	Error = 0.0509	[105]
Anoxic-aerobic MBR treatment of domestic wastewater	MLP	permeate flux, COD MBR, MLSS MBR, OLR	TMP	4–5-1	log-sigmoid	Levenberg-Marquardt	R^2 = 0.854	[31]
MBR pilot plant	RBFNN	contact angle of three probe liquids, zeta potential, separation distance	interfacial energy	5–8-15–160-15–110-1	Gaussian function	gram-Schmidt & adaptive gradient descent	R^2 = 1	[41]
Different MBR studies in literature	MLP	temperature, permeate flux, MLSS, TMP	fouling resistance	–	–	Levenberg-Marquardt	R^2 = 0.9682	[109]
air-lift multilevel circulation MBR treatment of marine domestic sewage	MLP	COD, HRT, MLSS, pH	effluent COD	4–10-1	tan-sigmoid	Gradient descent with momentum and adaptive learning rate	Avg. error = 3.67%	[118]
Ship Sewage Treatment using Aerobic-Anaerobic Micro-Sludge MBR System	MLPWNN	COD or TN loading rate, pH, F/M, salinity	effluent CODTN	4–3-14–3-1	sigmoidwavelet function	–	Avg. error (MLP) = 7%, 9.72%Avg. error (WNN) = 2.1%, 3.8%	[115]
Treatment of ship sewage using an air-lift multilevel circulation membrane reactor	WNN	influent COD, influent NH4 + -N, salinity	effluent COD effluent NH4 + -N	3–2-1	wavelet	–	Avg. error = 4.04%Avg. error = 7.90%	[117]
Treatment of ship sewage using an	WNN	TN & COD loading rate, HRT, pH, F/M	effluent CODTN	5–4-15–4-1	wavelet	–		[116]

(continued on next page)

Table 6 (*continued*)

Process	Method	Input	Output	Network Architecture	Activation	Training algorithm	Performance	References
air-lift multilevel circulation membrane reactor							Avg. error = 4.1%Avg. error = 7.1%	
Lab-scale MBR	RBFNN	contact angle of three probe liquids, zeta potential, separation distance, surface tensions	interfacial interactions	5–10-1	Gaussian function	Gram–Schmidt algorithm/gradient descent	$R^2 = 0.9744$	[119]

temporal and spatial pattern recognition and learning. The study presented two models, a semi-empirical model that described filtration resistance as a function of operating parameters for permeability decline calculation and an ENN model for predicting the flux. The pilot plant was run with a structured parameter variation leading to a highly accurate ENN model with a 2.7% average deviation. Similarly, Li et al. predicted the membrane flux from an MBR process for the treatment of sewage water using the Elman neural network [105]. The study presented a comparison between MLP and ENN model that showed ENN, due to its recursive nature, was more accurate than the back-propagation network. The feedback from the hidden layer improved the learning process.

Chen et al. presented a study of two-phase anaerobic digestion (TPAD) and MBR in sequence for the treatment of pharmaceutical wastewater, which was later modeled using MLP and linear regression techniques [106]. The study presented a unique use of three MLP used sequentially to predict the effluent chemical oxygen demand (COD) of MBR. The ANN model fitted to the laboratory data resulted in a mean square value of 0.008546, therefore, showing the feasibility of using ANN models to simulate the removal of COD. A similar MLP model was constructed by Chen and the group to simulate a combined membrane process consisting of UF, MBR, security filter, activated charcoal filter, and RO for the treatment of submarine wastewater [107]. Individual ANN models were evaluated sequentially corresponding to the following process variables: filtering precision, backwash time, backwashing intensity, illumination time, backwashing time. The ANN was trained from day 1 to 120 of operational data and validated using data from day 121 to 139. The highest average root mean square error value of 6.20% was obtained for the prediction of effluent NH_4^+-N concentration.

Liu et al. presented a particle swarm optimization-based MLP (PSO-MLP) model for the filtration of printing and dyeing wastewater to study membrane fouling [108]. To reduce the complexity and improve the performance of the neural network, principal component analysis (PCA) was used to reduce the original 12 input parameters to 5 key input parameters by realizing the mapping between the model variables and the flux. The authors proposed an algorithm to combine back-propagation with PSO to update the weights for the ANN model during the training process. The PSO-MLP model maintained high performance and accuracy while the converging time (epochs) was 12 times faster than the traditional MLP model. The research is supported by Hamedi et al., who presented a comparison between ANN-MLP and ANN-PSO models to study membrane fouling in MBR processes by predicting fouling resistance [109]. The data is extracted from multiple literature sources studying the MBR process. The application of a simple MLP model proved insufficient to accurately calculate the fouling resistance with only 0.58 and 0.68 values of R^2. Whereas MLP incorporated with PSO helped optimize neuron's weights and bias that proved to be superior in predicting the resistance (R^2 greater than 0.96). Fig. 11 shows the algorithm of the ANN-PSO to find the optimal weights and biases for the model.

In another study, Li et al. suggested improvements in an ANN model for the sewage treatment plant to predict membrane flux [110]. They identified disadvantages related to a BP neural network, i.e., slow convergence, falling into local minima, and difficulty in determining network architecture. The use of GA optimization showed independence from the initial weight, global search optimization ability, and fast convergence of the ANN training. Moreover, the comparison between the traditional BP network and GA-MLP showed the latter to be precise as well. Pendashteh et al. developed an MLP model for the MBR treatment of hypersaline oily wastewater [111] to predict several output parameters such as COD, total organic carbon (TOC), Mixed liquor suspended solids (MLSS) and oil in sludge. To train the network efficiently, the authors conducted training ANN using different algorithms such as gradient descent, Levenberg-Marquardt, and batch back-propagation (BBP). LM method resulted in oscillating error rates, due to which the convergence time slowed down. Therefore, BBP was found to be the best training algorithm with an R^2 value of 0.97339 for the system in their study. The relative importance of each input variable showed that the organic loading rate (OLR) and TDS had the most and least influence on the model outputs, respectively.

Mirbagheri et al. modeled MBR wastewater treatment using the two types of ANN models that are MLP and RBFNN, to predict the TMP and permeate flux [112]. The RBFNN model was trained with a spread value of 1 and designed in a loop to run 100 times to minimize the error. The network trained using GA for optimizing weights and network parameters illustrated better accuracy than trial and error. The comparison shows MLP to be more accurate (1% error) in predicting the TMP and flux, for both approximation and generalization, compared to a 2–3% error with the RBF model. Furthermore, ANN models were trained using single and join input variables to investigate the importance of each input variable on TMP and flux. Mirbagheri et al. later implemented an RBF neural network model for the filtration of combined municipal and industrial wastewater following a similar procedure [113]. The model was used to predict effluent concentration based on the influent

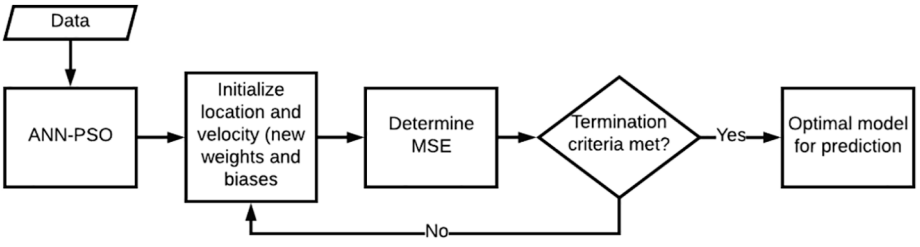

Fig. 11. ANN-PSO algorithm to find the optimal weights and biases for the network.

concentration, hydraulic retention time (HRT), and pH.

Hazrati et al. predicted TMP and COD concentration using the ANN model for the treatment of petrochemical wastewater [114]. A study conducted on 9 different training algorithms to train the MLP model revealed that the LM method had the least relative error and MSE. The sensitivity analysis of the weights of the ANN model found that HRT and MLSS both have a more significant effect on the outputs as compared to time. Schmitt et al. investigated membrane fouling in an anoxic-aerobic MBR process by studying the effect of different input parameters on TMP [31]. 10 consecutive ANN training showed variations in the performance, despite the same architecture, due to different initialization of weights and selection of data for training. Moreover, the relative importance of the variables was different in every training; therefore, the effect of each input type was studied separately by developing an MLP for each case. The analysis revealed that TMP was accurately predicted based on total nitrogen (TN) and total phosphorus concentration than typical parameters such as MLSS, COD, pH, and dissolved oxygen (DO).

Chen et al. developed a highly accurate RBFNN model for interfacial energy related to membrane fouling [41]. The interfacial energy/force is composed of three components: Lifshitz-van der Waals (LW), acid-base (AB) and electrostatic (EL) interactions. The study conducted three RBFNN for each type of interaction based on several input variables. The optimum network at 8 neurons was obtained for LW, while it took 160 and 110 neurons to minimize the error for AB and EL interfacial energies. The model compared to the extended Derjaguin-Landau-Verwey-Overbeek (XDLVO) approach for a case study found high correlation and accuracy in predictions. Zhao et al. conducted a similar study modeling a laboratory-scale MBR process to predict the interfacial interactions related to membrane fouling, also using RBFNN [35]. The authors developed one accurate network with a smaller number of neurons to predict all the interaction energies compared to the previous study by Chen et al. A comparison made with the XDLVO method showed that the computational time for XDLVO was 3 times higher as than the RBF network. A study conducted to show the impact of a number of training data (5, 10, 15, and 20 samples) revealed that the model was more efficient when the network was trained using more than 10 data samples.

Cai et al. modeled the Air-Lift Multilevel Circulation MBR process for the treatment of domestic marine sewage to study the effect of COD, HRT, MLSS, and pH on the reactor performance [115]. The weights of the input neurons were used to determine the importance of each factor, indicating pH and MLSS to be more influential variables. In addition, Cai et al. studied the effect of hydraulic retention [116], the effect of salinity [117], and the effect of pH [118] on the pollutant removal of ship sewage treatment. The authors conducted different studies developing wavelet neural networks to predict effluent COD and TN separately. A comparison study revealed that the WNN model was more accurate in predicting the output (up to 10% error) than the MLP models, with errors reaching up to 20%.

For MBR, ENN has shown better performance than the semi-empirical and MLP-ANN model, possibly due to its recursive nature. However, the literature lacks a comparison with completely physics-based models to identify the advantages of the ANN model over the MBR mathematical model. The unique feature of ANN to be used sequentially with another ANN was also explored. However, the use of ANN in this manner must be justified by comparing the sequential model with a single developed ANN model to justify the need. Similar to factor analysis, PCA may also be used to reduce the number of input parameters to only significant variables. The resulting ANN using this technique has shown high accuracy and prediction power. The comparative study between factor analysis and PCA may be helpful in identifying the best method to reduce the input variables of the ANN. Similar to GA optimization, PSO is another technique highlighted in the literature for the optimization of weights and biases of the neural network. This encourages another comparative study between the selection of GA and PSO for the optimization of weights and biases. LM optimization was compared to several other training algorithms and was found to be the most accurate. However, another study encountered oscillating error rates resulting in a high convergence time. A few other training methods have been suggested by the literature, such as gradient descent, batch backpropagation, BFGS quasi-Newton backpropagation and Bayesian regularization that must also be explored for the development of new ANN models. MBR modeling showed MLP-ANN to be more accurate than the RBFNN in approximation and generalization. The impact of training data was also investigated but only from 5 to 20 data samples. This is a small range to observe the effect of training data on the ANN model. It is recommended that larger sets of data may be trained to check improvement in the ANN training.

3.6. Membrane distillation

Membrane distillation is a relatively new technology used for brackish and seawater desalination using porous membranes. The separation is achieved based on the volatility of the separating constituent. The process is operated at atmospheric pressure and below 100 °C temperature. The driving force for the separation in an MD is the partial pressure difference created across the membrane due to the temperature difference at each side of the membrane [120]. The membrane allows vapor to pass through the pores while stopping the liquid. One side of the membrane is directly in contact with the hot saline water, while the other side is in contact with cold freshwater. Types of MD include direct contact membrane distillation, air gap membrane distillation, sweeping gas membrane distillation, and vacuum membrane distillation. Table 7 presents the characteristics of ANN models developed for the microfiltration process.

Tavakolmoghadam and Safavi investigated the performance of a vacuum MD process using a data-driven ANN model [121]. GA was used to optimize the model parameters such as the number of neurons, epochs, and other model coefficients to obtain the best training level. Cao et al. proposed an ANN model for a vacuum MD to optimize the operating parameters [122]. Using the BFGS algorithm, a set of 38 data samples was used to train the network. The simulation showed the effect of vacuum pressure was significant, followed by feed inlet temperature. Yang et al. also developed an ANN model for a vacuum MD to predict permeate flux and specific energy consumption under different operating conditions and module parameters [119]. Membrane length was used as one of the input parameters. Increasing the membrane length resulted in a decline in flux and an increase in specific energy consumption.

Khayet and Cojocaru studied air gap MD to develop a predictive ANN model to maximize the performance index [123]. Several experiments with a variation in the operating conditions were considered for the training for the network. The developed ANN model was used to optimize the input parameters using Monte Carlo simulation to predict the maximum performance index. The model was also statistically analyzed using ANOVA, which showed that 99% of the variability in the performance index was explained by ANN. The feed inlet temperature had the greatest effect on the response variable. Shirazian and Alibabaei predicted cold feed outlet temperature, distillate flux, and gained output ratio using ANN to model air gap MD [124]. The neutral network was combined with PSO to find the optimal weights for the network. Furthermore, using Volterra functional series theory, a mathematical relationship between input and output was developed. Khayet and Cojocaru also showed that complex configurations such as the sweeping gas MD process could be modeled using ANN to predict the performance index [10]. The optimal operating conditions were determined using a Monte Carlo simulation with the developed ANN model. The ANN technique compared to the RSM model resulted in a 1.5 times more accurate model. According to the study, inlet feed temperature had the most influence on the performance index.

Porrazzo et al. proposed an ANN model to optimize the feedforward

Table 7
Features of ANN models developed for the membrane distillation process.

Process	Method	Input	Output	Network Architecture	Activation	Training algorithm	Performance	References
Lab-scale desalination by vacuum membrane distillation	MLP	vacuum pressure, feed inlet temperature, feed salt concentration, feed flow rate	permeate flux	4–5-1	tan-sigmoid	–	Error < 1%	[121]
Desalination using air gap membrane distillation	MLP	air gap thickness, cooling inlet temperature, feed inlet temperature, feed flow rate	performance index	4–10-1	log-sigmoid	Levenberg-Marquardt	$R^2 = 0.9922$	[123]
Solar-powered membrane distillation plant for seawater desalination	MLP	radiation, feed flow rate, the inlet temperature	distillate flow rate	3–5-1	log-sigmoid	Levenberg-Marquardt	$R^2 = 0.985$	[125]
Desalination by sweeping gas membrane distillation	MLP	feed inlet temperature, feed liquid flow rate, airflow rate	performance index	3–9-1	log-sigmoid	Levenberg-Marquardt	$R^2 = 0.80$	[10]
Seawater desalination using vacuum membrane distillation	MLP	feed inlet temperature, vacuum pressure, feed flow rate, feed salt concentration	permeate flux	4–3-1	tan-sigmoid	BFGS	$R^2 = 0.998656$	[122]
Air gap membrane distillation systems for desalination process	MLP	cold feed inlet temperature, hot feed inlet temperature, feed flow rate	distillate flux, cold feed outlet temperature, gained output ratio	3–5-5–3	–	PSO	$R^2 = 0.9788$, 0.9800, 0.9879	[124]
Permeate gap membrane distillation using renewable energy sources	MLP	evaporator temperature, condensor temperature, feed flow rate	distillate flow rate	3–10-1	tan-sigmoid	Levenberg-Marquardt	$R^2 = 0.94996$	[34]
Desalting brines from Reverse Osmosis plant using Permeate-Gap Membrane Distillation	MLP	evaporator temperature, condenser temperature, feed flow rate, feed salt concentration	permeate flux, specific thermal energy consumption	4–7-2–2	log-sigmoid	Levenberg-Marquardt	$R^2 = 0.991$, 0.990	[126]
Vacuum membrane distillation	MLP	feed inlet temperature, feed flow rate, membrane length	permeate flux, specific thermal energy consumption	3–7-2	tan-sigmoid	Levenberg-Marquardt	$R^2 = 0.9936$, 0.9645	[119]

control system for a solar-powered MD due to its variable source of heat [125]. A dynamic non-recurring neural network was used in the study in which the response is dependent on the current inputs as well as previous values. The application of ANN helped identify the range of operating parameters for the maximum production of distillate. The authors also studied the effect of typical and sudden changes in the feed flow rate and radiation. Acevedo et al. implemented an ANN model on the permeate gap MD integrated into a trigeneration unit based on the hybrid production of desalted water using renewable energy sources [34]. The ANN model improved the predictions as compared to the linear regression model. The developed ANN model was then tested against real plant data, which showed an underestimation of only 1.01 L per day. Gil et al. presented ANN and RSM forecasting models for solar energy-powered commercial permeate gap MD for desalting of RO brine [126]. Both models predicted permeate production and thermal energy consumption as process performance parameters using a typical input parameter in addition to feeding salt concentration. ANN predicted the specific thermal energy consumption with more accuracy than the RSM model due to its non-linear relationship with the feed salt concentration. The study also included multi-objective optimization to maximize permeate production and minimize specific energy by optimizing the operating parameters.

Studies on ANN models for membrane distillation are few compared to the other membrane processes. The studies have shown the successful use of GA and PSO techniques for the optimization of the weights and biases of the ANN model, in addition to the traditional training algorithms. Two studies supported the use of Monte Carlo simulations to maximize the performance index by optimizing the input variables of the ANN model. ANN has been compared with RSM in multiple studies showing that ANN can model non-linearity better than the RSM technique.

4. Future aspects and conclusion

The study presents a review of the development of artificial neural network models for water treatment and desalination processes using membrane separation. The paper also presents a complete description of the MLP and RBF neural network development methodology.

In general, the ANN has shown good generalization capability by predicting the outputs comparable to conventional models or sometimes even better in terms of accuracy and computational time. Accuracy is reported and compared by most of the studies in the literature; however, only a few studies exist that compare the computational costs of ANN against the conventional model. One such study indicated that it reduced 50% of the computation cost for an ultrafiltration model. The comparison of computation cost and resources (such as experimental setup, size of the data for training, and generalization capabilities) may be able to highlight the advantage of ANN over conventional transport-based models. The evaluation of computational cost between different types of ANN would also assist in the selection of the best model for that process. The parameters which were mostly predicted using the ANN model were membrane fouling, permeate quality, permeate flux, and TMP. The application of these models has usually been to investigate the performance of the membrane filtration process. However, it has not been limited to that but also to predict the operational and maintenance cost. The developed ANN models have been combined with other optimization techniques such as particle swarm optimization, genetic algorithm, and Monte Carlo simulations to determine the optimum operating conditions maximizing the performance. To select the best process for optimization, a study is encouraged where the optimization outcomes from all three of the techniques are compared for accuracy and speed.

Most studies have used the MLP type neural network design to model the processes due to its simplicity and accuracy. However, according to some studies, RBFNN is found to be simpler in topology and faster in computation. This contradiction may be due to the type of training data

and training algorithm used to build the ANN models. In terms of ANN development, a few studies have suggested a more systematic way of determining the model parameters such as epochs, number of hidden layers, and number of neurons instead of using trial and error directly. These methods involve using GA and PSO to produce a different solution set that consists of the model parameters, thereby automating the process of finding the optimal architecture. This also helps to avoid settling on a local minima solution. The most popular training algorithm among all the studies is the Levenberg-Marquardt backpropagation algorithm, as it provides fast convergence and accurate results. However, studies show that conjugate gradient, quasi newton principles and regularization algorithm have performed better than LM. Moreover, LM encountered oscillating error rates leading to an increase in the convergence time. These inconsistencies may also be due to the specific training data used, types of inputs and outputs. Future studies may focus on the mathematical stability and computational cost of different training algorithms used for ANN models. Furthermore, other training algorithms such as batch backpropagation, BFGS quasi-Newton backpropagation and Bayesian regularization should be investigated. In addition, several studies have used PSO and GA to update the weight parameters and biases of the network for better convergence. Few studies have shown the success of these techniques in the literature. A comparative study between GA and PSO in terms of accuracy, time consumption, and ease of development will benefit future studies in the selection of the optimization method. The network architecture, activation function, and other parameters are dependent on the type of process and data used to train the ANN model. Therefore, the characteristics of the model need to be tested for the implementation of ANN on different processes.

ANN has been proven efficient in the case of interpolation, whereas some researchers have proven that ANN is not reasonable when it comes to extrapolation. One reason given for such a claim is the lack of training data which results in the ANN. The impact of training was observed from 5 to 20 data samples that suggested more than 10 data samples result in acceptable accuracy. Investigation of ANN for extrapolation should be carried out, especially increasing further the size of the training data to see its impact on the extrapolation capabilities. Factor analysis and PCA have been used by researchers to reduce the number of high input variables to the ANN model to only a few significant ones, thus achieving high accuracy and prediction power. The literature lacks a comparison between these factors or input reduction techniques which could provide necessary insights to make the selection between the two. Moreover, ANN models may be developed with all inputs and reduced inputs to highlight the significance resulting from the reduction.

Another insight gathered from the review for future consideration includes the scarcity of the studies on ANN models for the emerging forward osmosis technology. In other membrane processes, many models have been tested based on experimental data, pilot plant and even commercial plant studies. However, forward osmosis is limited to laboratory-scale experimental data, only consisting of a small number of data set for the development of the ANN model. Conventional ANN models like MLP, RBF or ENN are widely used in many of these studies discussed in this paper. It is also found that some researchers attempted hybrid models and multiple ANN models evaluated sequentially. These models are also scarce and may require more studies to prove their usefulness over conventional techniques.

To conclude, the ANN methodology proves to be a successful alternative to the conventional techniques that use transport-based models. The developed ANN models can be used to simulate, control, and optimize the operations of the membrane processes. In most cases, ANN is only limited to interpolation between the range of the data used to train the network; however, in a few cases, it has been seen to forecast and extrapolate as well. ANN models have often been combined with GA, PSO and other optimization methods to maximize the performance of the process. The review further highlights the challenges in modeling using ANN due to the scarcity of data and studies in the selection of optimization methods and training algorithms. From the review of several studies, some inconsistencies and contradictions have also been identified. These may be due to ANN being a black-box model with different types of training data and algorithms. The information presented in the paper may be helpful for engineers to develop neural networks considering relevant methods used over the years to improve the learning process of the model.

Declaration of Competing Interest

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

Acknowledgments

This publication was possible by an NPRP grant (NPRP10-0117-170176) from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors. In addition, the authors would like to thank Qatar University for the financial support through grant number QUCG-CAM-19/20-4.

References

[1] S. Al Aani, T. Bonny, S.W. Hasan, N. Hilal, Can machine language and artificial intelligence revolutionize process automation for water treatment and desalination? Desalination. 458 (2019) 84–96, https://doi.org/10.1016/j.desal.2019.02.005.

[2] Y.M. Kim, Y.S. Lee, Y.G. Lee, S.J. Kim, D.R. Yang, I.S. Kim, J.H. Kim, Development of a package model for process simulation and cost estimation of seawater reverse osmosis desalination plant, Desalination. 247 (2009) 326–335, https://doi.org/10.1016/j.desal.2008.12.035.

[3] N. Kress, Desalination Technologies, Mar. Impacts Seawater Desalin. (2019), https://doi.org/10.1016/b978-0-12-811953-2.00002-5.

[4] H. Chen, PREDICTION OF PERMEATE FLUX DECLINE IN CROSSFLOW MEMBRANE FILTRATION OF COLLOIDAL SUSPENSION: A RADIAL BASIS FUNCTION NEURAL NETWORKAPPROACH, UNIVERSITY OF HAWAI'I, 2005.

[5] A.L. Wei, G.M. Zeng, G.H. Huang, J. Liang, X.D. Li, Modeling of a permeate flux of cross-flow membrane filtration of colloidal suspensions: A wavelet network approach, Int. J. Environ. Sci. Technol. 6 (2009) 395–406, https://doi.org/10.1007/BF03326078.

[6] R. Badrnezhad, B. Mirza, Modeling and optimization of cross-flow ultrafiltration using hybrid neural network-genetic algorithm approach, J. Ind. Eng. Chem. 20 (2014) 528–543, https://doi.org/10.1016/j.jiec.2013.05.012.

[7] M. Dornier, M. Decloux, G. Trystram, A. Lebert, Dynamic modeling of crossflow microfiltration using neural networks, J. Memb. Sci. 98 (1995) 263–273, https://doi.org/10.1016/0376-7388(94)00195-5.

[8] B. Rahmanian, M. Pakizeh, S.A.A. Mansoori, R. Abedini, Application of experimental design approach and artificial neural network (ANN) for the determination of potential micellar-enhanced ultrafiltration process, J. Hazard. Mater. 187 (2011) 67–74, https://doi.org/10.1016/j.jhazmat.2010.11.135.

[9] M. Barello, D. Manca, R. Patel, I.M. Mujtaba, Neural network based correlation for estimating water permeability constant in RO desalination process under fouling, Desalination. 345 (2014) 101–111, https://doi.org/10.1016/j.desal.2014.04.016.

[10] M. Khayet, C. Cojocaru, Artificial neural network model for desalination by sweeping gas membrane distillation, Desalination. 308 (2013) 102–110, https://doi.org/10.1016/j.desal.2012.06.023.

[11] P. Vaija, J. Randon, A. Larbot, L. Cot, Prediction of the flux through an ultrafiltration membrane using fuzzy mathematics, J. Memb. Sci. 83 (1993) 173–179, https://doi.org/10.1016/0376-7388(93)85265-X.

[12] S.S. Madaeni, A.R. Kurdian, Fuzzy modeling and hybrid genetic algorithm optimization of virus removal from water using microfiltration membrane, Chem. Eng. Res. Des. 89 (2011) 456–470, https://doi.org/10.1016/j.cherd.2010.07.009.

[13] I. Noshadi, A. Salahi, M. Hemmati, F. Rekabdar, T. Mohammadi, Experimental and ANFIS modeling for fouling analysis of oily wastewater treatment using ultrafiltration, ASIA-PACIFIC, J Chem. Eng. (2012) 527–538, https://doi.org/10.1002/apj.

[14] F. Salehi, S.M.A. Razavi, Dynamic modeling of flux and total hydraulic resistance in nanofiltration treatment of regeneration waste brine using artificial neural networks, Desalin. Water Treat. (2012), https://doi.org/10.1080/19443994.2012.664683.

[15] B. Rahmanian, M. Pakizeh, M. Esfandyari, F. Heshmatnezhad, A. Maskooki, Fuzzy modeling and simulation for lead removal using micellar-enhanced ultrafiltration (MEUF), J. Hazard. Mater. 192 (2011) 585–592, https://doi.org/10.1016/j.jhazmat.2011.05.051.

[16] B. Rahmanian, M. Pakizeh, S.A.A. Mansoori, M. Esfandyari, D. Jafari, H. Maddah, A. Maskooki, Prediction of MEUF process performance using artificial neural

networks and ANFIS approaches, J. Taiwan Inst. Chem. Eng. 43 (2012) 558–565, https://doi.org/10.1016/j.jtice.2012.01.002.
[17] A. Rahimzadeh, F.Z. Ashtiani, A. Okhovat, Application of adaptive neuro-fuzzy inference system as a reliable approach for prediction of oily wastewater microfiltration permeate volume, J. Environ. Chem. Eng. 4 (2016) 576–584, https://doi.org/10.1016/j.jece.2015.12.011.
[18] A.R.S. Nejad, A.M. Ghaedi, S.S. Madaeni, M.M. Baneshi, A. Vafaei, D. Emadzadeh, W.J. Lau, Development of intelligent system models for prediction of licorice concentration during nanofiltration/reverse osmosis process, Desalin. Water Treat. 145 (2019) 83–95, https://doi.org/10.5004/dwt.2019.23731.
[19] F. Salehi, S.M.A. Razavi, Modeling of waste brine nanofiltration process using artificial neural network and adaptive neuro-fuzzy inference system, Desalin. Water Treat. (2016), https://doi.org/10.1080/19443994.2015.1063087.
[20] S.J. Kim, S. Oh, Y.G. Lee, M.G. Jeon, I.S. Kim, J.H. Kim, A control methodology for the feed water temperature to optimize SWRO desalination process using genetic programming, Desalination. 247 (2009) 190–199, https://doi.org/10.1016/j.desal.2008.12.024.
[21] J.S. Cho, H. Kim, J.S. Choi, S. Lee, T.M. Hwang, H. Oh, D.R. Yang, J.H. Kim, Prediction of reverse osmosis membrane fouling due to scale formation in the presence of dissolved organic matters using genetic programming, Desalin. Water Treat. (2010), https://doi.org/10.5004/dwt.2010.1675.
[22] H. Shokrkar, A. Salahi, N. Kasiri, T. Mohammadi, Prediction of permeation flux decline during MF of oily wastewater using genetic programming, Chem. Eng. Res. Des. 90 (2012) 846–853, https://doi.org/10.1016/j.cherd.2011.10.002.
[23] R. Goebel, M. Skiborowski, Machine-based learning of predictive models in organic solvent nanofiltration: Pure and mixed solvent flux, Sep. Purif. Technol. 237 (2019), 116363, https://doi.org/10.1016/j.seppur.2019.116363.
[24] H. Fazeli, R. Soleimani, M.A. Ahmadi, R. Badrnezhad, A.H. Mohammadi, Experimental study and modeling of ultrafiltration of refinery effluents using a hybrid intelligent approach, Energy and Fuels. 27 (2013) 3523–3537, https://doi.org/10.1021/ef400179b.
[25] K. Gao, X. Xi, Z. Wang, Y. Ma, S. Chen, X. Ye, Y. Li, Use of support vector machine model to predict membrane permeate flux, Desalin. Water Treat. (2016). 10.1080/19443994.2015.1086691.
[26] H. Adib, A. Raisi, B. Salari, Support vector machine-based modeling of grafting hyperbranched polyethylene glycol on polyethersulfone ultrafiltration membrane for separation of oil–water emulsion, Res. Chem. Intermed. 45 (2019) 5725–5743, https://doi.org/10.1007/s11164-019-03931-z.
[27] M. Al-Abri, N. Hilal, Artificial neural network simulation of combined humic substance coagulation and membrane filtration, Chem. Eng. J. 141 (2008) 27–34, https://doi.org/10.1016/j.cej.2007.10.005.
[28] M. Hamachi, M. Cabassud, A. Davin, M. Mietton Peuchot, Dynamic modelling of crossflow microfiltration of bentonite suspension using recurrent neural networks, Chem. Eng. Process. Process Intensif. 38 (1999) 203–210, https://doi.org/10.1016/S0255-2701(99)00004-5.
[29] 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, https://doi.org/10.1016/j.cherd.2016.01.030.
[30] A. Alver, Z. Kazan, Prediction of full-scale filtration plant performance using artificial neural networks based on principal component analysis, Sep. Purif. Technol. 230 (2020), 115868, https://doi.org/10.1016/j.seppur.2019.115868.
[31] F. Schmitt, R. Banu, I.T. Yeom, K.U. Do, Development of artificial neural networks to predict membrane fouling in an anoxic-aerobic membrane bioreactor treating domestic wastewater, Biochem. Eng. J. 133 (2018) 47–58, https://doi.org/10.1016/j.bej.2018.02.001.
[32] J.J. Moré, The Levenberg-Marquardt algorithm: Implementation and theory, in (1978), https://doi.org/10.1007/bfb0067700.
[33] S.S. Madaeni, M. Shiri, A.R. Kurdian, Modeling, optimization, and control of reverse osmosis water treatment in kazeroon power plant using neural network, Chem. Eng. Commun. 202 (2015) 6–14, https://doi.org/10.1080/00986445.2013.828606.
[34] L. Acevedo, J. Uche, A. Del-Amo, Improving the distillate prediction of a membrane distillation unit in a trigeneration scheme by using artificial neural networks, Water (Switzerland). 10 (2018), https://doi.org/10.3390/w10030310.
[35] Z. Zhao, Y. Lou, Y. Chen, H. Lin, R. Li, G. Yu, Prediction of interfacial interactions related with membrane fouling in a membrane bioreactor based on radial basis function artificial neural network (ANN), Bioresour. Technol. 282 (2019) 262–268, https://doi.org/10.1016/j.biortech.2019.03.044.
[36] G.B. Sahoo, C. Ray, Predicting flux decline in crossflow membranes using artificial neural networks and genetic algorithms, J. Memb. Sci. 283 (2006) 147–157, https://doi.org/10.1016/j.memsci.2006.06.019.
[37] J. Sargolzaei, M. Haghighi Asl, A. Hedayati Moghaddam, Membrane permeate flux and rejection factor prediction using intelligent systems, Desalination. 284 (2012) 92–99. 10.1016/j.desal.2011.08.041.
[38] Y. Zhao, J.S. Taylor, S. Chellam, Predicting RO/NF water quality by modified solution diffusion model and artificial neural networks, J. Memb. Sci. 263 (2005) 38–46, https://doi.org/10.1016/j.memsci.2005.04.004.
[39] A.M. Aish, H.A. Zaqoot, S.M. Abdeljawad, Artificial neural network approach for predicting reverse osmosis desalination plants performance in the Gaza Strip, Desalination. 367 (2015) 240–247, https://doi.org/10.1016/j.desal.2015.04.008.
[40] M. Bianchini, P. Frasconi, M. Gori, Learning without Local Minima in Radial Basis Function Networks, IEEE Trans. Neural Networks. (1995), https://doi.org/10.1109/72.377979.
[41] Y. Chen, G. Yu, Y. Long, J. Teng, X. You, B.Q. Liao, H. Lin, Application of radial basis function artificial neural network to quantify interfacial energies related to membrane fouling in a membrane bioreactor, Bioresour. Technol. 293 (2019), 122103, https://doi.org/10.1016/j.biortech.2019.122103.
[42] C. Aydiner, I. Demir, E. Yildiz, Modeling of flux decline in crossflow microfiltration using neural networks: The case of phosphate removal, J. Memb. Sci. 248 (2005) 53–62, https://doi.org/10.1016/j.memsci.2004.07.036.
[43] C. Charcosset, Membrane Processes Biotechnol. Pharmaceut. (2012), https://doi.org/10.1016/C2011-0-04261-8.
[44] K. Scott, Handbook Industr. Membr. (1995), https://doi.org/10.1016/0958-2118(95)90218-x.
[45] C. Aydiner, I. Demir, B. Keskinler, O. Ince, Joint analysis of transient flux behaviors via membrane fouling in hybrid PAC/MF processes using neural network, Desalination. 250 (2010) 188–196, https://doi.org/10.1016/j.desal.2009.06.025.
[46] E.R. Ziegel, P. Ross, Taguchi Techniques for Quality Engineering, Technometrics. (1997), https://doi.org/10.2307/1270793.
[47] S. Chellam, Artificial neural network model for transient crossflow microfiltration of polydispersed suspensions, J. Memb. Sci. 258 (2005) 35–42, https://doi.org/10.1016/j.memsci.2004.11.038.
[48] G.D. Garson, Interpreting Neural-Network Connection Weights, AI Expert. 6 (1991) 46–51.
[49] A.T.C. Goh, Seismic liquefaction potential assessed by neural networks, J. Geotech. Eng. (1994), https://doi.org/10.1061/(ASCE)0733-9410(1994)120:9(1467).
[50] Q.F. Liu, S.H. Kim, S. Lee, Prediction of microfiltration membrane fouling using artificial neural network models, Sep. Purif. Technol. 70 (2009) 96–102, https://doi.org/10.1016/j.seppur.2009.08.017.
[51] S. Strugholtz, S. Panglisch, J. Gebhardt, R. Gimbel, Neural networks and genetic algorithms in membrane technology modelling, J. Water Supply Res. Technol. - AQUA. 57 (2008) 23–34, https://doi.org/10.2166/aqua.2008.008.
[52] B.K. Nandi, A. Moparthi, R. Uppaluri, M.K. Purkait, Treatment of oily wastewater using low cost ceramic membrane: Comparative assessment of pore blocking and artificial neural network models, Chem. Eng. Res. Des. 88 (2010) 881–892, https://doi.org/10.1016/j.cherd.2009.12.005.
[53] H. Shokrkar, A. Salahi, N. Kasiri, T. Mohammadi, Mullite ceramic membranes for industrial oily wastewater treatment: Experimental and neural network modeling, Water Sci. Technol. 64 (2011) 670–676, https://doi.org/10.2166/wst.2011.655.
[54] Y. Liu, G. He, M. Tan, F. Nie, B. Li, Artificial neural network model for turbulence promoter-assisted crossflow microfiltration of particulate suspensions, Desalination. 338 (2014) 57–64, https://doi.org/10.1016/j.desal.2014.01.015.
[55] J. Koelmel, M.N.V. Prasad, G. Velvizhi, S.K. Butti, S.V. Mohan, Metalliferous Waste in India and Knowledge Explosion in Metal Recovery Techniques and Processes for the Prevention of Pollution, Environ. Mater. Waste Resour. Recover. Pollut. Prev. (2016), https://doi.org/10.1016/B978-0-12-803837-6.00015-9.
[56] H. Niemi, A. Bulsari, S. Palosaari, Simulation of membrane separation by neural networks, J. Memb. Sci. 102 (1995) 185–191, https://doi.org/10.1016/0376-7388(94)00314-O.
[57] N. Delgrange, C. Cabassud, M. Cabassud, L. Durand-Bourlier, J.M. Lainé, Modelling of ultrafiltration fouling by neural network, Desalination. 118 (1998) 213–227, https://doi.org/10.1016/S0011-9164(98)00132-5.
[58] N. Delgrange, C. Cabassud, M. Cabassud, L. Durand-Bourlier, J.M. Lainé, Neural networks for prediction of ultrafiltration transmembrane pressure - Application to drinking water production, J. Memb. Sci. 150 (1998) 111–123, https://doi.org/10.1016/S0376-7388(98)00217-8.
[59] M. Cabassud, N. Delgrange-Vincent, C. Cabassud, L. Durand-Bourlier, J.M. Lainé, Neural networks: A tool to improve UF plant productivity, Desalination. 145 (2002) 223–231, https://doi.org/10.1016/S0011-9164(02)00416-2.
[60] C.M. Chew, M.K. Aroua, M.A. Hussain, A practical hybrid modelling approach for the prediction of potential fouling parameters in ultrafiltration membrane water treatment plant, J. Ind. Eng. Chem. 45 (2017) 145–155, https://doi.org/10.1016/j.jiec.2016.09.017.
[61] C. Teodosiu, O. Pastravanu, M. Macoveanu, Neural network models for ultrafiltration and backwashing, Water Res. 34 (2000) 4371–4380, https://doi.org/10.1016/S0043-1354(00)00217-7.
[62] W.R. Bowen, M.G. Jones, H.N.S. Yousef, Prediction of the rate of crossflow membrane ultrafiltration of colloids: A neural network approach, Chem. Eng. Sci. 53 (1998) 3793–3802, https://doi.org/10.1016/S0009-2509(98)00183-3.
[63] C. Bhattacharya, M. Singh, Studies on the Applicability of Artificial Neural Network (ANN) in Continuous Stirred Ultrafiltration, Chem. Eng. Technol. 25 (2002) 1187–1192, https://doi.org/10.1002/1521-4125(20021210)25:12<1187::AID-CEAT1187>3.0.CO;2-T.
[64] H.K. Oh, M.J. Yu, E.M. Gwon, J.Y. Koo, S.G. Kim, A. Koizumi, KNT-artificial neural network model for flux prediction of ultrafiltration membrane producing drinking water, Water Sci. Technol. 50 (2004) 103–110, https://doi.org/10.2166/wst.2004.0499.
[65] M. Kabsch-Korbutowicz, M. Kutyłowska, Use of artificial intelligence in predicting the turbidity retention coefficient during ultrafiltration of water, Environ. Prot. Eng. 37 (2011) 75–84.
[66] R.S. Faibish, M. Elimelech, Y. Cohen, Effect of interparticle electrostatic double layer interactions on permeate flux decline in crossflow membrane filtration of colloidal suspensions: An experimental investigation, J. Colloid Interface Sci. (1998), https://doi.org/10.1006/jcis.1998.5563.
[67] Q.F. Liu, S.H. Kim, Evaluation of membrane fouling models based on bench-scale experiments: A comparison between constant flowrate blocking laws and

artificial neural network (ANNs) model, J. Memb. Sci. 310 (2008) 393–401, https://doi.org/10.1016/j.memsci.2007.11.020.
[68] M.J. Corbatón-Báguena, M.C. Vincent-Vela, J.M. Gozálvez-Zafrilla, S. Álvarez-Blanco, J. Lora-García, D. Catalán-Martínez, Comparison between artificial neural networks and Hermia's models to assess ultrafiltration performance, Sep. Purif. Technol. 170 (2016) 434–444, https://doi.org/10.1016/j.seppur.2016.07.007.
[69] W. Lin, L. Jing, Z. Zhu, Q. Cai, B. Zhang, Removal of Heavy Metals from Mining Wastewater by Micellar-Enhanced Ultrafiltration (MEUF): Experimental Investigation and Monte Carlo-Based Artificial Neural Network Modeling, Water. Air. Soil Pollut. 228 (2017), https://doi.org/10.1007/s11270-017-3386-5.
[70] R. Soleimani, N.A. Shoushtari, B. Mirza, A. Salahi, Experimental investigation, modeling and optimization of membrane separation using artificial neural network and multi-objective optimization using genetic algorithm, Chem. Eng. Res. Des. 91 (2013) 883–903, https://doi.org/10.1016/j.cherd.2012.08.004.
[71] E. Nagy, Nanofiltration, in: Basic Equations Mass Transp. through a Membr. Layer, 2012. 10.1016/b978-0-12-416025-5.00010-7.
[72] W.R. Bowen, M.G. Jones, J.S. Welfoot, H.N.S. Yousef, Predicting salt rejections at nanofiltration membranes using artificial neural networks, Desalination. 129 (2000) 147–162, https://doi.org/10.1016/S0011-9164(00)00057-6.
[73] H. Al-Zoubi, N. Hilal, N.A. Darwish, A.W. Mohammad, Rejection and modelling of sulphate and potassium salts by nanofiltration membranes: neural network and Spiegler-Kedem model, Desalination. 206 (2007) 42–60, https://doi.org/10.1016/j.desal.2006.02.060.
[74] G.R. Shetty, H. Malki, S. Chellam, Predicting contaminant removal during municipal drinking water nanofiltration using artificial neural networks, J. Memb. Sci. 212 (2003) 99–112, https://doi.org/10.1016/S0376-7388(02)00473-8.
[75] G.R. Shetty, S. Chellam, Predicting membrane fouling during municipal drinking water nanofiltration using artificial neural networks, J. Memb. Sci. 217 (2003) 69–86, https://doi.org/10.1016/S0376-7388(03)00075-9.
[76] M. Mousavi, A. Avami, Modeling and Simulation of Water Softening by Nanofiltration Using Artificial Neural Network, Iran. J. Chem. Chem. Eng. 25 (2006) 421–430.
[77] N.A. Darwish, N. Hilal, H. Al-Zoubi, A.W. Mohammad, Neural networks simulation of the filtraton of sodium chloride and magnesium chloride solutions using nanofiltration membranes, Chem. Eng. Res. Des. 85 (2007) 417–430, https://doi.org/10.1205/cherd06037.
[78] M.K. Purkait, V.D. Kumar, D. Maity, Treatment of leather plant effluent using NF followed by RO and permeate flux prediction using artificial neural network, Chem. Eng. J. 151 (2009) 275–285, https://doi.org/10.1016/j.cej.2009.03.023.
[79] Y. Ammi, L. Khaouane, S. Hanini, Prediction of the rejection of organic compounds (neutral and ionic) by nanofiltration and reverse osmosis membranes using neural networks, Korean J. Chem. Eng. 32 (2015) 2300–2310, https://doi.org/10.1007/s11814-015-0086-y.
[80] V. Yangali-Quintanilla, A. Verliefde, 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. Memb. Sci. 342 (2009) 251–262, https://doi.org/10.1016/j.memsci.2009.06.048.
[81] L. Khaouane, Y. Ammi, S. Hanini, Modeling the Retention of Organic Compounds by Nanofiltration and Reverse Osmosis Membranes Using Bootstrap Aggregated Neural Networks, Arab. J. Sci. Eng. 42 (2017) 1443–1453, https://doi.org/10.1007/s13369-016-2320-2.
[82] S. Park, S.S. Baek, J.C. Pyo, Y. Pachepsky, J. Park, K.H. Cho, Deep neural networks for modeling fouling growth and flux decline during NF/RO membrane filtration, J. Memb. Sci. 587 (2019), 117164, https://doi.org/10.1016/j.memsci.2019.06.004.
[83] M.T. Demeuse, Production and applications of hollow fibers, Handb. Text. Fibre Struct. (2009), https://doi.org/10.1533/9781845697310.3.485.
[84] K.A. Al-Shayji, Modeling, simulation, and optimization of large-scale commerical desalination plants, Virginia Polytechnic Institute and State University, 1998.
[85] K.A. Al-Shayji, Y.A. Liu, Predictive modeling of large-scale commercial water desalination plants: Data-based neural network and model-based process simulation, Ind. Eng. Chem. Res. 41 (2002) 6460–6474, https://doi.org/10.1021/ie020077r.
[86] M.M. Jafar, A. Zilouchian, Prediction of critical desalination parameters using radial basis functions networks, J. Intell. Robot. Syst. Theory Appl. 34 (2002) 219–230, https://doi.org/10.1023/A:1015620713975.
[87] Z.V.P. Murthy, M.M. Vora, Prediction of reverse osmosis performance using artificial neural network, Indian J. Chem. Technol. 11 (2004) 108–115.
[88] A. Abbas, N. Al-Bastaki, Modeling of an RO water desalination unit using neural networks, Chem. Eng. J. 114 (2005) 139–143, https://doi.org/10.1016/j.cej.2005.07.016.
[89] Y.G. Lee, Y.S. Lee, J.J. Jeon, S. Lee, D.R. Yang, I.S. Kim, J.H. Kim, Artificial neural network model for optimizing operation of a seawater reverse osmosis desalination plant, Desalination. 247 (2009) 180–189, https://doi.org/10.1016/j.desal.2008.12.023.
[90] P. Cabrera, J.A. Carta, J. González, G. Melián, Artificial neural networks applied to manage the variable operation of a simple seawater reverse osmosis plant, Desalination. 416 (2017) 140–156, https://doi.org/10.1016/j.desal.2017.04.032.
[91] R. Righton, Development of an artificial neural network model for predicting the performance of a reverse osmosis RO unit, Curtin University of Technology, 2009.
[92] D. Libotean, J. Giralt, F. Giralt, R. Rallo, T. Wolfe, Y. Cohen, Neural network approach for modeling the performance of reverse osmosis membrane desalting, J. Memb. Sci. 326 (2009) 408–419, https://doi.org/10.1016/j.memsci.2008.10.028.
[93] M. Khayet, C. Cojocaru, M. Essalhi, Artificial neural network modeling and response surface methodology of desalination by reverse osmosis, J. Memb. Sci. 368 (2011) 202–214, https://doi.org/10.1016/j.memsci.2010.11.030.
[94] M.C. Garg, H. Joshi, A new approach for optimization of small-scale RO membrane using artificial groundwater, Environ. Technol. (United Kingdom) 35 (2014) 2988–2999, https://doi.org/10.1080/09593330.2014.927928.
[95] A. Moradi, V. Mojarradi, M. Sarcheshmehpour, Prediction of RO membrane performances by use of artificial neural network and using the parameters of a complex mathematical model, Res. Chem. Intermed. 39 (2013) 3235–3249, https://doi.org/10.1007/s11164-012-0835-z.
[96] A. Salgado-Reyna, E. Soto-Regalado, R. Gómez-González, F.J. Cerino-Córdova, R. B. García-Reyes, M.T. Garza-González, M.M. Alcalá-Rodríguez, Artificial neural networks for modeling the reverse osmosis unit in a wastewater pilot treatment plant, Desalin. Water Treat. 53 (2015) 1177–1187, https://doi.org/10.1080/19443994.2013.862023.
[97] J. Farahbakhsh, M. Delnavaz, V. Vatanpour, Simulation and characterization of novel reverse osmosis membrane prepared by blending polypyrrole coated multiwalled carbon nanotubes for brackish water desalination and antifouling properties using artificial neural networks, J. Memb. Sci. 581 (2019) 123–138, https://doi.org/10.1016/j.memsci.2019.03.050.
[98] A. Ruiz-García, J. Feo-García, Operating and maintenance cost in seawater reverse osmosis desalination plants. Artificial neural network based model, Desalin. Water Treat. 73 (2017) 73–79, https://doi.org/10.5004/dwt.2017.20807.
[99] J. Jawad, A.H. Hawari, S. Zaidi, Modeling of forward osmosis process using artificial neural networks (ANN) to predict the permeate flux, Desalination. 484 (2020), 114427, https://doi.org/10.1016/j.desal.2020.114427.
[100] A.H. Hawari, N. Kamal, A. Altaee, Combined influence of temperature and flow rate of feeds on the performance of forward osmosis, Desalination. 398 (2016) 98–105, https://doi.org/10.1016/j.desal.2016.07.023.
[101] J. Jawad, A.H. Hawari, S.J. Zaidi, Modeling and sensitivity analysis of the forward osmosis process to predict membrane flux using a novel combination of neural network and response surface methodology techniques, Membranes (Basel). 11 (2021) 1–19, https://doi.org/10.3390/membranes11010070.
[102] M.A. Shannon, P.W. Bohn, M. Elimelech, J.G. Georgiadis, B.J. Marîas, A. M. Mayes, Science and technology for water purification in the coming decades, Nature. 452 (2008) 301–310, https://doi.org/10.1038/nature06599.
[103] S.J. Wang, J.J. Zhong, Bioreactor Engineering, Bioprocess. Value-Added Prod. from Renew. Resour. (2007), https://doi.org/10.1016/B978-044452114-9/50007-4.
[104] S. Geissler, T. Wintgens, T. Melin, K. Vossenkaul, C. Kullmann, Modelling approaches for filtration processes with novel submerged capillary modules in membrane bioreactors for wastewater treatment, Desalination. 178 (2005) 125–134, https://doi.org/10.1016/j.desal.2004.11.032.
[105] C. Li, X. Wang, Application of MBR Membrane Flux Prediction Based on Elman Neural Network, in: DEStech Trans. Eng. Technol. Res., 2018: pp. 365–372. 10.12783/dtetr/iccere2017/18308.
[106] Z. Chen, N. Ren, A. Wang, Z.P. Zhang, Y. Shi, A novel application of TPAD-MBR system to the pilot treatment of chemical synthesis-based pharmaceutical wastewater, Water Res. 42 (2008) 3385–3392, https://doi.org/10.1016/j.watres.2008.04.020.
[107] Z. Chen, A. Zhou, N. Ren, Y. Tian, D. Hu, Pollutants removal and simulation model of combined membrane process for wastewater treatment and reuse in submarine cabin for long voyage, J. Environ. Sci. 21 (2009) 1503–1512, https://doi.org/10.1016/S1001-0742(08)62447-X.
[108] Z. Liu, D. Pan, J. Wang, S. Yang, Modelling of membrane fouling by PCA-PSOBP neural network, in: 2010 Int. Conf. Comput. Control Ind. Eng. CCIE 2010, IEEE, 2010: pp. 34–37. 10.1109/CCIE.2010.16.
[109] H. Hamedi, M. Ehteshami, S.A. Mirbagheri, S. Zendehboudi, New deterministic tools to systematically investigate fouling occurrence in membrane bioreactors, Chem. Eng. Res. Des. 144 (2019) 334–353, https://doi.org/10.1016/j.cherd.2019.02.003.
[110] C. Li, Z. Yang, H. Yan, T. Wang, The application and research of the GA-BP neural network algorithm in the mbr membrane fouling, Abstr. Appl. Anal. 2014 (2014), https://doi.org/10.1155/2014/673156.
[111] A.R. Pendashteh, A. Fakhru'l-Razi, N. Chaibakhsh, L.C. Abdullah, S.S. Madaeni, Z. Z. Abidin, Modeling of membrane bioreactor treating hypersaline oily wastewater by artificial neural network, J. Hazard. Mater. 192 (2011) 568–575. 10.1016/j.jhazmat.2011.05.052.
[112] S.A. Mirbagheri, M. Bagheri, Z. Bagheri, A.M. Kamarkhani, Evaluation and prediction of membrane fouling in a submerged membrane bioreactor with simultaneous upward and downward aeration using artificial neural network-genetic algorithm, Process Saf. Environ. Prot. 96 (2015) 111–124, https://doi.org/10.1016/j.psep.2015.03.015.
[113] S.A. Mirbagheri, M. Bagheri, S. Boudaghpour, M. Ehteshami, Z. Bagheri, Performance evaluation and modeling of a submerged membrane bioreactor treating combined municipal and industrial wastewater using radial basis function artificial neural networks, J. Environ. Heal. Sci. Eng. 13 (2015) 1–15, https://doi.org/10.1186/s40201-015-0172-4.
[114] H. Hazrati, A.H. Moghaddam, M. Rostamizadeh, The influence of hydraulic retention time on cake layer specifications in the membrane bioreactor: Experimental and artificial neural network modeling, J. Environ. Chem. Eng. 5 (2017) 3005–3013, https://doi.org/10.1016/j.jece.2017.05.050.

[115] Y. Cai, X. Li, A.A. Zaidi, Y. Shi, K. Zhang, P. Sun, Z. Lu, Processing efficiency, simulation and enzyme activities analysis of an air-lift multilevel circulation membrane bioreactor (AMCMBR) on marine domestic sewage treatment, Period. Polytech. Chem. Eng. 63 (2019) 448–458, https://doi.org/10.3311/PPch.13389.
[116] Y. Cai, X. Li, A.A. Zaidi, Y. Shi, K. Zhang, R. Feng, A. Lin, C. Liu, Effect of hydraulic retention time on pollutants removal from real ship sewage treatment via a pilot-scale air-lift multilevel circulation membrane bioreactor, Chemosphere. 236 (2019), 124338, https://doi.org/10.1016/j.chemosphere.2019.07.069.
[117] Y. Cai, A.A. Zaidi, Y. Shi, K. Zhang, X. Li, S. Xiao, A. Lin, Influence of salinity on the biological treatment of domestic ship sewage using an air-lift multilevel circulation membrane reactor, Environ. Sci. Pollut. Res. 26 (2019) 37026–37036, https://doi.org/10.1007/s11356-019-06813-4.
[118] Y. Cai, T. Ben, A.A. Zaidi, Y. Shi, K. Zhang, A. Lin, C. Liu, Effect of pH on pollutants removal of ship sewage treatment in an innovative aerobic-anaerobic micro-sludge MBR system, Water. Air. Soil Pollut. 230 (2019), https://doi.org/10.1007/s11270-019-4211-0.
[119] C. Yang, X. Peng, Y. Zhao, X. Wang, J. Fu, K. Liu, Y. Li, P. Li, Prediction model to analyze the performance of VMD desalination process, Comput. Chem. Eng. 132 (2020), 106619, https://doi.org/10.1016/j.compchemeng.2019.106619.
[120] V. Belessiotis, S. Kalogirou, E. Delyannis, V. Belessiotis, S. Kalogirou, E. Delyannis, Chapter four – membrane distillation, Therm. Sol. Desalin. (2016), https://doi.org/10.1016/B978-0-12-809656-7.00004-0.
[121] M. Tavakolmoghadam, M. Safavi, An optimized neural network model of desalination by vacuum membrane distillation using genetic algorithm, Procedia Eng. 42 (2012) 106–112, https://doi.org/10.1016/j.proeng.2012.07.400.
[122] W. Cao, Q. Liu, Y. Wang, I.M. Mujtaba, Modeling and simulation of VMD desalination process by ANN, Comput. Chem. Eng. 84 (2016) 96–103, https://doi.org/10.1016/j.compchemeng.2015.08.019.
[123] M. Khayet, C. Cojocaru, Artificial neural network modeling and optimization of desalination by air gap membrane distillation, Sep. Purif. Technol. 86 (2012) 171–182, https://doi.org/10.1016/j.seppur.2011.11.001.
[124] S. Shirazian, M. Alibabaei, Using neural networks coupled with particle swarm optimization technique for mathematical modeling of air gap membrane distillation (AGMD) systems for desalination process, Neural Comput. Appl. 28 (2017) 2099–2104, https://doi.org/10.1007/s00521-016-2184-0.
[125] R. Porrazzo, A. Cipollina, M. Galluzzo, G. Micale, A neural network-based optimizing control system for a seawater-desalination solar-powered membrane distillation unit, Comput. Chem. Eng. 54 (2013) 79–96, https://doi.org/10.1016/j.compchemeng.2013.03.015.
[126] J.D. Gil, A. Ruiz-Aguirre, L. Roca, G. Zaragoza, M. Berenguel, Prediction models to analyse the performance of a commercial-scale membrane distillation unit for desalting brines from RO plants, Desalination. 445 (2018) 15–28, https://doi.org/10.1016/j.desal.2018.07.022.