

Development of machine learning and stepwise mechanistic models for performance prediction of direct contact membrane distillation module- A comparative study

Pooria Behnam, Abdellah Shafieian, Masoumeh Zargar, Mehdi Khiadani *

School of Engineering, Edith Cowan University, 270 Joondalup Drive, Joondalup, Perth, WA 6027, Australia



ARTICLE INFO

Keywords:

Direct contact membrane distillation
Mechanistic modeling approach
Machine learning
Artificial neural network
Support vector regression
Random forest

ABSTRACT

The development of accurate and fast modeling tools to predict the performance of direct contact membrane distillation (DCMD) modules can result in their performance improvement. Conventional models ignoring the variations of operational parameters along the membrane's length may cause inaccuracy. Moreover, models considering these variations are often complex and computationally expensive. To propose an accurate and fast modeling tool, the possibility of using machine learning models for the performance prediction of the DCMD module has been studied for the first time in this study. The robustness of three machine learning models (ANN, SVR, and RF) has been thoroughly compared with the stepwise mechanistic modeling approach in terms of models' accuracy, trend predictability, and computational time. The results show that ANN and SVR models exhibit an enhanced performance over the mechanistic model, possessing a MAPE_{test} of 3.46% and 4.78% as compared to the mechanistic model with a MAPE_{test} of 7.31%. Further, compared to the mechanistic model, the machine learning models have the privilege of simplicity, enhanced accuracy, and significantly lowered computational time. The feature importance analysis also showed that the feed flow temperature is the most influencing parameter on permeate flux in the DCMD system.

1. Introduction

Desalination methods can provide an effective solution to the water scarcity issue. Among various desalination methods, membrane distillation (MD) possesses distinctive advantages including low operating temperature (50–80 °C), simple /compact design, and the capability to treat high saline water [1,2]. MD is known as a hybrid desalination method that benefits from both thermal and membrane processes. In this process, salt is removed from saline water by vaporizing the saline water and transferring the vapor through a thin hydrophobic membrane. Based on the method of vapors collection, MD modules are divided into four basic categories: direct contact membrane distillation (DCMD), air-gap membrane distillation (AGMD), vacuum membrane distillation (VMD), and sweep gas membrane distillation (SGMD). DCMD modules have been mostly studied by researchers compared to other configurations because of their simplicity in design and operation [3].

Heretofore, the performance of the DCMD system has been extensively investigated through the experimentation. Elmarghany et al [4]. investigated the effects of temperature and mass flow rate of feed

solution on the performance of a lab-scale DCMD system. In another study [5], DCMD systems equipped with various spacers were studied, and a detailed comparison in terms of mass transfer characteristics was made with available correlations from the literature. Khalifa et al [6]. showed that DCMD can be used effectively for the desalination process even with high saline water. Some researchers have sought to couple solar energy with the DCMD system due to the low operating temperature of the DCMD system and therefore the capability of powering with solar energy. Shafieian and Khiadani [7] analyzed the performance of a tubular DCMD system equipped with heat pipe evacuated tube solar collectors. The possibility of direct use of solar energy for supplying the required energy of DCMD was examined by Bamasag et al [8], where DCMD modules were inserted inside evacuated tube collectors acting as feed containers and solar receivers at the same time. Further, the findings from an experimental investigation by Kabeel et al [9]. showed that the use of cooling units in the solar-driven DCMD systems led to a significant increase in freshwater productivity, an increase about 1.25 times compared to the system without the cooling unit.

Different modeling approaches have also been applied to examine the performance of DCMD modules [10]. In these investigations, heat

* Corresponding author.

E-mail address: m.khiadani@ecu.edu.au (M. Khiadani).

Nomenclature			
A	Area (m^2)	t	Transpose operator, time (s)
AGMD	Air-gap membrane distillation	V	Velocity (m/s)
ANN	Artificial neural network	VMD	Vacuum membrane distillation
b	Bias term	W	Weights
C	Penalty parameter	w	Transferred molecules mean free path (m)
C_m	Membrane specific mass transfer coefficient ($\text{kg}/(\text{m}^2 \cdot \text{s} \cdot \text{Pa})$)	x	Input parameter
d	Pore dimension (m)	Y	Specific enthalpy (J/kg)
DCMD	Direct contact membrane distillation	y_i	Experimental target value
E	Energy rate (W)	\hat{y}_i	Predicted target value
H	Hat matrix	\bar{y}	Mean value of the experimental data
H^*	Warning leverage		
H_v	Evaporation enthalpy (J/kg)		
h	Convective heat transfer coefficient ($\text{W}/(\text{m}^2 \cdot \text{K})$)		
J	Permeate flux ($\text{kg}/(\text{m}^2 \cdot \text{s})$)		
K	Kernel, Film mass transfer coefficient (m/s)		
k	Number of input features		
kn	Knudsen number		
m	Mass flow rate (kg/s)		
MAE	Mean absolute error	a	Assumption
MAPE	Mean absolute percentage error	f	Feed flow
MD	Membrane distillation	f,m	Membrane surface, feed side
N, n	The number of experimental samples	i	Inlet
NTU	Number of transfer units	in	Inlet
P	Pressure (Pa)	j	J'th step
PGMD	Permeate Gap membrane distillation	max	Maximum value
Q	Heat flux (W/m^2)	min	Minimum value
R^2	Coefficient of determination	m,j	Membrane,step
RF	Random forest	norm	Normalized value
RMSE	Root mean square error	n	Last step
RSM	Response surface methodology	p	Permeate
S	Salinity (g/l)	p,m	Membrane surface, permeate side
SGMD	Sweep gas membrane distillation	v	Vapor
SVR	Support vector regression	w	Water
T	Temperature (°C)	wa	Water, assumption

and mass transfer equations have been solved from different viewpoints. Several simplified studies have ignored the variations of operational parameters (temperature, salinity, and mass flow rate) along the membrane's length and only considered the inlet bulk or average operational values [11,12]. Although these models had the privilege of simplicity and low computational cost, they may lead to significant errors due to mentioned simplified assumptions, especially in large-scale applications [13]. To address this problem, more accurate modeling of the DCMD module was performed using "stepwise model", whereby variations of operational parameters were considered by solving heat and mass equations for a number of segments along the membrane length. In other words, in the stepwise modeling approach, the module was discretized into smaller steps positioned perpendicular to the stream direction to provide the distribution of process variables along the membrane length. The results of the calculations for each step (i.e. outputs) are considered as the inlet boundary conditions for the next step. The module is discretized from the inlets of the feed channels towards their outlets for both feed and permeate channels [13,14]. Elzahaby et al [15], applied the stepwise mechanistic modeling approach to examine the performance of a solar-driven tubular DCMD. The stepwise model was then employed to investigate the effect of membrane's length on the performance of multi-stage DCMD systems [16,17]. It was concluded that the optimum performance of multi-stage DCMD systems could be achieved by varying the feed to permeate flow ratio [17]. Recently, Noamani et al [18], proposed a novel mechanistic modeling approach based on the combination of stepwise and ϵ -NTU models. It was reported that there

was a 10% deviation between the values obtained from the mechanistic and experimental analyses. In another study, Shafieian et al [14], modeled a lab-scale tubular DCMD module using the stepwise methodology, and the developed mechanistic model was verified with experimental results.

The stepwise mechanistic modeling of MD systems still suffers from high computational cost and complexity. This is mainly due to the fact that heat and mass transfer phenomena occur simultaneously in MD systems, resulting in complex mechanistic modeling. Moreover, solving integral heat and mass equations for each segment of MD modules leads to a higher execution time. Data-driven models can play an important role in mitigating the shortcomings of mechanistic modeling approach, but still offer a robust and accurate modeling tool. Over the last few years, the application of data-driven models in desalination systems has received more interest [19–21]. Amongst different data-driven models including machine learning techniques and statistical models, artificial neural network (ANN) and response surface methodology (RSM) have respectively gained more attention to predict the performance of MD modules [10,22]. In this study, there was no special experimental design before conducting the experimental tests and therefore it was not possible to develop the RSM model. Moreover, with reference to the application of the RSM method for performance modeling of the DCMD module [23–26], the development of the RSM model was not in the scope of the current study. It has been reported that ANN enjoyed a better generalization capability compared to RSM and there is no special experimental design to develop the ANN model [27,28]. Khayet and

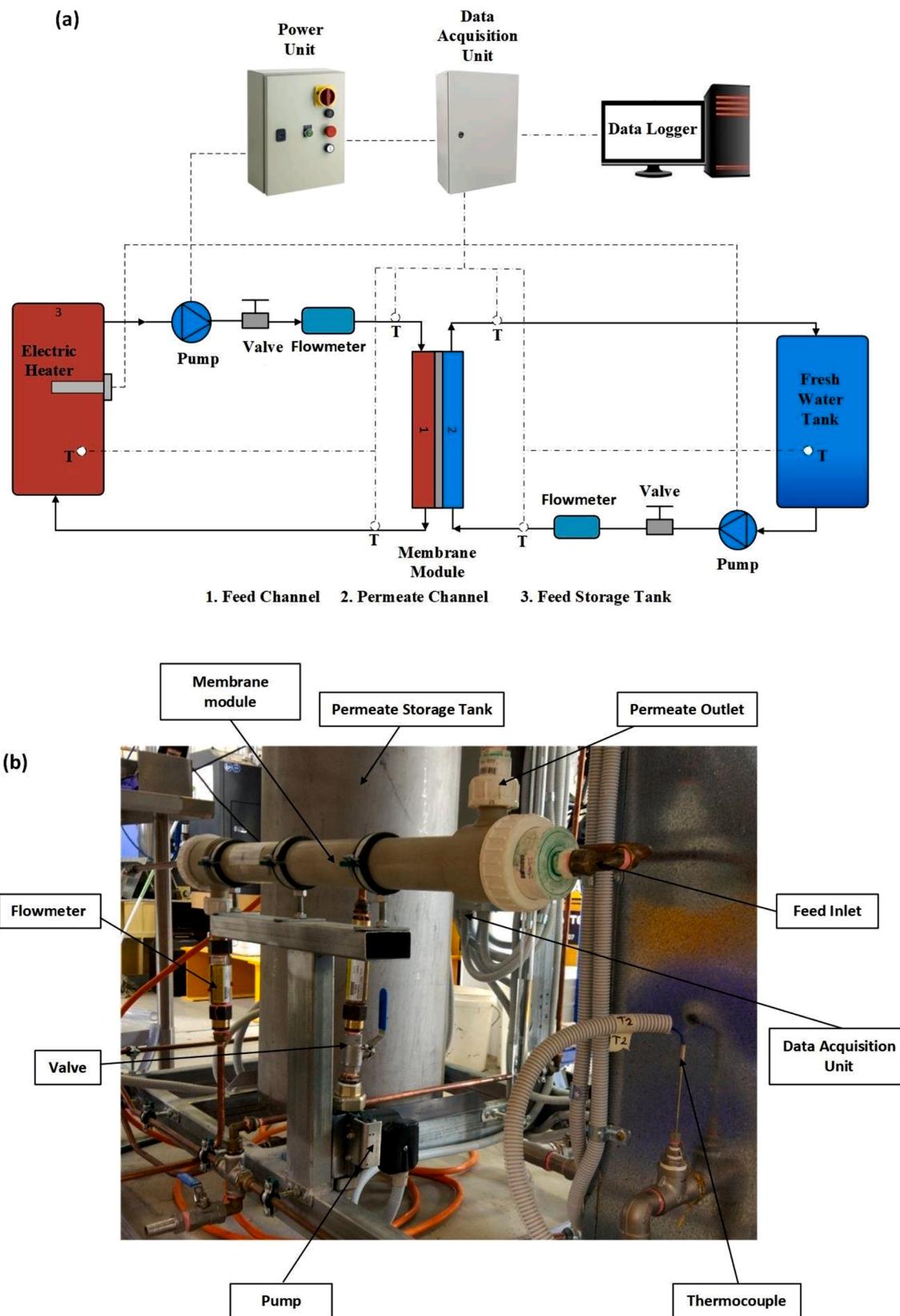


Fig. 1. Experimental rig: (a). Schematic view (b). Photographic view.

Cojocaru [29] employed ANN to predict the performance of the AGMD module, and it was reported that the feed flow temperature had the most influential effect on the distillate flux. The performance of the AGMD system was also investigated by Shirazian and Alibabaei [30] using ANN and particle swarm optimization techniques. Further, Khayet and

Cojocaru [31] applied ANN to 53 experimental tests and employed the Monte Carlo simulation technique to model and optimize an SGMD system. By utilizing 252 experimental tests, Tavakolmoghadam and Safavi [32] used ANN to predict the permeate flux of the VMD configuration. Similarly, the application of the ANN model to predict the

Table 1

The DCMD module characteristics.

Characteristic	Value	Characteristic	Value
Model type	MD 090 TP 2 N ANSI	Membrane porosity	75%
Membrane material	Polypropylene	Outer diameter of membrane module	8.5 mm
Inner diameter of membrane module	5.5 mm	Nominal module diameter	9 cm
Membrane thickness	1.5 mm	Membrane area	0.2 m ²
Membrane module length	75 cm	Average pore size	0.2 μm
Outer shell material	Polypropylene	Potting material	Polypropylene

Table 2

Values of operational parameters in the experimental dataset.

Parameter	Dimension	Value
Feed flow temperature	°C	35–40–45–50–55–60–65
Feed flow rate	l/min	3–5–7.5–10
Salinity of feed flow	g/l	0–10–20–35
Permeate flow temperature	°C	25–30–35–40–45

permeate flux of the VMD system was investigated by Cao et al [33].. To consider both operational and design parameters of a VMD module, Yang et al [34], then employed the ANN model to predict the performance indicators of the VMD module including permeate flux and specific energy consumption. Recently, Kim et al [35], utilized ANN and RSM models to predict wetting phenomenon in the DCMD module for the treatment of synthetic wastewater.

Proposing a fast and accurate modeling tool for analyzing the effects of operational parameters on the permeate flux of the DCMD module can lead to saving time and cost compared to experimental investigations. Moreover, performance improvement of DCMD modules can be achieved by coupling a fast and accurate modeling tool with optimization algorithms. It can be inferred from the literature that existing conventional modeling approaches for the performance prediction of DCMD modules use simplified assumptions which may result in insufficient accuracy. Moreover, models in which the variations of operational parameters along the membrane's length are considered are often complex and computationally expensive. To propose an accurate and fast modeling tool, a limited number of researchers have sought to employ machine learning models for the performance projection of different MD configurations [10]. To the best of the authors' knowledge, only ANN has been utilized as the machine learning model and the application of other promising machine learning models such as support vector regression model (SVR) and random forest (RF) has not been investigated in previous studies for all configurations of MD systems. Regarding the DCMD configuration, the literature review showed that the application of machine learning models for performance analysis of the DCMD module was studied only in one investigation by Kim et al [35].. They employed the ANN model to investigate the membrane wetting in the DCMD module. To develop the ANN model, the chemical species in feed water (NaCl, CaSO₄, humic acid, alginate, and sodium dodecyl sulfate (SDS)) were considered as the influential parameters (inputs) and time required to observe the wetting and the maximum recovery of the permeate were chosen as the outputs. Therefore, the development of machine learning models (ANN, SVR, and RF models) to analyze the effect of main working parameters (i.e., feed flow temperature, feed flow rate, salinity of feed flow, and permeate flow temperature) on the permeate flux of the DCMD module has not been investigated yet. Further, there is no comprehensive investigation that compares the performance of the stepwise mechanistic modeling approach with machine learning models in terms of models' accuracy, trend predictability, and computational time. This comparison can highlight the merits of machine learning models over the conventional mechanistic modeling approach and leads to more application of

machine learning models in analyzing the DCMD modules.

This study aims to determine the possibility of using machine learning models for the prediction of permeate flux in the DCMD system. It is expected that computational time is reduced while maintaining the modeling accuracy. Unlike previous investigations that only used the ANN to model the MD system, the robustness of SVR and RF models have also been examined in this study. This is particularly important because machine learning models have been used to predict the performance of the DCMD system for the first time in this investigation. To develop the machine learning models, an experimental rig was constructed, and a dataset consisting of 70 experimental tests was gathered, where the impact of four operational parameters including feed flow rate, salinity, and feed and permeate temperatures on permeate flux of the DCMD module was evaluated. The robustness of the machine learning models was then compared with the estimated values from the stepwise mechanistic modeling approach in terms of computational cost as well as accuracy. Further, a detailed parametric study was conducted to examine the trend predictability of machine learning models compared with the mechanistic modeling approach model. In addition, the impact of each operational parameter on the permeate flux of the DCMD module was estimated using the RF model.

2. Experimental set-up and data acquisition

The schematic and photographic views of the experimental rig are shown in Fig. 1. The main component of the system is a tubular DCMD module (Type MD 090 TP 2 N ANSI-Microdin-Nadir company) whose effective membrane area and average pore size are 0.2 m² and 0.2 μm, respectively. Further information regarding the characteristics of the membrane module can be found in Table 1. Other components of the experimental rig include a National Instrument Data Acquisition (NI-DAQ) unit; hot seawater and cold permeate water storage tanks; two pumps (Davey company); a power unit; thermocouples (TC Direct company); a computer, two flow meters (Omega company); and pipes and fittings. Seawater is synthesized by dissolving Sodium Chloride (NaCl) in tap water and a conductivity meter is used to constantly monitor its salt concentration. A 2-kW electric heater is utilized to continuously increase the temperature of the synthesized seawater inside the storage tank, whose temperature is monitored by a T-type thermocouple and controlled using a thermostat. The module is operated in counter-flow condition at different feed and permeate mass flow rates. The hot and cold streams are circulated via two pumps and their mass flow rates are regulated using the valves located after the pumps. The mass flow rates of the feed and permeate streams are controlled and monitored using two flowmeters. The temperatures at different locations are measured by six T-type thermocouples which are connected to the NI-DAQ system. An application-based program interface is developed in LabVIEW software to record the data.

The effect of four operational parameters namely feed flow temperature, feed flow rate, salinity of feed flow, and permeate flow temperature on the permeate flux was analyzed by changing the operational parameters in intervals as shown in Table 2. A list of all performed experimental tests is presented in Appendix A. Moreover, to analyze the effect of each operational parameter on the permeate flux of the DCMD module (trend analysis), the one-factor-at-a-time approach was employed for experimental tests. In other words, when one operational parameter was changed, the others were remained constant at a base case condition (feed flow temperature = 45 °C, feed flow rate= 10 l/min, salinity = 35 g/l, and permeate flow temperature = 25 °C). The permeate flux was measured indirectly by checking the water level increase in the freshwater tank over a specific period and the concentration was regularly measured using a conductivity meter type Multi 3410 made by WTW company.

As shown in Table 3, the uncertainty analysis is also performed according to the methods employed by Moffat [36] and Holman [37] to determine the uncertainties associated with measured and calculated

Table 3

The measured and calculated parameters uncertainties.

Parameter	Measurement instrument	Range	Uncertainty (+%)		
			Systematic	Random	Total
Flow rate	Flowmeter	0 – 0.07 kg/s	1.34	0.45	± 2
Salinity	Conductivity meter	0–200 g/l	2.15	0.65	± 2.24
Temperature	T-Type Thermocouple	–150 – 300 °C	1.42	0.32	± 1.7
Permeate flux	–	–	–	–	± 4.05

parameters.

3. Descriptions of stepwise mechanistic modeling approach and machine learning models

3.1. Stepwise mechanistic modeling approach

The configuration of the DCMD module including the feed and permeate channels, inlet and outlet ports, and the housing is shown in Fig. 2. To include the changing nature of the involved parameters including mass flow rate, temperature, and salinity, the module was discretized into smaller segments positioned perpendicular to the stream direction (Fig. 2). In addition to heat and mass transfer equations, which need to be solved for each segment, the energy and mass balance equations are supposed to be satisfied. The results of the calculations for each segment (i.e. outputs) are considered as the inlet boundary conditions for the next segment. The module is discretized from the inlets of the feed channels towards their outlets for both feed and permeate

channels. The permeate flow is considered to be unidirectional which can be justified by the large length to diameter ratio of the module and its small diameter.

The computation process of the mechanistic modeling approach is illustrated in Fig. 3. Furthermore, detailed descriptions of the energy and mass balance equations as well as heat and mass transfer equations are provided in Appendix B. The steady-state condition has been considered to solve these equations and The main reason for the validity of this assumption is that the ratio of the changes in important parameters over time is insignificant [15,18]. Application of heat and mass transfer equations plus mass and energy balance equations leads to an equation including salinity and the outlet temperature as the unknown parameters. To eliminate one of these parameters, the following equation (i.e. the salt mass balance equation) can be applied:

$$\frac{m_{f,j} S_{f,j}}{\rho_{f,j}} = \frac{m_{f,j+1} S_{f,j+1}}{\rho_{f,j+1}} \quad (1)$$

The program is initiated by entering the initial guesses for the mass flux

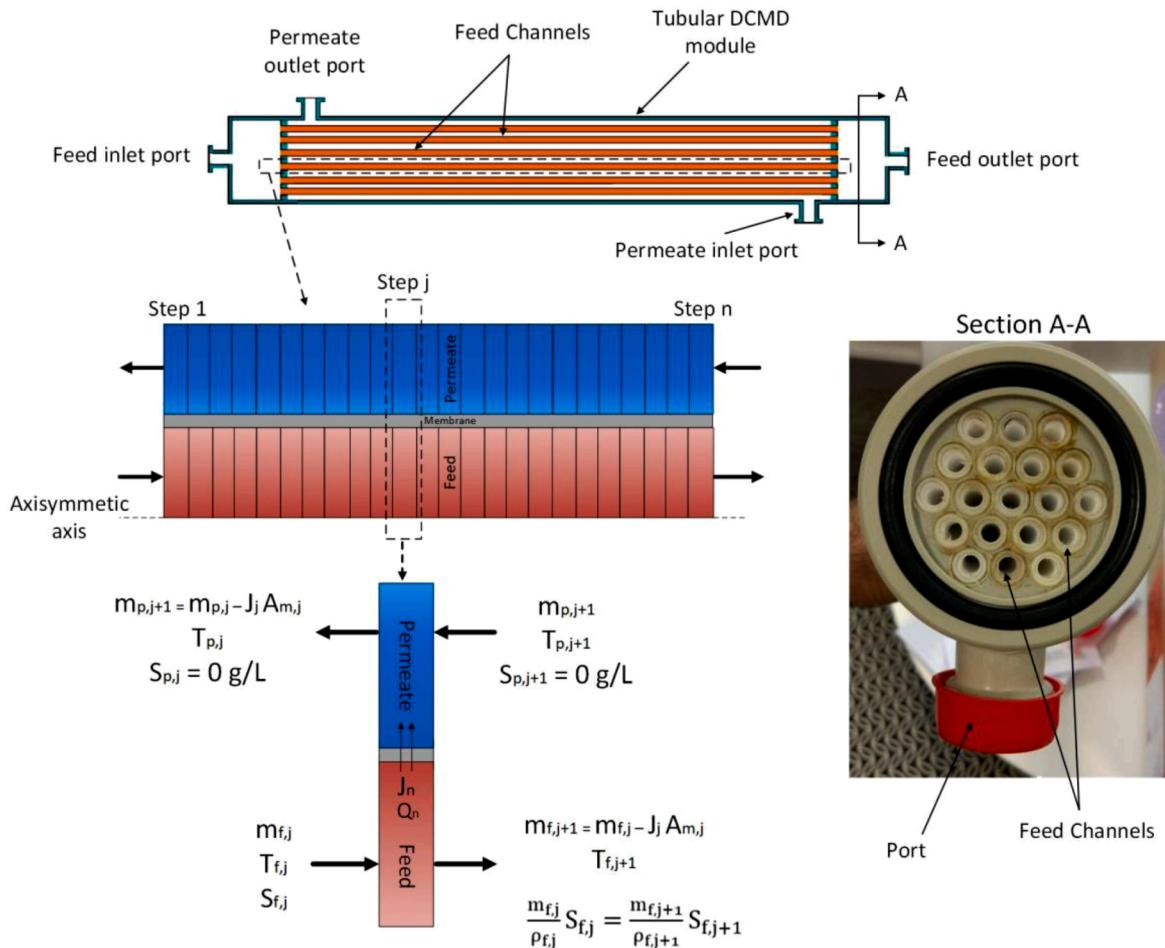


Fig. 2. Discretized membrane plus the boundary conditions.

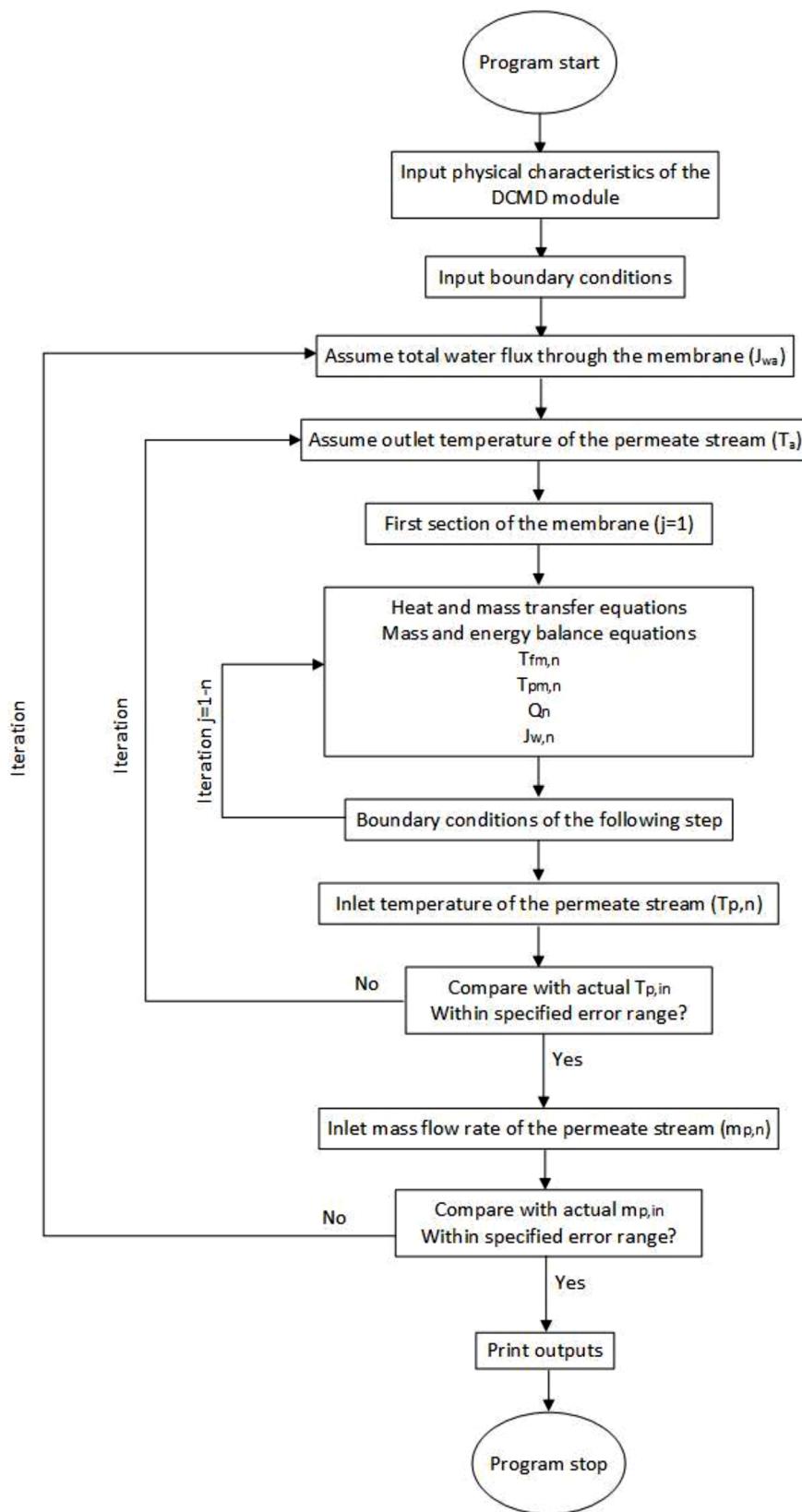


Fig. 3. flowchart of computation procedure.

through the membrane (J_{wa}) and the permeate outlet temperature (T_a). The calculation process is followed until the inlet temperature of the permeate stream is calculated. This needs to be compared with the actual permeate inlet temperature. The process is iterated until the difference is below 0.001 °C. The mentioned process is repeated to

determine the permeate flux.

3.2. Machine learning models

Three models including artificial neural network (ANN), support

Table 4
Input and output values for the machine learning models.

Parameter	Dimension	Range
Inputs		
Feed flow temperature	°C	35–65
Feed flow rate	l/min	3–10
Salinity of feed flow	g/l	0–35
Permeate flow temperature	°C	25–45
Output		
Permeate flux	g/m ² .min	5.7–82.34

Table 5
The main hyper-parameters for machine learning models.

Model	Hyper-parameters
ANN	Number of neurons in the hidden layer Type of activation function
SVR	Kernel coefficient (γ) Penalty parameter (C)
RF	Radius (ϵ) Number of trees in the forest Maximum depth of each tree Minimum sample data for splitting Maximum number of variables chosen for the best split

Table 6
Size of datasets in previous investigations on MD modules.

Reference	MD configuration	Machine learning model	Dataset size
[27]	PGMD	ANN	88
[29]	AGMD	ANN	72
[30]	AGMD	ANN	15
[31]	SGMD	ANN	53
[32]	VMD	ANN	252
[33]	VMD	ANN	38
[34]	VMD	ANN	36
[35]	DCMD	ANN	57
[21]	VMD	ANN	149

vector regression (SVR), and random forest (RF) have been considered as the machine learning models in this study. To predict the permeate flux of the DCMD module as the output, each machine learning model possessed four inputs including feed flow temperature, feed flow rate, the salinity of feed flow, and permeate flow temperature **Table 4** summarizes the inputs and the output and their ranges. Moreover, the main hyper-parameters used for developing the machine learning models have been listed in

Table 5. The scikit-learn library of python language [38] was used to implement the SRV and RF models, while the ANN model was developed using MATLAB. The developed machine learning models are also provided in the supplementary material section. In this study, the performance of the DCMD module was experimentally tested under different operational conditions and a 70 input-output experimental dataset was achieved. 48 experimental samples (69% of the whole data) were selected randomly to train the machine learning models. Afterwards, the remaining 22 experimental data (31% of the whole data), which act as unseen data and were not fed into the models in the training process, were utilized to validate the accuracy of models (please refer to experimental data in **Appendix A**). Furthermore, compared to previous investigations conducted on machine learning models for performance prediction of different MD configurators listed in **Table 6**, a dataset with an acceptable size (70 experimental data) was used to develop the machine learning models, covering a wide range of DCMD's working conditions. Although the selection of training and testing data was performed randomly, fixed training and testing datasets were used to train and test each machine learning model for performing a fair performance comparison among the models. Further, the min-max normalization method was used to decrease the convergence time of ANN and SVR models and therefore input data has been fallen into the

[-1 1] range as follows:

$$x_{\text{norm}} = \frac{(x - x_{\text{min}})}{(x_{\text{max}} - x_{\text{min}})} - 1 \quad (2)$$

Here, x_{norm} represents the normalized data, and x is the experimental data. x_{max} and x_{min} are the maximum and minimum value of experimental data, respectively.

3.2.1. Artificial neural network (ANN)

In this study, the multi-layer perceptron network was chosen as the artificial neural network which has frequently been applied for performance projection of desalination technologies [19,39]. **Fig. 4** indicated the architecture of the ANN model used in this study, having 4 neurons in the input layer and 1 neuron in the output unit for the prediction of permeate flux of the DCMD module. There are several hyper-parameters for the ANN model that should be selected for an acceptable generalization capability, namely the number of hidden layers, the number of units in each hidden layer, and the type of activation function (**Appendix C**). The feed-forward backpropagation approach was utilized for training the ANN model whereby considering random initial values for weights and bias terms, the feed-forward propagation process happens. Then, the appropriate value of these parameters is achieved using the backpropagation method. In this study, the Levenberg-Marquardt backpropagation method was used whereby the cost function (mean squared error) was minimized and the optimum values for parameters (weights and bias) were obtained. The reader is referred to [40,41] for more information about the ANN model and its training procedure.

3.2.2. Support vector regression (SVR)

The SVR model follows the risk minimization principle [42] and the errors of predictions that are close to target values are ignored. Therefore, the SVR model has the privilege of accurate approximation capability even with a small-sized training dataset. The following equation is used in the SVR model to predict the outputs:

$$\hat{y} = f(X) = W \cdot \phi(X) + b \quad (3)$$

Where \hat{y} denotes the outputs, W is the weights, $\phi(X)$ represents kernel function, and b is the bias value. As shown in **Fig. 5**, the input values are fed into the input layer, and using the kernel layer the non-linearity is made and the input data are moved to a higher dimensional space, named feature space [43]. A commonly-used type of kernel function i.e. radial basis function (RBF) was utilized as the kernel function in this study. Considering x and x' as two samples, the kernel function is defined as follows:

$$\phi(X) = K(x_i, x_j) = \exp(-\gamma x_i - x_j^2) \quad (4)$$

Where γ in **Eq. (4)** is the kernel coefficient and is a hyper-parameter. By minimizing the following cost function, the best values of W and b are calculated.

$$\text{Cost function} = \frac{1}{2} W^2 + \frac{C}{N} \sum_{i=1}^N L(y_i, f(x_i)) \quad (5)$$

Where C , N , y_i are the penalty parameter, the number of training samples, and the actual target value, respectively. The penalty parameter (C) is another hyper-parameter that should be selected accurately to prevent underfitting and overfitting issues. The term $L(y_i, f(x_i))$ is the empirical error and formulated as:

$$L(y_i, f(x_i)) = \begin{cases} 0 & |y_i - f(x_i)| \leq \epsilon \\ |y_i - f(x_i)| - \epsilon & \text{Otherwise} \end{cases} \quad (6)$$

According to **Eq. (6)**, the empirical error is zero when the predicted value is placed in the ϵ tube. However, the slack parameters (ζ , ζ^*) are used to estimate errors for the predictions outside of the ϵ tube. To formulate the **Eqs. (5)** and **(6)**, the following equation can be considered

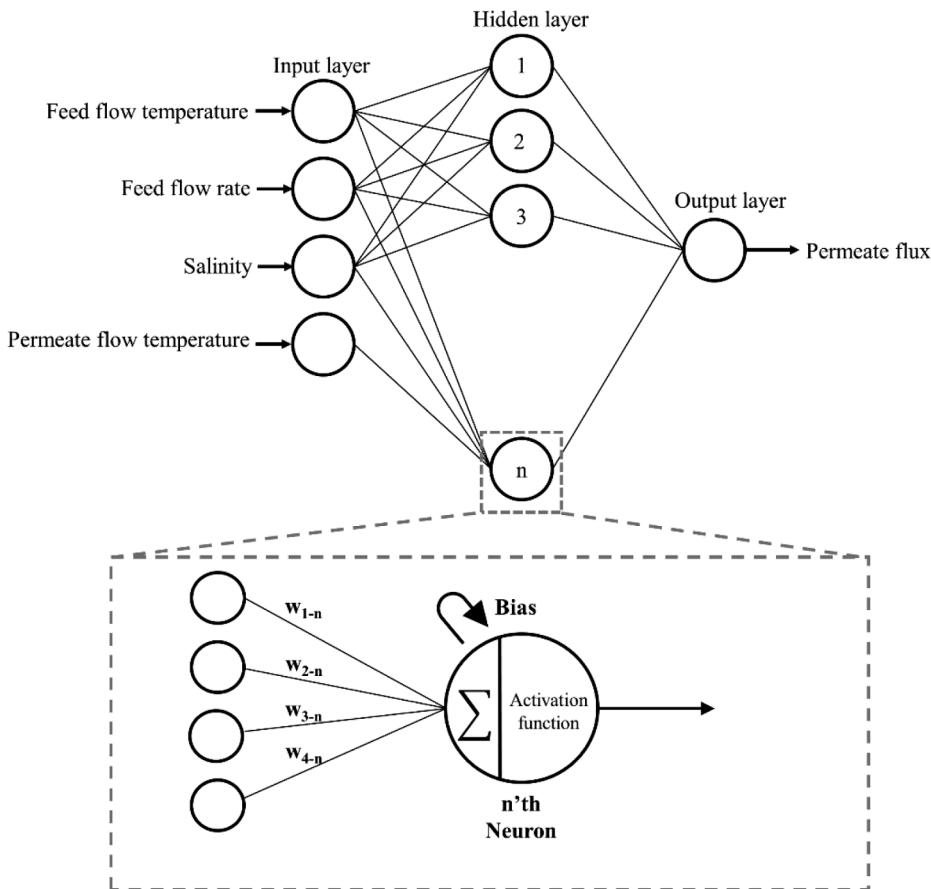


Fig. 4. Architecture of the ANN model.

as the overall optimization function:

$$\begin{aligned} \text{minimize}_{\zeta, \zeta^*, W, b} & \frac{1}{2} \|W\|^2 + C \sum_{i=1}^N L(y_i, f(x_i)) \text{ Subject to} \\ & : \begin{cases} y_i - (W \cdot \phi(X) - b) \leq \varepsilon + \zeta_i \\ (W \cdot \phi(X) - b) - Y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \quad (7) \end{aligned}$$

The final regression function is then achieved by using Lagrange multipliers (α_i , α_i^*) and Kernel function in the previous equation [44]:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) k(x_i, x_j) + b \quad (8)$$

3.2.3. Random forest (RF)

Random forest is a tree-based model that benefits from the simplicity of tree-based models while solving the weaknesses of conventional regression or classification trees. The decision tree model mainly is vulnerable to overfitting even with appropriate hyper-parameter tuning, meaning that the decision tree model has undesirable predictive performance on test data [45]. However, the overfitting issue has been solved in the RF model mainly due to randomness in selecting the bootstrapped datasets and creating the trees based on a random subset of features. RF model has received significant attention as a robust predictive machine learning model for the performance prediction of various energy systems [45,46]. This model is developed based on constructing a number of simple trees and the RF model's performance is estimated by the average of predictions. The training process of the RF model consists of three main stages as follows:

- Stage 1: Several subsets entitled “bootstrapped datasets” are formed using the main training dataset. The number of subsets is equal to the number of trees in the forest that is a hyper-parameter (Appendix C). It should be also noted that the size of training and bootstrapped datasets are equal meaning that the identical sample can be selected more than once.
- Stage 2: A bootstrapped dataset is used to create the tree and a random subset of features (i.e., feed flow temperature, feed flow rate, salinity of feed flow, and permeate flow temperature) is used to form the tree at each step.
- Stage 3: Stage 2 is repeated to create a new tree by employing a different bootstrapped dataset and accidentally selected features. This process is repeated until all the trees are created.

3.3. Performance indicators

Mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) were applied to compare the performance of developed machine learning models with the mechanistic modeling approach as follows [47]:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (10)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{(\hat{y}_i - y_i)}{y_i} \right| * 100 \quad (11)$$

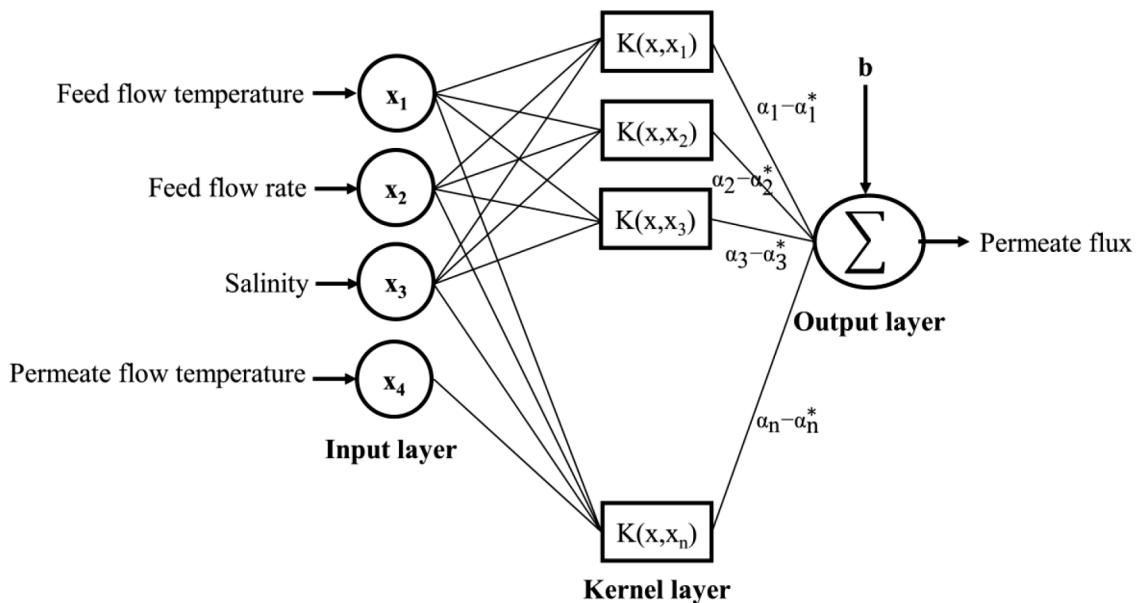


Fig. 5. The SVR model's schematic plot.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (12)$$

where y_i , \hat{y}_i , N and \bar{y} are corresponded to experimental data, the predicted value, the number of samples, and the mean value of the experimental data, respectively. The MAE measures the average error without consideration of the magnitude of errors whereas the RSME takes into consideration the magnitude of errors by placing more weights on larger errors since errors are squared before averaging. Compared to MAE and RMSE metrics, the MAPE expresses the deviation between the actual and predicted values in terms of percentage by normalizing the errors. Further, R^2 indicates the degree to which the variance in the features can be explained by the independent variables and represent the goodness of the fit [48]. Lower values of MAE, RMSE, MAPE, and higher values of R^2 represent the better performance of models.

4. Results and discussions

As mentioned before, MATLAB was used to develop the ANN and mechanistic models, whereas the SVR and RF models have been implemented using the scikit-learn library of python language. The hyper-parameter tuning process is comprehensively explained for obtaining the best generalization ability of each machine learning model and the results are presented in Appendix C. In this section, the outlier detection analysis is performed to detect the suspected data and verify the reliability of the machine learning models. A detailed comparison is then made between mechanistic modeling approach and machine learning models in terms of models' accuracy, trend predictability, and computational time. Finally, the most influential operational parameters are determined by feature importance analysis of the RF model.

4.1. Outlier detection

The validity of the developed machine learning models depends strongly on the collected experimental dataset. Hence, suitable experimental datasets should be employed for developing machine learning models by diagnosing suspected data samples (outliers) and removing them from the experimental dataset. In this investigation, the Leverage approach has been applied to detect the outliers and William's plot has been drawn correspondingly [49]. William's plot illustrates standardized residual of predicted and actual target values versus Hat values. The

Hat values are obtained from diagonal arrays of the hat matrix (H) defined as [50]:

$$H = X (X^T X)^{-1} X^T \quad (13)$$

X is an $n \times k$ matrix, where n and k represent the number of data samples and the number of input features, respectively. According to the Leverage approach, the acceptable domain is restricted by $0 < \text{hat values} < H^*$ and $-3 < \text{standardized residuals} < 3$. Moreover, H^* is the Warning leverage and is defined as [50]:

$$H^* = 3 \frac{k+1}{n} \quad (14)$$

In this study, the number of data samples (n) and the number of input features (k) are 70 and 4, respectively, and according to Eq. (14), H^* is equal to 0.21 Fig. 6. illustrates William's plot for the prediction of permeate flux using the three machine learning models. It is observed that the vast majority of the data are within the ranges $-3 < \text{standardized residuals} < 3$ and $0 < \text{hat values} < 0.21$, which confirms the validity of the developed machine learning models. It is noteworthy that removing the outliers could lead to the better performance of the developed machine learning models, however, all experimental data samples have been used in this study for training and testing the models.

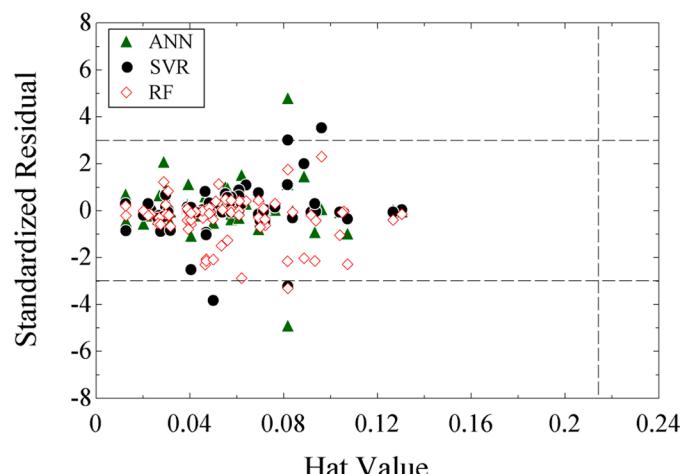


Fig. 6. Detection of outliers using William's plot.

Table 7

Performance of machine learning and mechanistic modeling approach in the prediction of permeate flux.

Model	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
ANN	0.51	1.28	1.62	0.99	0.88	1.10	3.46	0.99
SVR	0.76	1.70	2.21	0.99	1.56	2.31	4.78	0.98
RF	1.11	1.93	3.39	0.99	2.78	3.47	9.92	0.95
Mechanistic	1.89	2.81	6.18	0.98	2.55	4.53	7.31	0.92

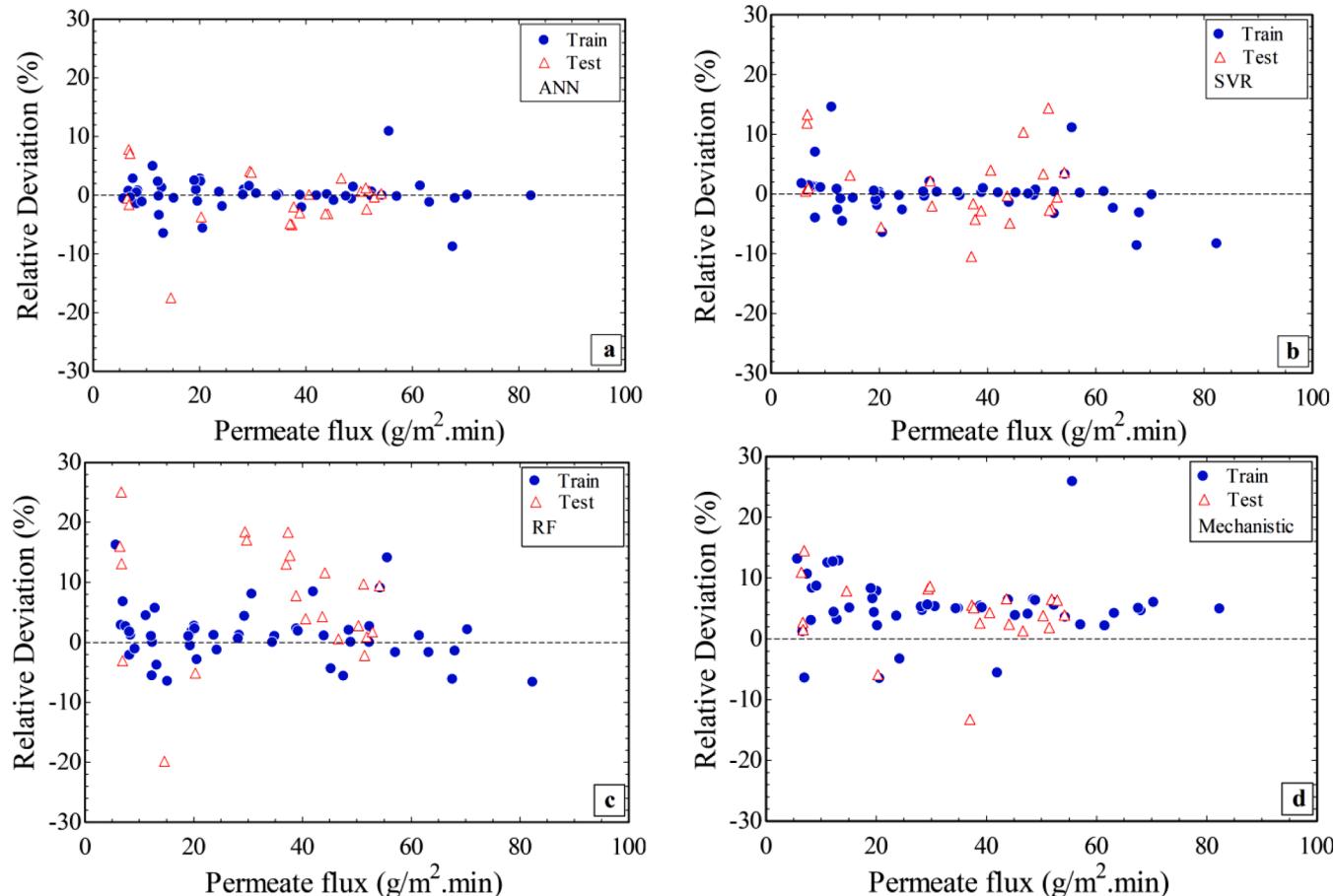


Fig. 7. Performance of developed models based on the relative deviation.

4.2. Performance comparison of the developed models

Table 7 indicates the predictive performance of three machine learning models compared with the stepwise mechanistic modeling approach in terms of statistical performance indicators. High values of R² and low errors (MAE, RMSE, and MAPE) represent the capability of the developed models for the accurate prediction of permeate flux of the DCMD module. This should be mentioned that training and test datasets for the mechanistic modeling approach represent the similar experimental training and testing datasets used for training and testing the machine learning models. This provided a fair comparison between the predictive performance of mechanistic modeling approach and machine learning models based on identical experimental data. As seen in Table 7, all models exhibit acceptable predictive performance for both training and testing datasets, possessing a maximum MAPE_{test} of 9.92% for the RF model. Amongst machine learning models, ANN, SVR, and RF models ranked first, second, and third in terms of predictive ability on the test dataset. This is mainly attributed to the higher R² and lower errors of ANN, SVR, and RF models for the prediction of the test data, respectively. Further, ANN and SVR models show much better predictive performance as compared to the mechanistic modeling approach. As

shown in Table 7, the MAPE_{test} for the mechanistic modeling approach is 7.31%, while that for ANN and SVR models are 3.46% and 4.78%, respectively. It can be also seen that MAE_{test} for ANN and SVR models stood at 0.88 and 1.56 whereas MAE_{test} for mechanistic modeling approach was 2.55, meaning that compared with the mechanistic modeling approach, MAE_{test} for ANN and SVR models were lower by approximately 65.3% and 38.61%, respectively.

To provide a better comparison amongst the machine learning models and the mechanistic modeling approach, the relative deviation of models' predictions from the experimental data is graphically illustrated in Fig. 7. Training the machine learning models using a small number of data may cause the underfitting problem which is known as the inaccurate predictive performance of machine learning models on training and testing datasets. As mentioned before, in this study, 48 data samples have been used to train the machine learning models. However, as shown in Fig. 7, three machine learning models (ANN, SVR, and RF) exhibited acceptable performance on the training dataset. Further, amongst three machine learning models, ANN and SVR models show superior performance on the training dataset, where relative deviations of training samples are around zero. This can be justified by the excellent capability of ANN and SVR models in the training process even with the

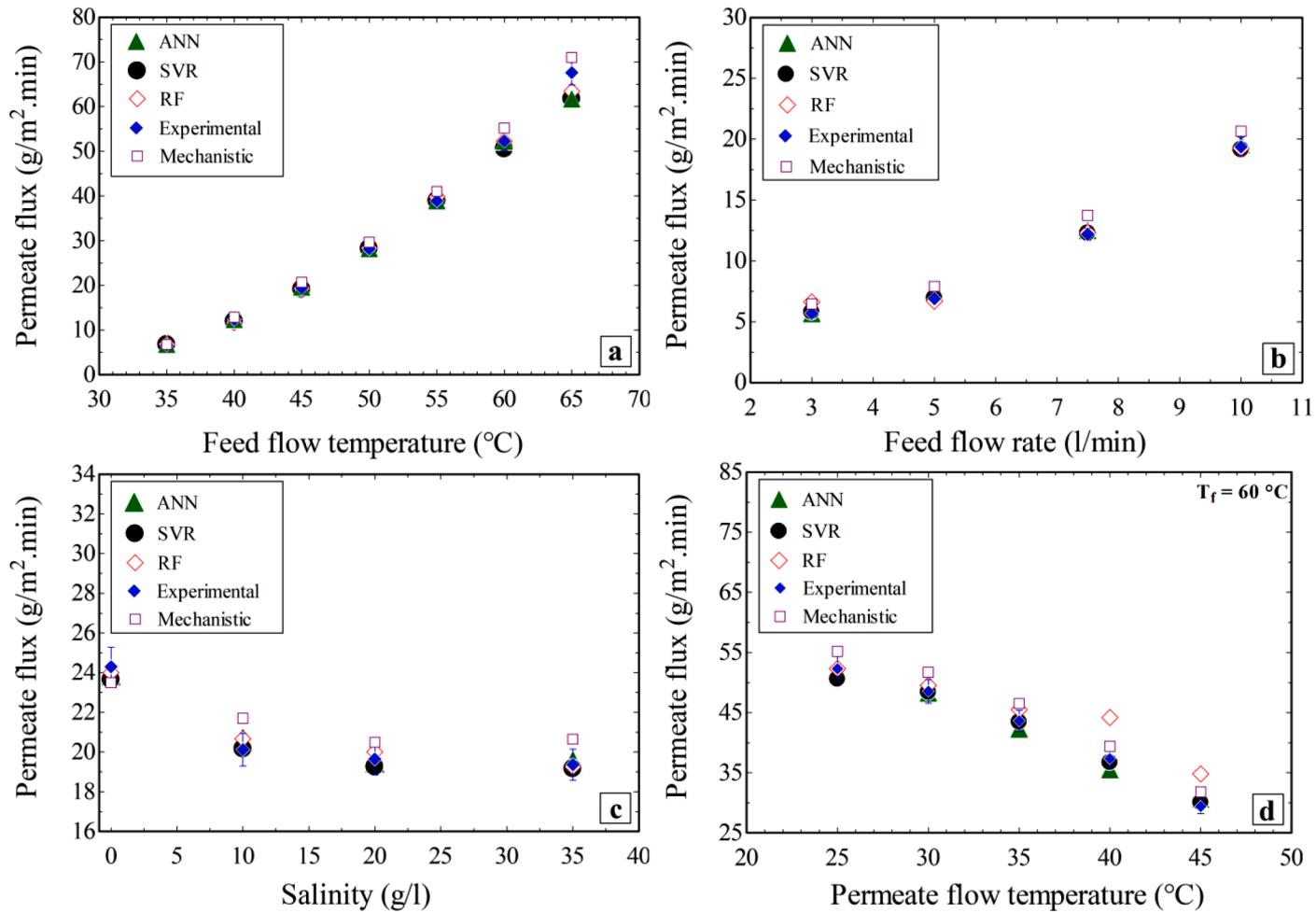


Fig. 8. Effect of operational parameters on the permeate flux of DCMD module.

Table 8

Computation time for machine learning models and the mechanistic modeling approach.

Model	Computation time for test dataset (ms)
ANN	95.05
SVR	1
RF	2.98
Mechanistic	343,000

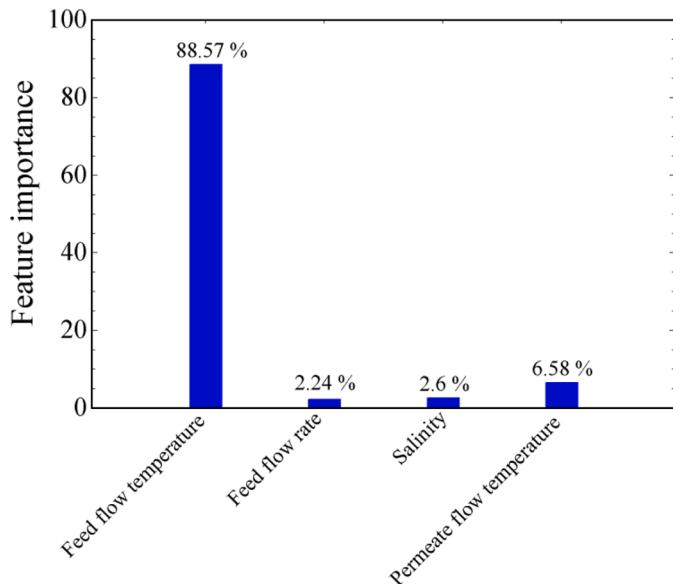


Fig. 9. Feature importance analysis conducted by the RF model.

small size of datasets [46,51]. Further, the ANN model shows the best predictive performance based on the accurate prediction of testing data samples. In contrast, Fig. 7(c) illustrates that the testing data samples predicted by the RF model have widely scattered that also justifies the reported results in Table 7. Higher relative deviations of testing data samples can also be seen (12–18%) in the prediction of permeate flux by RF model in the range of 25–35 g/m².min. It can be inferred from Fig. 7(d) that the mechanistic modeling approach tends to overestimate permeate flux. Moreover, relative deviations of the majority of the predicted values have been restricted to 0–12% approximately which confirms the relative acceptable predictive performance of the stepwise mechanistic modeling approach.

Apart from the predictive performance investigation of the developed models based on the statistical performance indicators and relative deviation analysis, the predictive ability of the models is also determined using trend analysis. Fig. 8 compares the ability of three machine learning models with that of the mechanistic modeling approach for predicting the effect of each operational parameter on the permeate flux of the DCMD module. To conduct the trend analysis, when an operational parameter is varied, the others are remained constant at a base case condition (feed flow temperature = 45 °C, feed flow rate = 10 l/min, salinity = 35 g/l, and permeate flow temperature = 25 °C). It should be also mentioned that feed flow temperature was considered to be 60 °C in the case of analyzing the effect of permeate flow temperature (Fig. 8(d)). It can be inferred from Fig. 8(a-d) that the mechanistic modeling approach mostly tends to overestimate the permeate flux as compared to the machine learning models. Further, there is a close agreement between the predictions of all the developed models and experimental data samples, highlighting the robustness of the developed models. According to Fig. 8(a), the permeate flux increases significantly as feed flow temperature rises. The mechanistic modeling approach shows more deviations from the experimental data at higher feed flow

temperatures. Moreover, except for T_f = 65 °C, the ANN and SVR models showed a better performance for analyzing the effect of feed flow temperature on the permeate flux. As illustrated in Fig. 8(b), the three machine learning models mostly exhibit better performance as compared to the mechanistic modeling approach for the prediction of permeate flux, especially at higher feed flow rates. It can also be seen that with an increase in the feed flow rate, the permeate flux rises mainly due to the lower concentration polarization effect. Fig. 8(c) indicates that an increase in salinity of feed flow leads to an insignificant decrease in the permeate flux and ANN and SVR models show superiority over the mechanistic modeling approach for studying the effect of salinity. As can be observed from Fig. 8(d), higher permeate flow temperatures lead to a lower permeate flux due to lower vapor pressure difference as the main driving force of the vaporization phenomenon. Further, Fig. 8(d) shows that the RF model had a higher deviation from experimental data at higher permeate flow temperatures.

Apart from models' accuracy and trend predictability, the computational time is also considered as an important indicator for a modeling tool. In the case of the stepwise mechanistic modeling approach, computational time rises significantly with increasing the membrane length since integral heat and mass transfer equations are solved simultaneously along the membrane length. As a result, accurate performance prediction of DCMD modules using stepwise mechanistic modeling approach requires significant time, especially for large-scale and industrialized DCMD modules. Furthermore, to optimize the permeate flux of the DCMD module under different working conditions, there is a need for coupling a fast and accurate modeling tool with an optimization algorithm. However, coupling the stepwise mechanistic modeling tool with an optimization algorithm (such as the genetic algorithm) requires extensive computational time. As shown in Table 8, in this study, a comparison was made between the required computational time for the prediction of permeate flux by machine learning models and the stepwise mechanistic modeling approach under different operational conditions. It should be mentioned this comparison has been performed on the test dataset, which acts as unseen data and was not fed into the machine learning models in the training process. Moreover, the stepwise mechanistic modeling approach and the three machine learning models have been developed using a personal computer having 2 GHz CPU and 8 GB RAM. It can be observed that the machine learning models had much lower computational time as compared to the mechanistic modeling approach. Further, amongst machine learning models, SVR, RF, and ANN had respectively superiority in terms of computational time. The results obtained from Table 8 reveal that machine learning models require much lower computational time than the stepwise mechanistic modeling approach for performance prediction of the DCMD system. Therefore, machine learning models can be considered as a fast tool for performance prediction of DCMD modules, particularly for large-scale modules where the application of the mechanistic modeling approach seems time-consuming and tedious.

4.3. Importance of influencing parameters

The performance of machine learning models is affected by the features (inputs variables) that models are built on. Therefore, identifying the most influential inputs can facilitate the development of more effective machine learning models for the performance prediction of the DCMD module. In this study, the contribution of each operational parameter towards the prediction of the permeate flux of the DCMD module is estimated using the RF model. To develop a tree in the forest, the selection of variables for split nodes is based on the variables that lead to greater variance reduction, and feature importance is then determined by averaging all over the trees in the forest. As can be observed in Fig. 9, the feed flow temperature shows the most significant effect on the prediction of the permeate flux, possessing a feature importance of 88.57%. This is mainly due to the fact that the feed flow temperature plays a prominent role in the vaporization process as the

main driving force in the DCMD system [25]. However, the permeate flow temperature, salinity, and feed flow rate come in the second, third, and fourth ranks, with feature importance values of 6.58%, 2.6%, and 2.24%, respectively.

5. Conclusion

In this investigation, the permeate flux of the DCMD system has been predicted using the stepwise mechanistic modeling approach and three machine learning models namely artificial neural network (ANN), support vector regression (SVR), and Random forest (RF). An experimental dataset consisting of 70 samples was acquired for validating the mechanistic modeling approach as well as training and testing the machine learning models. Results showed that three machine learning models developed by 70 samples had an acceptable performance on both training data and unseen data (test data) which highlights their validity and accuracy. A comprehensive comparison was then performed to demonstrate the predictive performance of the proposed models in terms of models' accuracy, trend predictability, and computational time. The main findings of this study are:

- Statistical indicators and relative deviation analysis showed that amongst machine learning models, the ANN, SVR, and RF models, respectively, possessed the best accuracy to predict the permeate flux of the DCMD module. Further, the ANN and SVR models showed better performance over the mechanistic modeling approach where $MAPE_{test}$ for ANN, SVR, and mechanistic modeling approach reached 3.46%, 4.78%, and 7.31%, respectively.
- The parametric study results revealed that the mechanistic modeling approach mostly tend to overestimate the permeate flux compared with the experimental data.
- The mechanistic modeling approach showed more deviations from the experimental data at higher feed flow temperatures. Moreover, except for $T_f = 65^\circ\text{C}$, the ANN and SVR models showed better performance for analyzing the effect of feed flow temperature on the permeate flux.
- In the case of investigating the effect of feed flow rate on the permeate flux, the three machine learning models mostly exhibited better performance as compared to the mechanistic modeling approach, especially at higher feed flow rates.
- ANN and SVR models showed better performance over the mechanistic modeling approach for studying the effect of salinity. However, the RF model had a higher deviation from experimental data at higher permeate flow temperatures.
- The machine learning models had by far lower computational time than the stepwise mechanistic modeling approach. This highlights the significant superiority of the machine learning models over the stepwise mechanistic modeling approach in terms of computational cost, especially for large-scale DCMD modules.
- Feature analysis made by the RF model showed that the most influential operational parameters on the permeate flux were feed flow temperature, permeate flow temperature, salinity, and feed flow rate, respectively.

This study demonstrated that the machine learning models can be

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.cep.2022.108857](https://doi.org/10.1016/j.cep.2022.108857).

Appendix A. Experimental data and predictions made by the machine learning models

Table A.1

considered as robust predictive models for the prediction of the permeate flux of the DCMD system. Hence, developing machine learning models considering both design and operational parameters is a promising research potential. The development of machine learning models using design parameters such as module length and membrane materials along with operational parameters can play an important role in designing more compact DCMD modules with higher productivity. Further, the possibility of employing machine learning models for the prediction of other important performance indicators of DCMD modules such as gained output ratio (GOR), cost of freshwater, and temperature polarization coefficient can be investigated for future studies, resulting in efficiency enhancement and cost reduction of DCMD modules. Moreover, investigating the predictive performance of RSM and machine learning models for performance analysis of the DCMD module appears a viable research area that deserves more attention in future studies.

Authorship contributions

Please indicate the specific contributions made by each author (list the authors' initials followed by their surnames, e.g., Y.L. Cheung). The name of each author must appear at least once in each of the three categories below.

Category 1

Conception and design of study: P Behnam and M Khiadani, Acquisition of data: P Behnam, and A Shafieian

Analysis and/or interpretation of data: P Behnam, A Shafieian, and M Khiadani

Category 2

Drafting the manuscript: P Behnam and A Shafieian

Category 3

Approval of the version of the manuscript to be published (the names of all authors must be listed):

P Behnam, A Shafieian, M Khiadani, and M Zargar.

Revising the manuscript critically for important intellectual content P Behnam, A Shafieian,

M Khiadani, and M Zargar

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.

Acknowledgement

All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements and have given us their written permission to be named. If we have not included an Acknowledgements, then that indicates that we have not received substantial contributions from non-authors.

Table A.1

Experimental data and predictions made by the machine learning models.

Sample	Type	Feed flow temperature (°C)	Feed flow rate (l/min)	Salinity of feed flow (g/l)	Permeate flow temperature (°C)	Experimental permeate flux (g/m².min)	Permeate flux prediction by ANN (g/m².min)	Permeate flux prediction by SVR (g/m².min)	Permeate flux prediction by RF (g/m².min)
1	Train	65	10	20	25	68.05	67.70	65.91	67.07
2	Train	45	3	0	25	8.4	8.46	8.50	8.5
3	Train	35	10	35	25	6.67	6.71	6.77	6.86
4	Train	60	10	35	30	48.54	48.23	48.43	49.52
5	Train	45	5	20	25	7.5	7.71	7.60	7.7
6	Train	45	10	10	25	20.12	20.68	20.17	20.66
7	Train	65	10	35	25	67.6	61.67	61.78	63.44
8	Train	50	10	20	25	28.37	28.64	28.26	28.71
9	Train	40	10	20	25	12.4	11.98	12.50	12.4
10	Train	45	5	10	25	8.2	8.08	7.87	8.02
11	Train	45	10	20	25	19.65	19.44	19.27	20
12	Train	65	10	35	25	55.6	61.67	61.78	63.44
13	Train	60	10	10	45	30.67	30.76	30.76	33.14
14	Train	60	10	20	30	48.9	49.60	49.24	48.91
15	Train	60	10	0	25	63.23	62.47	61.74	62.17
16	Train	45	7.5	10	25	11.2	11.75	12.83	11.7
17	Train	55	10	35	25	38.9	38.91	39.00	39.78
18	Train	45	10	10	25	20.21	20.68	20.17	20.66
19	Train	60	10	0	45	34.9	34.95	34.79	35.25
20	Train	60	10	10	25	54.31	54.30	56.08	59.24
21	Train	60	10	0	25	61.48	62.47	61.74	62.17
22	Train	45	10	20	25	20.6	19.44	19.27	20
23	Train	40	10	10	25	12.89	13.05	12.79	13.62
24	Train	65	10	0	25	82.34	82.28	75.50	76.88
25	Train	60	10	0	30	57.1	56.99	57.19	56.14
26	Train	45	7.5	20	25	13.2	12.34	12.60	12.7
27	Train	35	10	0	25	8.2	8.23	8.77	8.34
28	Train	65	10	35	40	42	41.98	42.09	45.54
29	Train	65	10	10	25	70.36	70.40	70.26	71.85
30	Train	55	10	0	25	47.56	47.48	47.57	44.87
31	Train	60	10	20	35	43.97	44.03	43.35	44.45
32	Train	55	10	20	25	39.19	38.38	39.55	39.93
33	Train	45	3	10	25	7	6.97	7.09	7.47
34	Train	45	10	35	25	19.37	19.55	19.17	19.25
35	Train	40	10	0	25	15.18	15.10	15.07	14.19
36	Train	50	10	35	25	28.17	28.19	28.27	28.34
37	Train	45	5	0	25	9.2	9.09	9.29	9.1
38	Train	45	10	35	25	19.07	19.55	19.17	19.25
39	Train	50	10	0	25	34.48	34.46	34.58	34.48
40	Train	45	10	0	25	24.3	23.84	23.65	23.99
41	Train	60	10	20	25	52.37	52.68	52.55	53.76
42	Train	45	3	35	25	5.7	5.66	5.80	6.62
43	Train	40	10	35	25	12.32	12.30	11.99	11.63
44	Train	60	10	35	25	52.3	52.22	50.58	52.3
45	Train	50	10	10	25	29.36	29.81	29.96	30.64
46	Train	60	10	10	35	45.25	44.84	45.35	43.26
47	Train	45	7.5	35	25	12.2	12.48	12.29	12.32
48	Train	45	10	0	25	23.71	23.84	23.65	23.99
49	Test	45	7.5	0	25	14.6	12.04	15.05	11.7
50	Test	60	10	35	45	29.39	30.60	30.03	34.80
51	Test	60	10	0	40	44.1	42.71	41.93	49.21
52	Test	60	10	35	35	43.63	42.25	43.47	45.49
53	Test	60	10	35	25	51.83	52.22	50.58	52.3
54	Test	35	10	10	25	6.67	7.19	7.46	8.34
55	Test	60	10	10	30	50.28	50.61	51.96	51.67
56	Test	55	10	10	25	40.55	40.60	42.16	42.14
57	Test	60	10	20	25	52.84	52.68	52.55	53.76
58	Test	60	10	35	40	37.34	35.46	36.73	44.18
59	Test	65	10	35	45	37	35.21	33.12	41.82
60	Test	60	10	20	40	37.68	36.93	36.07	43.14
61	Test	45	3	20	25	6.4	6.36	6.42	7.42
62	Test	60	10	0	35	51.4	50.18	49.99	50.26
63	Test	35	10	20	25	6.7	6.59	7.59	7.57
64	Test	60	10	10	25	54.14	54.30	56.08	59.24
65	Test	65	10	35	30	51.2	51.84	58.56	56.16
66	Test	45	5	35	25	6.9	7.38	6.96	6.68
67	Test	45	10	35	25	20.3	19.55	19.17	19.25
68	Test	60	10	10	40	38.8	37.63	37.71	41.80
69	Test	60	10	20	45	29.73	30.88	29.13	34.80
70	Test	65	10	35	35	46.6	47.94	51.41	46.85

Appendix B. Equations used in the mechanistic modeling approach

• Energy and mass balance

The transferred energies by heat and mass comprise the feed and permeate steady state energy balance equations. These equations have been developed by considering a steady-state condition with no energy input in form of work and constant potential energy across each section:

$$E_{f-j,in} - E_{f-j,out} = \frac{dE_{f-cv}}{dt} \quad (B.1)$$

$$E_{f-j,in} = E_{f-j,out} \quad (B.2)$$

$$m_{f,j} \left(Y_{f,j} + \frac{V_{f,j}^2}{2} \right) = m_{f,j+1} \left(Y_{f,j+1} + \frac{V_{f,j+1}^2}{2} \right) + Q_j A_{m,j} \quad (B.3)$$

$$m_{p,j} \left(Y_{p,j} + \frac{V_{p,j}^2}{2} \right) = m_{p,j+1} \left(Y_{p,j+1} + \frac{V_{p,j+1}^2}{2} \right) + Q_j A_{m,j} \quad (B.4)$$

As mass transfer across the membrane occurs, the mass flow rate passing each section is different. The feed and permeate mass flow rates of each section can be determined by:

$$m_{f,j+1} = m_{j,n} - J_{w,j} A_{m,j} \quad (B.5)$$

$$m_{p,j+1} = m_{p,j} - J_{w,j} A_{m,j} \quad (B.6)$$

• Heat and mass transfer

To model the heat and mass transfer in the DCMD module, it was assumed that: (a) heat loss to the surroundings is insignificant; (b) feed and permeate streams are in steady state and incompressible conditions; (c) the momentum of the permeate through the membrane is negligible; and (d) Dusty-Gas model applies for the mass transfer of fluid mixtures through the membrane [52]. The mass flux through the membrane in a DCMD module is linearly related to the vapor pressure difference across the membrane [13]:

$$J_w = C_m [P_v(T_{f,m}, S_{f,m}) - P_v(T_{p,m}, S_{p,m})] \quad (B.7)$$

The equations proposed by Sharqawy et al [53] were the main source for calculating the thermo-physical characteristics of feed and permeate streams. The temperature and salinity at the membrane surfaces need to be determined first before Eq. (B.7) can be used to calculate the mass flux through the membranes. The equations to find these values will be covered in the next sections.

The heat transfer in a DCMD module consists of three processes: (a) Q_f : the thermal energy transfer between the bulk feed stream and the boundary layer forming adjacent to the membrane surface; (b) Q_m : combination of membrane conductive heat transfer and water vapor movement through the membrane; (c) Q_p : the thermal energy transfer between the permeate boundary layer and the permeate stream. Q_f , Q_m , and Q_p are defined as:

$$Q_f = h_f (T_f - T_{f,m}) \quad (B.8)$$

$$Q_m = h_m (T_{f,m} - T_{p,m}) + J_w H_v (T_{f,m}, S_{f,m}) \quad (B.9)$$

$$Q_p = h_p (T_p - T_{p,m}) \quad (B.10)$$

The readers are referred to [12,54] for more information regarding the calculation processes of heat transfer coefficients.

Film theory on the feed channel boundary layer along with the mass balance equation was used to determine the salinity at the membrane wall [55]:

$$S_{f,m} = S_f \exp \left(\frac{J_w}{\rho_f K} \right) \quad (B.11)$$

More information regarding the calculation processes and equations for determining the film mass transfer coefficient (K), Sherwood number (S_h), and Schmidt number (S_c) can be found in [56,57].

The flow type which occurs inside the membrane can be evaluated quantitatively by the Knudsen number [57]:

$$Kn = \frac{W}{d} \quad (B.12)$$

where W and d are respectively the transferred molecules mean free path and pore diameter. Depending on the range of Knudsen number, the flow can be categorized as: (a) molecular diffusion (<0.01) where molecule-molecule collisions are more noticeable than molecule-pore surface collisions; (b) Knudsen mechanism ($Kn>1$) which is opposite of molecular diffusion; and (c) Knudsen-molecular diffusion ($0.01<Kn<1$) which is the transition process. It is worth noting that under specific conditions, the transition range can go up to 10 [58]. The readers are referred to [59,60] for detailed calculation processes.

$T_{f,m}$ and $T_{p,m}$ can be calculated using iterative computer modeling. First, the feed and permeate side surface temperatures are assumed to be equal to the bulk temperatures of feed and permeate streams (i.e., T_f and T_p). Then, new values of $T_{f,m}$ and $T_{p,m}$ are calculated by:

$$T_{f,m} = \frac{h_m \left(T_p + \left(\frac{h_p}{h_f} \right) T_f \right) + h_f T_f + J_w H_v}{h_m + h_f \left(1 + \frac{h_m}{h_p} \right)} \quad (B.13)$$

$$T_{p,m} = \frac{h_m \left(T_f + \left(\frac{h_p}{h_f} \right) T_p \right) + h_p T_p + J_w H_v}{h_m + h_p \left(1 + \frac{h_m}{h_f} \right)} \quad (B.14)$$

The mentioned process is repeated until an acceptable accuracy is achieved. The estimated values in the last iteration are used to determine the water productivity of the module.

Appendix C. Hyper-parameter tuning

Appropriate selection of hyper-parameters plays a key role in enhancing the accuracy of machine learning models and their generalization capability. Both optimization and trial and error methods have been broadly applied to tune the hyper-parameters for developing machine learning models [44,46]. In this study, a stepwise method based on trial and error has been applied for the hyper-parameter tuning process. Regarding the ANN model, the number of hidden layers, the number of units in each hidden layer, and the type of activation functions are considered as the main hyper-parameters. To analyze the effect of the number of hidden layers, four neurons were considered in each hidden layer and log-sigmoid activation functions were used in the hidden layer. As shown in Table C.1, creating the ANN model by one hidden layer resulted in the best performance. Further, the effect of the activation function of the hidden layer was estimated by choosing two different types i.e. log-sigmoid and tansig (Table C.2). Results showed that developing the ANN model based on one hidden layer and choosing log-sigmoid activation functions could lead to the acceptable performance of the ANN model.

As a result, one hidden layer was considered for developing the ANN model and sigmoid type was chosen as the activation function. The selection of the optimum number of neurons for the hidden layer of the ANN model is shown in Table C.3. According to Table C.3, the enhanced training process of the ANN model was achieved by increasing the number of neurons, whereas the best model's generalization capability can be seen in the case of eight neurons. The selection of eight neurons prevented the ANN model from underfitting and overfitting issues and therefore the 4–8–1 architecture was chosen as the best architecture for developing the training and testing the model.

In the case of the SVR model, penalty parameter (C), coefficient of the kernel (γ), and radius (ϵ) should be properly tuned as the main hyper-parameters. To tune the C value, the two other hyper-parameters were fixed at their default value in the scikit-learn library ($\gamma = 0.25$, $\epsilon = 0.1$) [38] Table C.4. shows the effect of penalty parameter (C) on the predictive performance of the SVR model for training and testing datasets. It can be observed that $C = 150$ produced the highest R^2 and lowest RMSE for testing datasets and therefore was chosen as the best penalty parameter value. Further, it was revealed that variations in γ and ϵ had no significant effect on the predictive performance of the SVR model and therefore these hyper-parameters remained constant at their default values. As a result, $C = 150$, $\gamma = 0.25$, and $\epsilon = 0.1$ were selected for training and testing the SVR model.

The RF model possesses four hyper-parameters namely the number of trees in the forest, maximum depth of each tree, the minimum sample data for splitting, and the maximum number of variables that are chosen for the best split [45]. The trial and error method showed that the best performance of the RF model was achieved when the maximum number of variables was fixed at four (Table C.5), and the minimum sample data for splitting was set at its minimum value i.e. two (Table C.6). It should be mentioned that to investigate the effect of the maximum number of variables, the other three hyper-parameters were fixed at their default values in scikit learn library (number of tree= 10, maximum depth of each tree = none, and the minimum sample data for splitting = 2). Also, the effect of the minimum sample data for splitting on the performance of RF model was estimated when the number of trees and the maximum depth of each tree were fixed at their default values, whereas the maximum number of variables was fixed at its optimum value (i.e. four).

There existed optimum values for the number of trees as well as the maximum depth of each tree in the forest Table C.7. depicts the impact of the maximum depth of trees in the forest on the performance indicators for both training and testing datasets. It should be noted that the number of trees was initially fixed at default value in scikit learn library (i.e. ten), and the minimum sample data for splitting and the maximum number of variables

Table C.1
Selection of the optimum number of hidden layers of the ANN model.

Number of Hidden layer	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE	R^2
1	0.82	1.44	3.21	0.99	1.28	1.62	4.23	0.99
2	0.69	1.39	2.40	0.99	1.62	2.41	5.11	0.97
3	0.55	1.29	1.75	0.99	1.85	3.06	12.36	0.96
4	0.40	1.25	1.09	0.99	3.61	6.00	12.58	0.86

Table C.2
Selection of the activation function of the hidden layer for the ANN model.

Activation function	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE	R^2
Log-sigmoid	0.82	1.44	3.21	0.99	1.28	1.62	4.23	0.99
Tansig	1.49	1.97	6.26	0.99	1.85	2.23	6.39	0.98

Table C.3

Selection of the optimum number of neurons for the hidden layer of the ANN model.

Number of Neurons	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
3	0.90	1.54	3.89	0.99	1.50	1.88	6.43	0.98
4	0.82	1.44	3.21	0.99	1.28	1.62	4.23	0.99
5	0.66	1.36	2.07	0.99	0.94	1.25	5.14	0.99
6	0.47	1.27	1.21	0.99	0.96	1.19	6.91	0.99
7	0.57	1.32	1.89	0.99	0.93	1.25	5.19	0.99
8	0.51	1.28	1.62	0.99	0.88	1.10	3.46	0.99
9	0.36	1.24	0.86	0.99	1.61	2.24	10.72	0.98
10	0.38	1.24	0.87	0.99	2.01	2.59	8.7	0.97

Table C.4

Effect of C on SVR model performance.

Penalty parameter (C)	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
100	0.87	1.83	2.72	0.99	1.72	2.33	5.95	0.97
150	0.76	1.70	2.21	0.99	1.56	2.31	4.78	0.98
200	0.68	1.62	1.92	0.99	1.45	2.36	4.33	0.97
250	0.64	1.59	1.83	0.99	1.39	2.41	3.98	0.97
300	0.60	1.57	1.75	0.99	1.39	2.48	4.02	0.97
350	0.57	1.62	1.67	0.99	1.44	2.63	3.96	0.97

Table C.5

Effect of maximum number of variables on the performance of RF model.

Maximum number of variables	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
1	1.75	2.60	7.67	0.98	4.05	5.11	22.05	0.90
2	1.56	2.40	6.07	0.98	2.78	3.73	14.35	0.94
3	1.29	2.02	5.68	0.99	3.02	3.54	15.37	0.95
4	1.09	1.87	4.02	0.99	2.94	3.52	11.7	0.95

Table C.6

Effect of the minimum sample data for splitting on the performance of RF model.

Minimum sample	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
2	1.09	1.87	4.02	0.99	2.94	3.52	11.7	0.95
3	1.23	2.10	5	0.98	3.06	3.71	13	0.94
4	1.50	2.31	5.97	0.98	3.07	3.86	12.93	0.94

Table C.7

Effect of maximum depth of trees on the performance of RF model.

Maximum depth	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
4	1.98	3.01	8.63	0.97	3.26	4.19	15.28	0.93
6	1.09	1.81	4.24	0.99	3.06	3.78	12.6	0.94
8	1.09	1.88	4.02	0.99	2.99	3.61	11.86	0.95
10	1.09	1.87	4.02	0.99	2.94	3.52	11.7	0.95
12	1.09	1.87	4.02	0.99	2.94	3.52	11.7	0.95

Table C.8

Effect of the number of trees on the performance of the RF model.

Number of Trees	Training dataset				Test dataset			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
5	1.13	1.95	3.78	0.99	3.46	4.70	12	0.91
6	1.21	2.01	3.71	0.99	3.27	4.08	11.4	0.93
7	1.20	2.02	3.66	0.99	3.07	3.78	10.77	0.94
8	1.11	1.93	3.39	0.99	2.78	3.47	9.92	0.95
9	1.15	1.98	4.25	0.99	2.87	3.43	11.66	0.95
10	1.09	1.87	4.02	0.99	2.94	3.52	11.7	0.95
12	1.19	1.941	5.16	0.99	2.98	3.58	12.86	0.95

were chosen as two and four, respectively. As can be seen from Table C.7, the best training and testing performance can be seen in the case of selecting ten for the maximum depth of trees and a further increase in the maximum depth of trees had no improving effect, thus selecting the maximum depth of trees at ten. To determine the optimum number of trees, the obtained results listed in Table C.8 depict that developing the forest by eight trees led to the best generalization capability. It can be seen from Table C.8 that in the case of eight trees, the lowest MAE and MAPE errors and acceptable high values of performance indicators for the training dataset can be achieved. Finally, the number of trees, maximum depth of each tree, the minimum sample data for splitting, and the maximum number of variables that are chosen for the best split were selected as eight, ten, two, and four, respectively.

References

- [1] N. Ghaffour, S. Soukane, J.G. Lee, Y. Kim, A. Alpatova, Membrane distillation hybrids for water production and energy efficiency enhancement: A critical review, *Appl. Energy* 254 (2019), 113698.
- [2] C.-K. Chiam, R. Sarbatly, Heat transfer in the rectangular cross-flow flat-sheet membrane module for vacuum membrane distillation, *Chem. Eng. Process.* 79 (2014) 23–33.
- [3] D. González, J. Amigo, F. Suárez, Membrane distillation: Perspectives for sustainable and improved desalination, *Renewable Sustainable Energy Rev.* 80 (2017) 238–259.
- [4] M.R. Elmarghany, A.H. El-Shazly, M.S. Salem, M.N. Sabry, N. Nady, Thermal analysis evaluation of direct contact membrane distillation system, *Case Studies in Thermal Engineering* 13 (2019), 100377.
- [5] Q.L. Ve, R. Koirala, M. Bawahab, H. Faqeja, M.C. Do, Q.L. Nguyen, A. Date, A. Akbarzadeh, Experimental investigation of the effect of the spacer and operating conditions on mass transfer in direct contact membrane distillation, *Desalination* 500 (2021), 114839.
- [6] A. Khalifa, H. Ahmad, M. Antar, T. Laoui, M. Khayet, Experimental and theoretical investigations on water desalination using direct contact membrane distillation, *Desalination* 404 (2017) 22–34.
- [7] A. Shafeian, M. Khiadani, A novel solar-driven direct contact membrane-based water desalination system, *Energy Convers. Manage.* 199 (2019), 112055.
- [8] A. Bamasag, T. Alqahtani, S. Sinha, N. Ghaffour, P. Phelan, Experimental investigation of a solar-heated direct contact membrane distillation system using evacuated tube collectors, *Desalination* 487 (2020), 114497.
- [9] A.E. Kabeel, M. Abdelgaiad, E.M.S. El-Said, Study of a solar-driven membrane distillation system: Evaporative cooling effect on performance enhancement, *Renewable Energy* 106 (2017) 192–200.
- [10] I. Hitsov, T. Maere, K. De Sitter, C. Dotremont, I. Nopens, Modelling approaches in membrane distillation: A critical review, *Sep. Purif. Technol.* 142 (2015) 48–64.
- [11] Y.M. Manawi, M. Khrasheh, A.K. Fard, F. Benyahia, S. Adham, Effect of operational parameters on distillate flux in direct contact membrane distillation (DCMD): Comparison between experimental and model predicted performance, *Desalination* 336 (2014) 110–120.
- [12] M. Qtaishat, T. Matsuura, B. Kruczak, M. Khayet, Heat and mass transfer analysis in direct contact membrane distillation, *Desalination* 219 (2008) 272–292.
- [13] R.D. Gustafson, J.R. Murphy, A. Achilli, A stepwise model of direct contact membrane distillation for application to large-scale systems: Experimental results and model predictions, *Desalination* 378 (2016) 14–27.
- [14] A. Shafeian, M. Khiadani, A. Nosrati, Performance analysis of a thermal-driven tubular direct contact membrane distillation system, *Appl. Therm. Eng.* 159 (2019), 113887.
- [15] A.M. Elzahaby, A. Kabeel, M. Bassuoni, A.R. Abd Elbar, Direct contact membrane water distillation assisted with solar energy, *Energy Convers. Manage.* 110 (2016) 397–406.
- [16] A. Ali, C.A. Quist-Jensen, F. Macedonio, E. Drioli, Optimization of module length for continuous direct contact membrane distillation process, *Chem. Eng. Process.* 110 (2016) 188–200.
- [17] A. Ali, J.-H. Tsai, K.-L. Tung, E. Drioli, F. Macedonio, Designing and optimization of continuous direct contact membrane distillation process, *Desalination* 426 (2018) 97–107.
- [18] S. Noamani, S. Niroomand, M. Rastgar, A. McDonald, M. Sadrzadeh, Development of a self-sustained model to predict the performance of direct contact membrane distillation, *Sep. Purif. Technol.* (2021), 118407.
- [19] 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.
- [20] M. Faegh, P. Behnam, M.B. Shafii, M. Khiadani, Development of artificial neural networks for performance prediction of a heat pump assisted humidification-dehumidification desalination system, *Desalination* 508 (2021), 115052.
- [21] S. Mittal, A. Gupta, S. Srivastava, M. Jain, Artificial Neural Network based modeling of the vacuum membrane distillation process: Effects of operating parameters on membrane fouling, *Chemical Engineering and Processing-Process Intensification* 164 (2021), 108403.
- [22] A. Ruiz-Aguirre, J. Andrés-Manas, J. Fernández-Sevilla, G. Zaragoza, Experimental characterization and optimization of multi-channel spiral wound air gap membrane distillation modules for seawater desalination, *Sep. Purif. Technol.* 205 (2018) 212–222.
- [23] D. Cheng, N. Li, H. Bai, J. Zhang, Z. Wang, F. Zeng, J. Sun, Z. Xie, Simulation and multi-objective optimization of heat and mass transfer in direct contact membrane distillation by response surface methodology integrated modeling, *Chem. Eng. Res. Des.* 159 (2020) 565–581.
- [24] S.T. Bouguecha, A. Boubakri, S.E. Aly, M.H. Al-Beiruty, M.M. Hamdi, Optimization of permeate flux produced by solar energy driven membrane distillation process using central composite design approach, *Water Sci. Technol.* 74 (2016) 87–98.
- [25] D. Cheng, W. Gong, N. Li, Response surface modeling and optimization of direct contact membrane distillation for water desalination, *Desalination* 394 (2016) 108–122.
- [26] M. Khayet, C. Cojocaru, C. García-Payo, Application of response surface methodology and experimental design in direct contact membrane distillation, *Ind. Eng. Chem. Res.* 46 (2007) 5673–5685.
- [27] 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.
- [28] M. Khayet, C. Cojocaru, M. Essalhi, Artificial neural network modeling and response surface methodology of desalination by reverse osmosis, *J. Membr. Sci.* 368 (2011) 202–214.
- [29] M. Khayet, C. Cojocaru, Artificial neural network modeling and optimization of desalination by air gap membrane distillation, *Sep. Purif. Technol.* 86 (2012) 171–182.
- [30] 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 Computing and Applications* 28 (2017) 2099–2104.
- [31] M. Khayet, C. Cojocaru, Artificial neural network model for desalination by sweeping gas membrane distillation, *Desalination* 308 (2013) 102–110.
- [32] M. Tavakolmoghadam, M. Safavi, An optimized neural network model of desalination by vacuum membrane distillation using genetic algorithm, *Procedia Engineering* 42 (2012) 106–112.
- [33] 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.
- [34] 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.
- [35] B. Kim, Y. Choi, J. Choi, Y. Shin, S. Lee, Effect of surfactant on wetting due to fouling in membrane distillation membrane: Application of response surface methodology (RSM) and artificial neural networks (ANN), *Korean J. Chem. Eng.* 37 (2020) 1–10.
- [36] R.J. Moffat, Describing the uncertainties in experimental results, *Exp. Therm Fluid Sci.* 1 (1988) 3–17.
- [37] J. Holman, *Experimental Methods For Engineers*, 8th ed ed., McGraw-Hill, 2011.
- [38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: Machine learning in Python, the, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [39] M. Hamdan, H.R. Khalil, E. Abdelhafez, Comparison of neural network models in the estimation of the performance of solar still under jordanian climate, *Journal of Clean Energy Technologies* 1 (2013) 238–242.
- [40] H.K. Ghritlahre, R.K. Prasad, Application of ANN technique to predict the performance of solar collector systems-A review, *Renewable Sustainable Energy Rev.* 84 (2018) 75–88.
- [41] A.H. Elsheikh, S.W. Sharshir, M. Abd Elaziz, A. Kabeel, W. Guilan, Z. Haiou, Modeling of solar energy systems using artificial neural network: A comprehensive review, *Sol. Energy* 180 (2019) 622–639.
- [42] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, *Statistics and Computing* 14 (2004) 199–222.
- [43] J. Yuan, C. Wang, Z. Zhou, Study on refined control and prediction model of district heating station based on support vector machine, *Energy* 189 (2019), 116193.
- [44] M. Sadi, H. Fakharian, H. Ganji, M. Kakavand, Evolving artificial intelligence techniques to model the hydrate-based desalination process of produced water, *Journal of Water Reuse and Desalination* 9 (2019) 372–384.
- [45] M.W. Ahmad, J. Reynolds, Y. Rezgui, Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees, *J. Cleaner Prod.* 203 (2018) 810–821.
- [46] P. Behnam, M. Faegh, M.B. Shafii, M. Khiadani, A comparative study of various machine learning methods for performance prediction of an evaporative condenser, *Int. J. Refrig.* (2021).
- [47] M. Moghadasi, H.A. Ozgoli, F. Farhani, Steam consumption prediction of a gas sweetening process with methyldiethanolamine solvent using machine learning approaches, *Int. J. Energy Res.* 45 (2021) 879–893.
- [48] Y. Sun, F. Haghghat, B.C. Fung, A review of the-state-of-the-art in data-driven approaches for building energy prediction, *Energy Build.* 221 (2020), 110022.
- [49] A.H. Mohammadi, A. Eslamimanesh, F. Gharagheizi, D. Richon, A novel method for evaluation of asphaltene precipitation titration data, *Chem. Eng. Sci.* 78 (2012) 181–185.

- [50] M.H. Ahmadi, A. Baghban, M. Sadeghzadeh, M. Zamen, A. Mosavi, S. Shamshirband, R. Kumar, M. Mohammadi-Khanaposhani, Evaluation of electrical efficiency of photovoltaic thermal solar collector, *Engineering Applications of Computational Fluid Mechanics* 14 (2020) 545–565.
- [51] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, *Energy Build.* 37 (2005) 545–553.
- [52] Ó. Andrjesdóttir, C.L. Ong, M. Nabavi, S. Paredes, A.S.G. Khalil, B. Michel, D. Poulikakos, An experimentally optimized model for heat and mass transfer in direct contact membrane distillation, *Int. J. Heat Mass Transfer* 66 (2013) 855–867.
- [53] M.H. Sharqawy, J.H. Lienhard, S.M. Zubair, Thermophysical properties of seawater: a review of existing correlations and data, *Desalin. Water Treat.* 16 (2010) 354–380.
- [54] G.-P. Li, L.-Z. Zhang, Conjugate heat and mass transfer in a cross-flow hollow fiber membrane bundle used for seawater desalination considering air side turbulence, *J. Membr. Sci.* 533 (2017) 321–335.
- [55] Y. Yun, R. Ma, W. Zhang, A. Fane, J. Li, Direct contact membrane distillation mechanism for high concentration NaCl solutions, *Desalination* 188 (2006) 251–262.
- [56] A. Alkhudhiri, N. Darwish, N. Hilal, Membrane distillation: A comprehensive review, *Desalination* 287 (2012) 2–18.
- [57] S.O. Olatunji, L.M. Camacho, Heat and mass transport in modeling membrane distillation configurations: a review, *Frontiers in Energy Research* 6 (2018) 130.
- [58] R.A. Johnson, M.H. Nguyen, Understanding Membrane Distillation and Osmotic Distillation, Wiley Online Library, 2017.
- [59] M.R. Qtaishat, F. Banat, Desalination by solar powered membrane distillation systems, *Desalination* 308 (2013) 186–197.
- [60] M. Khayet, Membranes and theoretical modeling of membrane distillation: a review, *Adv. Colloid Interface Sci.* 164 (2011) 56–88.