JOURNAL OF WATER AND LAND DEVELOPMENT

e-ISSN 2083-4535

Polish Academy of Sciences (PAN) Institute of Technology and Life Sciences – National Research Institute (ITP – PIB)

JOURNAL OF WATER AND LAND DEVELOPMENT
DOI: 10.24425/jwld.2025.155304
2025, No. 66 (VII–IX): 79–88

Predictive modelling of aerobic membrane bioreactors using machine learning algorithms for the treatment of wastewater from fish canneries

Hafida Ayyoub*[1], Mohammed Kaicer[2], Ismail Elabbassi[3], Mohamed Taky[1]

[1] Ibn Tofail University, Faculty of Sciences, Laboratory of Advanced Materials and Process Engineering, P.O. Box 1246, Kenitra, Morocco

[2] Ibn Tofail University, Faculty of Sciences, Department of Mathematics, Laboratory of Analysis Geometry and Applications, P.O. Box 1246, Kenitra, Morocco

[3] Mohammed V University, National School of Computer Science and Systems Analysis (ENSIAS), Madinat Al Irfane, BP 713, Agdal Rabat, Morocco

* Corresponding author

RECEIVED 13.01.2025 ACCEPTED 10.05.2025 AVAILABLE ONLINE 14.08.2025

Abstract: Process modelling is an effective tool for describing and predicting the performance of an aerobic membrane bioreactor (AeMBR) for fish canning wastewater (FCWW) treatment under different operating conditions. Three machine learning (ML) algorithms were developed, random forest (RF), decision tree regressor (DTR) and adaptive boosting regression (AdaBoost-R), based on various physico-chemical characteristics of the influent and operating conditions, including hydraulic retention times (*HRT*), organic loading rates (*OLR*), total dissolved solids (TDS), aeration rate and permeate volumetric rates. Predicted values for chemical oxygen demand (COD), biochemical oxygen demand (BOD_5), total Kjeldahl nitrogen (TKN), and nitrate (NO_3^-) are compared with those reported from the experiment. As regards the quantitative assessment of the three predictive models, the DTR model demonstrated a modest determination coefficient (R^2) value of 0.654, the AdaBoost-R model achieved 0.739, whereas the RF model showed the highest performance at 0.98. Due to its robustness and accuracy, the RF model was chosen for its superior ability to predict the performance of the AeMBR. Based on *OLR* of 4.27 (kg COD)·(m^3·d)$^{-1}$, a *HRT* of 24 h, a TDS of 3 g·dm^{-3}, an aeration rate of 1,300 Ndm^3·h^{-1} and a permeate volumetric rate of 15 dm^3·h^{-1}, the average effluent characteristics comply with discharge and reuse limits.

Keywords: effluent characteristics, fish canning wastewater treatment, machine learning algorithms optimisation, membrane bioreactor performance, operating conditions

INTRODUCTION

Fish canning wastewater (FCWW) represents a challenge in terms of its management and treatment. Before discharging, such effluents must be effectively treated to reduce the impact on the aquatic environment; alternatively, a way must be found to recover and reuse these effluents (Correa-Galeote *et al.*, 2021; Ayyoub *et al.*, 2022; Ayyoub *et al.*, 2023). In recent years, the use of the aerobic membrane bioreactor (AeMBR) technology has proven to be advantageous as it improves effluent quality and disposal efficiency (Hao *et al.*, 2018; Paul *et al.*, 2023). The AeMBR has many advantages over traditional activated sludge (AS), including high treatment efficiency, reduced pollution, simpler operation, high separation quality, and solids-free effluent (Elmoutez *et al.*, 2024). One of the disadvantages of this system is the high operating costs due to the fact that aeration promotes microbial growth in the wastewater (Al-Asheh, Bagheri and Aidan, 2021; Asante-Sackey *et al.*, 2022). The reduction in

AeMBR operating costs depends on operating parameters that influence their performance (Terna, Ahmed and Joo Hwa, 2016; Checa Fernández *et al.*, 2021).

Efforts to model AeMBR systems for wastewater (WW) treatment have traditionally focused on biological processes (treatment quality objectives) and various engineering aspects (cost-effective design and operation) (Saunders, Drew, and Brink, 2021). These traditional models are based on well-defined rules and theories, such as mass balance equations or reaction kinetics, but they may not provide accurate predictions when crucial factors are missing, or when key coefficients, such as reaction rates or partition constants, are inaccurate or incomplete (Li *et al.*, 2021; Guo and Cui, 2022). By providing advanced analytics, real-time monitoring, predictive capabilities, and optimisation strategies, machine learning (ML) can solve complex problems and support advanced modelling to deliver significant benefits in WW treatment (Sundui *et al.*, 2021; Kumar *et al.*, 2023). These benefits contribute to improved treatment performance, enhanced operational efficiency, and cost savings in the WW treatment process (Kitanou *et al.*, 2021b).

In recent years, AeMBR performance control using various ML algorithms has attracted particular attention, including the development of decision trees (DT), artificial neural networks (ANN), support vector machines (SVM), and random forests (RF) (Schmitt *et al.*, 2018; Mavani *et al.*, 2022). A fundamental principle of ML is to generalise relationships between input and output using inductive reasoning, and then use these generalisations to guide decision making in new contexts (Andrade Cruz *et al.*, 2022; Zhong *et al.*, 2022). The AdaBoost is an ensemble learning technique that works well for complex regression and classification tasks by combining multiple weak learners using weighted summation to improve prediction accuracy (Nguyen and Seidu, 2022). Similarly, decision tree regressor is a non-parametric model that predicts continuous variables by recursive date partitioning. It offers automated feature selection and flexibility in handling non-linear interactions (Qambar and Khalidy, 2022). More specifically, RF is an ensemble ML technique that is used for classification and regression (Umoh *et al.*, 2022). It improves the performance of regression and classification trees by combining multiple decision trees (Vigneau *et al.*, 2018). The RF method uses bootstrapping and randomised variable selection to reduce the correlation between individual trees, thereby reducing the variance of the aggregated trees (Salem *et al.*, 2022). It proves valuable in assessing the impact of operating parameters on AeMBR performance (Chang *et al.*, 2022). By utilising an ensemble of decision trees, RF effectively predicts outcomes and identifies significant variables (Kovacs *et al.*, 2022). Operational factors in the AeMBR can have a significant impact on the effectiveness of the treatment system in terms of contaminant removal, membrane fouling, and overall system performance. Using the RF model, the complex relationships and interactions between these operational parameters and performance indicators can be analysed (Chang *et al.*, 2022). The algorithm can handle non-linearities, missing data, and the relative importance of each operational parameter influencing the AeMBR performance (Kamali *et al.*, 2021). The RF model is a ML algorithm that has gained significant attention in artificial intelligence research due to its strong adaptive learning ability and nonlinear mapping capabilities (AlSawaftah *et al.*, 2021).

A review of the literature indicates that there are few modelling studies using data-driven machine learning algorithms to support the treatment of water produced by the AeMBR that consider the impact of operating parameters on the performance of this innovative technology (Reza *et al.*, 2011; Yusuf, Wahab and Sudin, 2019). However, considering that the majority of generated models are trained using only a few operational settings, it is unclear how they would perform in various contexts and whether they can be generalised (Kazemi *et al.*, 2020; Abouzari *et al.*, 2021). Assessing all impact parameters is time-consuming and involves difficult experiments and the use of hazardous materials, as specified in the standard methodology for water and wastewater (WW) assessment (Aghdam *et al.*, 2023). However, several critical parameters, such as BOD_5 and COD, are expensive and difficult to measure using sensors. This necessitates the development of mathematical prediction models to determine their values based on historical data. The ML can model complex nonlinear connections using mathematical or chemical formulas without explicitly specifying the treatment process. Unlike with traditional models, this enables us to explore new perspectives on WW behaviours that are difficult to detect. The integration of ML into WW treatment processes has been successfully used as a flexible computational tool capable of enhancing environmental preservation, optimising plant performance, and improving the treatment process (Arthur *et al.*, 2022; Mekaoussi *et al.*, 2023). Consequently, the present study focused on developing three regression algorithms DT, AdaBoost, and RF to determine the most appropriate model for different operational parameters at the pilot scale to assess the effectiveness of the AeMBR for FCWW treatment. Due to its impressive accuracy and ability to handle complex nonlinear interactions, the RF model was selected as the best method for studying the effect of operational parameters on treatment efficiency. This decision is based on the model's enhanced predictive capabilities, which provide more reliable and accurate information about elements that affect performance, thus offering a solid basis for improving FCWW treatment processes.

MATERIALS AND METHODS

EXPERIMENTAL SETUP

An AeMBR pilot scale system uses an ultrafiltration (UF) membrane fed with industrial FCWW. The layout of the pilot plant (AeMBR) is shown in Figure 1. The main characteristics of the UF membranes and the general operating conditions for the experimental studies are presented in Table 1. The experiment involves two stages. In the first stage, the FCWW undergoes pre-treatment by screening and then primary treatment by sedimentation and flotation lasting 2 h. In the second stage, the FCWW is treated in a bioreactor in contact with AS and filtered through an UF membrane (Kitanou *et al.*, 2021a; Kitanou *et al.*, 2021b). The bioreactor is equipped with anoxic and aeration tanks. The total suspended solids (TSS) concentration in the second stage is maintained at about 10 $g \cdot dm^{-3}$, with periodic sludge removal and sludge retention time of approximately 12 days (Tab. 1).

This study shows that sedimentation and flotation are effective methods for the removal of TSS and oil and grease (O&G), but not effective for the removal of organic matter due to

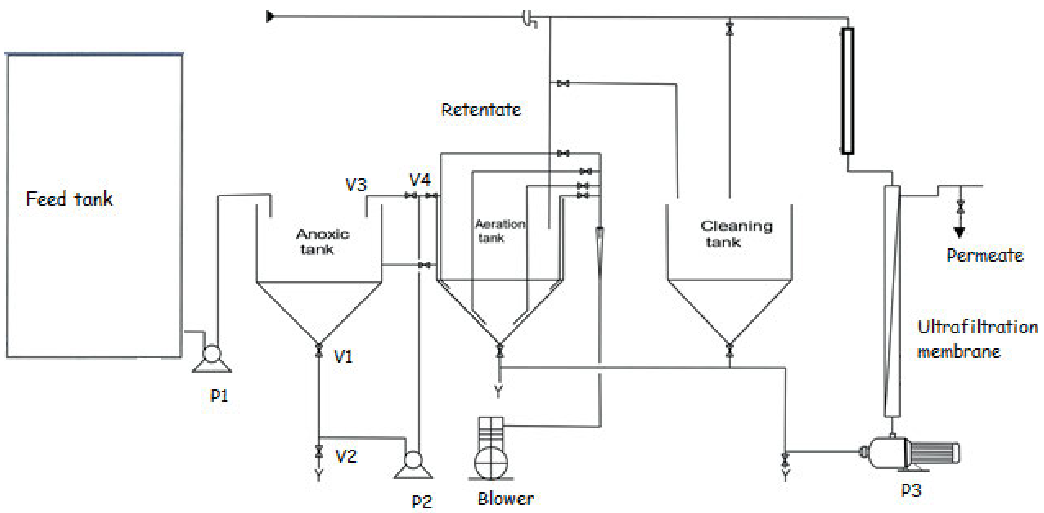


Fig. 1. The layout of the pilot-scale AeMBR; source: Ayyoub *et al.* (2023)

Table 1. Characteristics of the ultrafiltration (UF) membrane and operating conditions of the pilot-scale aerobic membrane bioreactor (AeMBR)

Parameter	Feature/Value
UF membrane	
Membrane material	ceramic
Module	Tubular type P10
Membrane area	0.45 m^2
Cut off	15 kDa / 10–20 nm
Membrane length	1.178 m
Diameter of the channels	6 mm
AeMBR operating conditions	
pH	6.5–8.0
Temperature	21.5 ±4°C
Solids retention time (*SRT*)	12 d
Ratio of feed to microorganisms (F/M)	(0.24 kg BOD_5)·(kg VSS·d)$^{-1}$
Rate of volumetric loading	(2.23 kg BOD_5)·(m^3·d)$^{-1}$
Biomass concentration in the mixed liquor	10 g·dm^3 TSS and 8.5 g·dm^{-3} VSS
Retentate flow rate	124 dm^3·h^{-1}
Generated sludge	1.46 (g TSS)·h^{-1}
Dissolved oxygen	2–4 mg·dm^{-3}
Permeate flux	11.1–51.1 dm^3·(m^2·h)$^{-1}$
Transmembrane pressure	5–150 kPa

Explanations: BOD_5 = biochemical oxygen demand, VSS = volatile suspended solids, TSS = total suspended solids.
Source: Ayyoub *et al.* (2023), modified.

chemical oxygen demand (COD) and biochemical oxygen demand (BOD_5). This requires further treatment using biological and membrane processes. The bioreactor is sequentially aerated; the recovered permeate is collected through a ceramic UF membrane module. The membrane operates at varying flow rates and transmembrane pressures (TMPs), and it is regularly cleaned.

ANALYTICAL METHODS

Standard methods are used to determine quality parameters such as COD (Hach DR2800 spectrophotometer) and BOD_5 (OxiTop WTW) (Eaton *et al.* (eds.), 2005; Rodier, 2009). Total dissolved solids (TDS) are measured using a multi-parameter conductivity meter (inoLab) with an electrode consisting of two platinum strips. An electrode (Sension MM 340) is used to measure the nitrate (NO_3^-) content. The Kjeldahl method is used to calculate total Kjeldahl nitrogen (TKN) (VELP Scientifica, 2013).

OPERATIONAL DATA

In this study, the effect of operating conditions is used to control the performance of the biological process and membrane separation, including *HRT*, *OLR*, aeration rate, TDS, and permeate volumetric rate, that affect the effluent quality of the pilot AeMBR in terms of COD, BOD_5, TKN and NO_3^-. The AeMBR was operated at different *OLR*s (3.0, 4.27 and 5.0 (kg COD)·m^{-3}·d^{-1}), *HRT*s (12, 15, 20 and 24 h), TDS (2.5, 3.0 and 5.0 g·dm^{-3}), aeration rate (700 and 1300 N dm^3·h^{-1}) and permeate volumetric rates (15 and 20 dm^3·h^{-1}). Experience shows that the permeate quality is significantly affected by these ranges of operating conditions.

MACHINE LEARNING MODELS

The three ML algorithms developed include RF, DTR, and AdaBoost-R. Their development required 223 samples from the study on the effect of operating parameters on the AeMBR performance in FCWW treatment. To obtain representative samples, a training dataset was generated using 90% of the original data, while the remaining 10% was used to form the corresponding test dataset.

Decision tree regressor

The DT regression is an iterative binary splitting technique used to predict continuous outcomes. The proposed method performs an intensive search for optimal node splits across features by repeatedly selecting splitting rules to minimise prediction error. This process is repeated until a pre-specified error threshold or minimum sample requirement at the nodes is met (Bishop, 2006).

Adaptive boosting regression

The AdaBoost is an ensemble learning technique that employs adaptive resampling to increase expected performance by correcting errors in the underlying algorithm. The basic principle of the AdaBoost algorithm is to develop models in iterations, with models built in subsequent iterations to correct errors in the previous model (Kumar, Kalita and Ramachandran, 2021). When this procedure reaches a finite state, the final model is created by adding the weights of the underlying models.

Random forest

The RF is a type of algorithm based on ensemble learning that trains several decision trees on random subsets of data and then combines their predictions to obtain a final prediction (Salman, Kalakech and Steiti, 2024). This can enhance model performance over the use of a single decision tree. The process of construction can be summarised as follows: the bootstrap approach is used to create new training data sets from the initial training set. Each new training data set results in the creation of a regression tree. Once trained, each of the regression trees produces a projected value, and the average of these values is used to make the final forecast (Salem *et al.*, 2022).

Verification of the model

The prediction error rates and model performance in regression analysis, which are defined in Equations (1)–(4), are evaluated using the determination coefficient (R^2), mean absolute error (*MAE*), mean squared error (*MSE*), and root mean squared error (*RMSE*):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y}_i| \tag{1}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2 \tag{2}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum (y_i - \widehat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{4}$$

where: y_i = value of the i^{th} sample of the target variable, $\hat{y}$ = predicted value, and $\bar{y}$ = average of collected samples (Zhuang *et al.*, 2021).

RESULTS AND DISCUSSION

EVALUATION OF PREDICTION MODELS

This study aimed to build on recent developments in data-driven machine learning algorithms to predict the performance of a pilot-scale AeMBR for treating FCWW. For this reason, we constructed a three-model optimisation of the operating parameters using the three ML algorithms: DT, AdaBoost, and random forest (RF). This model integrates various physic-chemical characteristics of the influent (CODin, BOD_5in, TKNin and NO_3^-in) and all operating parameters such as *HRT*, *OLR*, aeration rate, TDS and permeate volumetric rate, taking into account the dynamics of the interrelation between these different parameters. Influent characteristics at various stages are shown in Table 2.

Table 2. Characteristics of fish canning wastewater (FCWW) at different hydraulic retention time (*HRT*), organic loading rates (*OLR*), aeration rate, total dissolved solids (TDS) and permeate volumetric rate

***HRT* (h)**	**Average CODin**	**Average BOD_5in**	**Average TKNin**	**Average NO_3^-in**
	mg·dm^{-3}			
12	4,255.87	2,011.06	17.29	119.89
15	4,191.53	2,061.07	19.77	122.25
20	4,308.10	2,353.99	20.46	120.36
24	3,942.01	2,026.14	21.46	123.35

Explanations: BOD_5 = biological oxygen demand; COD = chemical oxygen demand; TKN = total Kjeldahl nitrogen. The organic loading rates (*OLR*) used in the experiments were 3.0, 4.27, and 5.0 (kg COD)·m^{-3}·d^{-1}, and the total dissolved solids (TDS) concentrations were 2.5, 3.0, and 5.0 g·dm^{-3}, respectively. These values are reported where applicable in the table.
Source: own study.

The performance metrics of the predictive models, including DTR and AdaBoost-R, are summarised in Table 3. The table provides a clear overview of each model's R^2, *MSE*, *RMSE* and *MAE*, enabling a direct comparison of their predictive accuracy and error levels. In terms of predictive accuracy, DTR (R^2 = 0.654) performs modestly against AdaBoost (R^2 = 0.739). The correlations between experimental data and values predicted by the final 4 output models using the DT, AdaBoost, and RF algorithms are shown in Figure 2a, 2b, and 2c, respectively.

Table 3. Performance metrics for decision tree regressor (DTR) and adaptive boosting regression (AdaBoost-R)

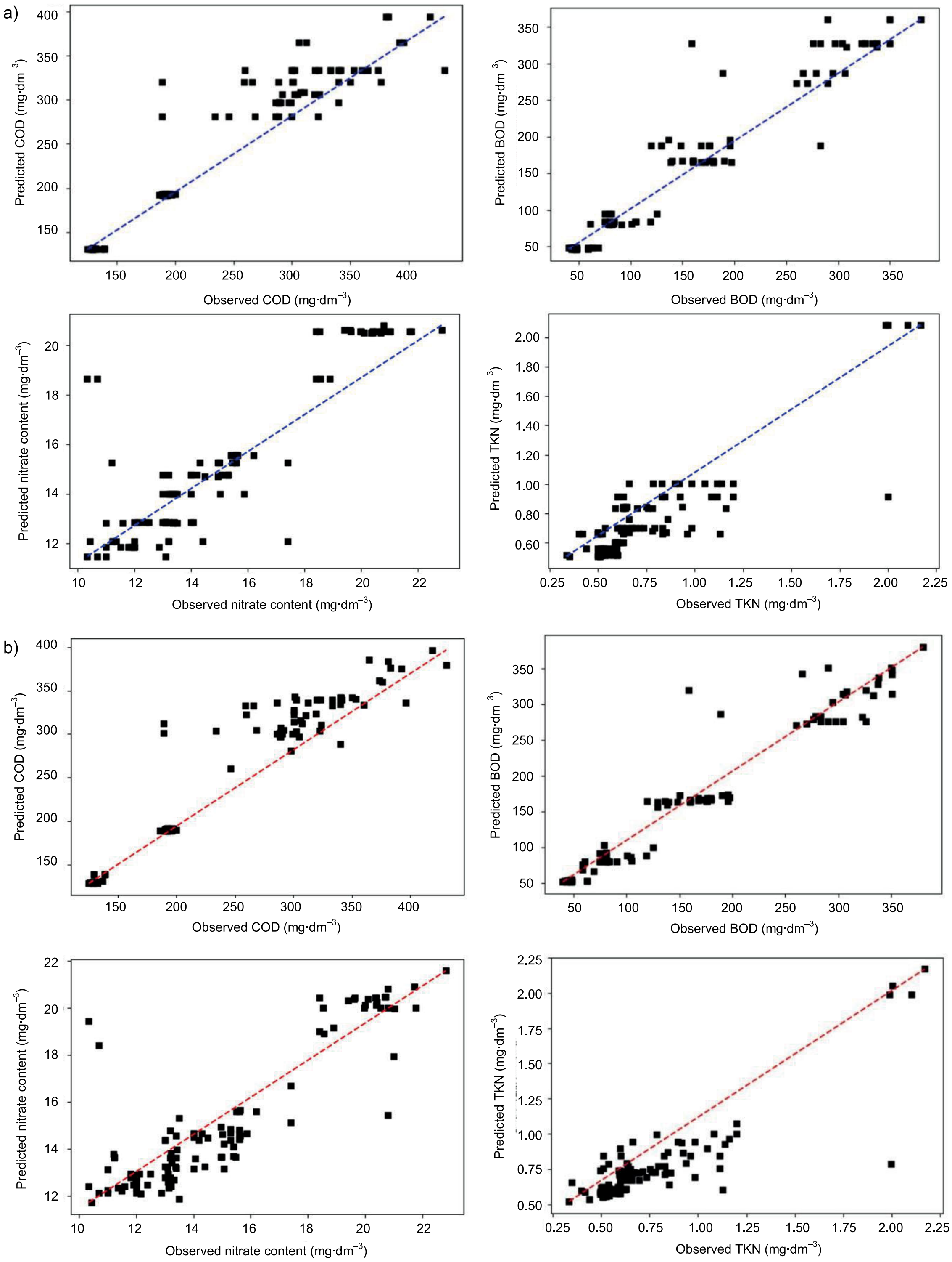
Model	***RMSE***	***MAE***	***MSE***	**R^2**
DTR	0.1315	0.0737	0.0179	0.654
AdaBoost-R	0.1176	0.0719	0.0145	0.739

Explanations: *RMSE* = root mean squared error, *MAE* = mean absolute error, *MSE* = mean squared error, R^2 = determination coefficient.

The scatterplots for the four parameters, such as COD, BOD_5, TKN, and nitrate, show that the predictive capability is adequate and that the linearity between the predicted and actual values is close to 1. A further comparison is made using *RMSE*, *MAE*, *MSE*, and R^2 to provide a quantitative assessment of the RF model's performance. The observed and predicted values gave an R^2 of 0.98, an *MSE* of 0.014, an *MAE* of 1.34, and an *RMSE* of 2.36. A low *RMSE* value means that the simulated and observed data are close to each other, indicating better accuracy (Vigneau *et al.*, 2018). This exceptional accuracy surpasses both DT and AdaBoost regressor, which presented lower R^2 scores and larger errors. The higher performance of the RF model is due to its ensemble nature, which reduces overfitting while efficiently capturing complex nonlinear relationships between the data. With low *MSE*, *RMSE*, and *MAE*, the RF model is the most reliable and stable for predicting the effect of the operational parameters on the performance of the AeMBR in treating FCWW, making it an excellent choice for high-precision applications in environmental monitoring and management.

Overall, our results revealed the importance of certain parametric data. However, we find that most of studies that generated models were trained using only few operational parameters. For example, (Reza *et al.*, 2011) trained the ANN network using 193 operational data including TDS, *OLR*, and *HRT* at the inlet and effluent COD, total organic carbon (TOC), and O&G concentrations at the outlet. According to Aghdam *et al.* (2023), the RF algorithm for BOD_5 estimation showed a better performance level.

Zhong *et al.* (2022b) recently investigated effluent quality modelling in an AeMBR that treated an ammoniacal nitrogen influx with high salt content. They used a variety of methodo-



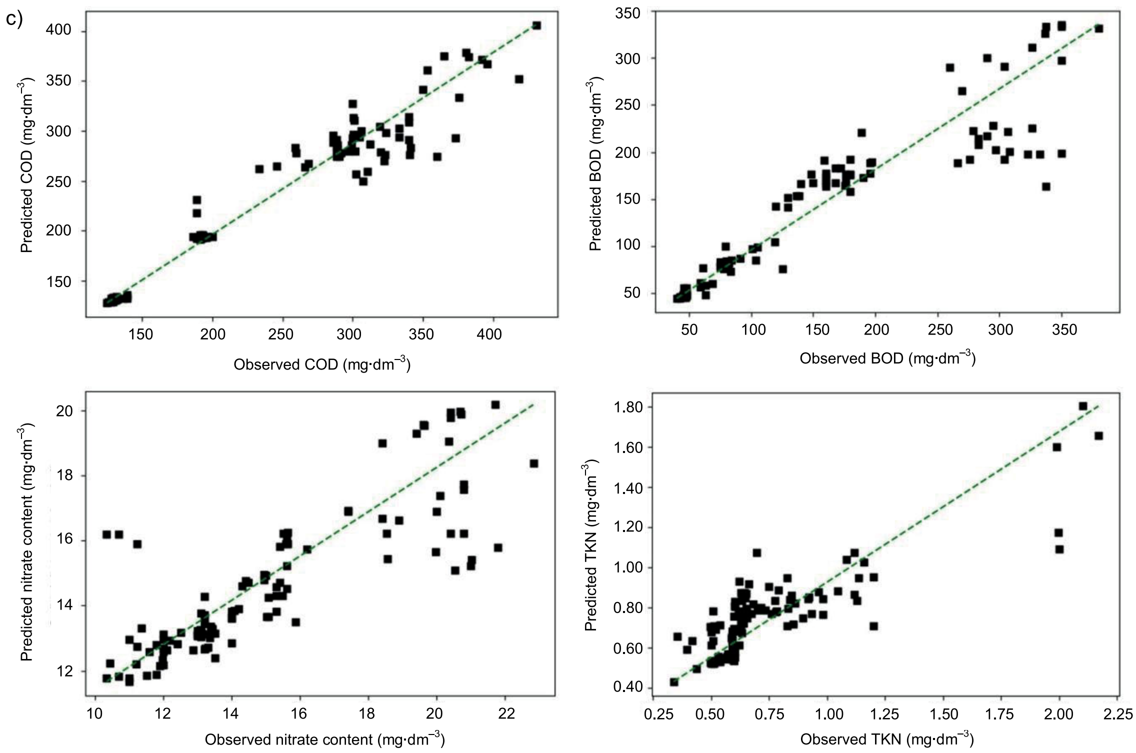


Fig. 2. Combined scatter plot of predicted versus actual values for chemical oxygen demand (COD), biological oxygen demand (BOD_5), total Kjeldahl nitrogen (TKN) and NO_3^- using: a) the decision trees (DT), b) adaptive boosting (AdaBoost), c) random forest (RF); source: own study

logies, including linear regression (LR), regularised linear regression (RR), kernel-peak regression (KRR), polynomial regression (PR), nearest neighbour (KNN), support vector machines (SVM), gradient boosting (GB), and random forest (RF). They have chosen operational input variables such as salinity, dissolved oxygen (DO), *HRT*, pH, temperature, CODin, NH_4^+-Nin, C/N and NH_4^+-Nout with outputs such as NH_4^+-Nout, NO_3^--Nout, NO_2^--Nout, CODout and TNout. The algorithms effectively simulated AeMBR operation in high-salinity wastewater, with RF and GB offering the best results, although RF required the highest computing capacity. The authors stressed the need for a variety of data sets, as well as long-term data, to improve model accuracy. In their study, Li, Li and Wang (2020) used principal component analysis (PCA) by selecting three input parameters, such as mixed liquid suspended solids (MLSS), resistance and pressure, to predict membrane flux. They evaluated RF, backpropagation neural networks and SVMs on the Hadoop massive data platform, and found that the RF-based model had the lowest *RMSE* and the highest accuracy.

UNDERSTANDING PARAMETER EFFECTS WITH RF MODELLING

Python software was used to create a three-dimensional diagram that visualises the interactions between parameter effects and their impact on water quality characteristics in an organic form, such as COD and BOD_5, and in nutrient form, such as NO_3^- and TKN. The impact of parameters and their impact on permeate characteristics is shown in a three-dimensional diagram in Figure 3.

The effects of *OLR*, TDS, and their interrelations with the permeate COD characteristics are presented in Figure 3a. The lowest COD concentration was obtained with the respective lower values of TDS and *OLR* (Fig. 3a). An increase in TDS led to an increase in COD in the permeate, indicating that high TDS levels can negatively impact treatment efficiency (Chandrasekhar *et al.*, 2022). Similarly, as *OLR* levels increase, the concentration of COD in the permeate can also increase, indicating that high *OLR* levels can lead to higher levels of COD in the permeate (Vo *et al.*, 2021). Increasing the *OLR* can enable a more consistent treatment efficiency with active biomass in the membrane module. However, significant membrane fouling is expected as the *OLR* increases (Schmitt *et al.*, 2018; Burman and Sinha, 2020).

Analysis of the effects of *HRT* and *OLR* and their impact on TKN concentrations (Fig. 3b) showed that as *HRT* increased up to 24 h, TKN in the permeate decreased, indicating that a longer *HRT* may promote better nitrogen removal. As *HRT* increases, the TKN concentration in the permeate may also increase, indicating that high *OLR*s may lead to higher TKN levels in the permeate. Both parameters influence effluent quality (Zhu, Huang and Chen, 2022; Rahman *et al.*, 2023).

The influence of *HRT* and aeration rate and their impact on BOD_5 in the permeate is shown in Figure 3c. As *HRT* increases, BOD_5 decreases in the permeate, indicating that longer *HRT*s give biological processes more time to remove organic pollutants. As the aeration rate increases, the BOD_5 concentration in the permeate may decrease, indicating that higher aeration rates can improve organic matter decomposition and reduce BOD_5 levels (Wang *et al.*, 2021). With a 24-h retention time and an aeration rate of 1,300 $Ndm^3 \cdot h^{-1}$, the BOD_5 concentration reached a value

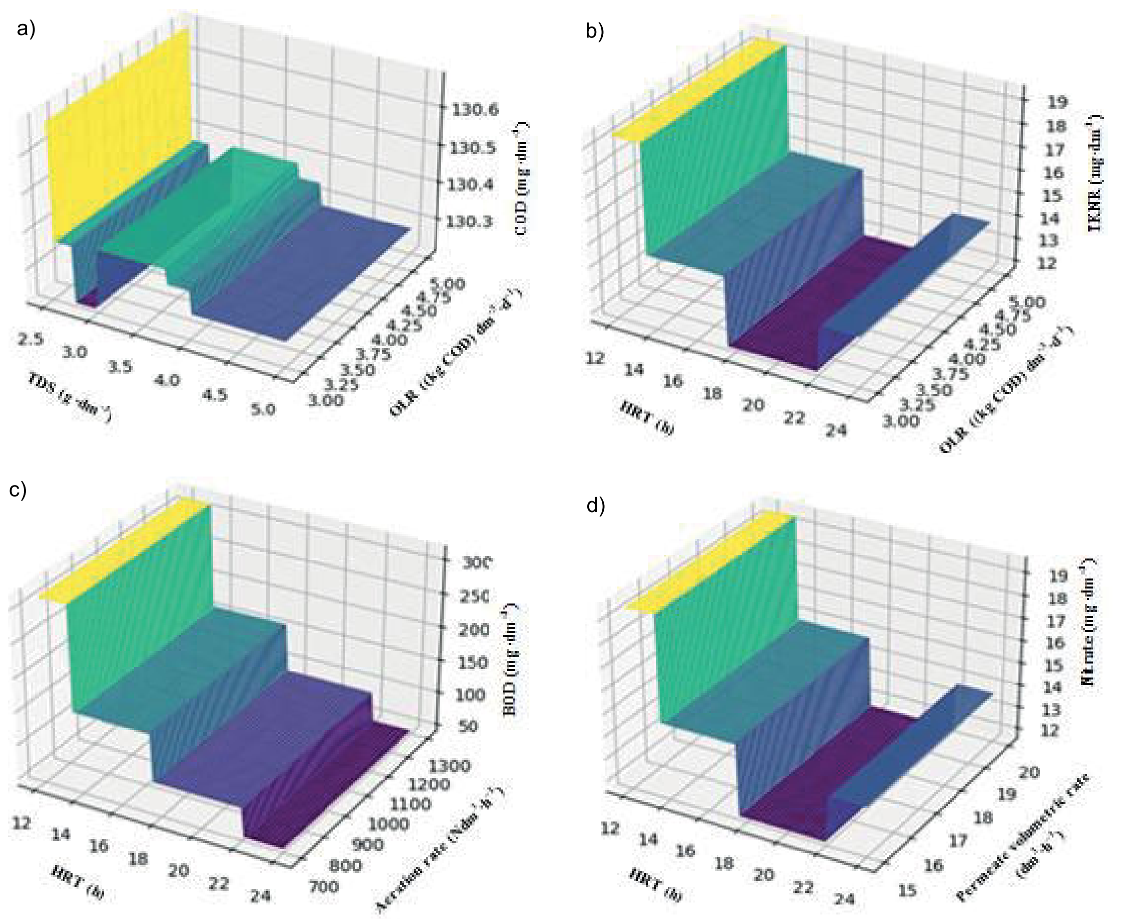


Fig. 3. Influence of parameters and their interactions: a) effects of organic loading rates (*OLR*), total dissolved solids (TDS) and their interrelations on chemical oxygen demands (COD) in the permeate, b) the effects of hydraulic retention times (*HRT*) and *OLR* and their interactions on total Kjeldahl nitrogen (TKN) concentrations, c) influence of *HRT* and aeration rate and their interaction on biochemical oxygen demand (BOD_5) in the permeate, d) influence of *HRT* and permeate volumetric rate on the nitrate concentration; source: own study

of 40 mg·dm^{-3}, indicating a better performance of the AeMBR for treating FCWW.

As the *HRT* increases, the nitrate in the permeate decreases, indicating that longer *HRT* may promote better nitrate removal, potentially through denitrification (Hoover *et al.*, 2016). As the permeate volumetric flow rate increases, the nitrate concentration in the permeate may also decrease, indicating that higher permeate flow rates may enable nitrate concentration to be reduced (Breida *et al.*, 2018). Furthermore, the results of the membrane separation process using different permeate volumetric flow rates (15 and 20 dm^3·h^{-1}) show that the permeate volumetric flow rate of 15 dm^3·h^{-1} gives the best results (Fig. 3d).

OPTIMISATION OF EFFLUENT QUANTITY

Based on the discharge and reuse limits for treated wastewater, organic matter and nutrient standards must comply with these limits. In this study, we took into account various influent characteristics and operating conditions, which are detailed in Table 2. These data formed the basis for the development of the RF algorithm, optimising effluent quality conditions to meet discharge and reuse standards. Some optimum conditions for achieving by AeMBR treatment optimal concentrations of COD, BOD_5, TKN and NO_3^- are shown in Table 4. It can be seen that under operating conditions with a high *HRT* of 24 h, *OLR* of 4.27 (kg COD)·m^{-3}·d^{-1}, TDS of 3 g·dm^{-3}, aeration rate of over 1,300 Ndm^3·h^{-1}, and a permeate volumetric rate of 15 dm^3·h^{-1}, a good correlation between the actual and predicted values was obtained with an R^2 value of 0.98 (values observed during various iterations of RF model varied between 0.96 and 0.99). The mean COD concentrations in the AeMBR permeate was 128 mg·dm^{-3}. These results are very close to the standards for reuse of about 100 mg·dm^{-3}. Furthermore, the mean effluent BOD_5 was 40 mg·dm^{-3}, which is below standards for discharge to the environment (100 mg·dm^{-3}) and close to reuse standards. In line with Moroccan reuse standards, NO_3^- content and TKN concentrations were 12.1 mg·dm^{-3} and 0.34 mg·dm^{-3}, respectively. Furthermore, under these conditions, high removal efficiencies were achieved for 97% of the organic carbon and over 96% of the measured ions (Ayyoub *et al.*, 2022).

In a previous study (Reza *et al.*, 2011), ANN has been used effectively to predict submerged membrane behaviour in the treatment of hypersaline effluents with a range of TDS content (35 to 250 g·dm^{-3}) and *OLRs* (0.281 to 3.372 (kg COD)·m^{-3}·d^{-1}). The network was trained using 193 operational data. The resulting model predicted a COD removal of 98% with an *OLR* of 2.44 (kg COD)·m^{-3}·d^{-1}, TDS of 78 g·dm^{-3}, and a reaction time of 40 h, corresponding to an *HRT* of 80 h.

Table 4. Optimal concentrations under different operating conditions

Inputs					Outputs							
HRT (h)	*OLR* (kg COD) $\cdot m^{-3} \cdot d^{-1}$)	TDS ($g \cdot dm^{-3}$)	aeration rate ($Ndm^3 \cdot h^{-1}$)	permeate volumetric rate ($dm^3 \cdot h^{-1}$)	predicted COD	observed COD	predicted BOD_5	observed BOD_5	predicted TKN	observed TKN	predicted NO_3^-	observed NO_3^-
					$mg \cdot dm^{-3}$							
24	4.27	3.0	1300	15.0	130.46	128	45.63	40	0.53	0.34	13.67	12.10
20	5.0	2.5	1300	20	192.57	192.31	76.96	84.77	0.72	0.73	11.30	12.55
15	5.0	5.0	1300	20	272.36	288.03	171.85	168.19	0.65	0.78	15.36	14.86

Explanations: *HRT* = hydraulic retention time, *OLR* = organic loading rates, TDS = total dissolved solids, COD = chemical oxygen demand, BOD_5 = biological oxygen demand, TKN = total Kjeldahl nitrogen.
Source: own study.

CONCLUSIONS

In this study, the random forest (RF) model was developed to evaluate the performance of aerobic membrane bioreactor (AeMBR) in fish canning wastewater (FCWW) treatment. This model includes hydraulic rate times (*HRT*), organic loading rates (*OLR*), total dissolved solids (TDS), aeration rate, and permeate volumetric rate as inputs, as well as physic-chemical characteristics of untreated water, in terms of chemical oxygen demand (COD), biological oxygen demand (BOD_5), total Kjeldahl nitrogen (TKN), and nitrate content. For a quantitative assessment of the model's performance, the observed and predicted values gave a determination coefficient (R^2) of 0.98, mean squared error (*MSE*) of 0.014, mean absolute error (*MAE*) of 1.34, and a root mean squared error (*RMSE*) of 2.36. According to the model created, the optimum results for *HRT*, *OLR*, and permeate volumetric rate are 24 h, 4.27 (kg COD)$\cdot m^{-3} \cdot d^{-1}$ and 15 $dm^3 \cdot h^{-1}$ respectively. These results clearly show that RF is very useful for assessing the effect of operating parameters on the AeMBR performance.

CONFLICT OF INTERESTS

All authors declare that they have no conflict of interests.

REFERENCES

Abouzari, M. *et al.* (2021) "Chemosphere estimating the chemical oxygen demand of petrochemical wastewater treatment plants using linear and nonlinear statistical models – A case study," *Chemosphere*, 270, 129465. Available at: https://doi.org/10.1016/j.chemosphere.2020.129465.

Aghdam, E. *et al.* (2023) "Predicting quality parameters of wastewater treatment plants using artificial intelligence techniques," *Journal of Cleaner Production*, 405, 137019. Available at: https://doi.org/10.1016/j.jclepro.2023.137019.

Al-Asheh, S., Bagheri, M. and Aidan, A. (2021) "Membrane bioreactor for wastewater treatment: A review," *Case Studies in Chemical and Environmental Engineering*, 4, 100109. Available at: https://doi.org/10.1016/j.cscee.2021.100109.

AlSawaftah, N. *et al.* (2021) "A comprehensive review on membrane fouling: Mathematical modelling, prediction, diagnosis, and mitigation," *Water*, 13(9), 1327. Available at: https://doi.org/10.3390/w13091327.

Andrade Cruz, I. *et al.* (2022) "Bioresource technology application of machine learning in anaerobic digestion: Perspectives and challenges," *Bioresource Technology*, 345, 126433. Available at: https://doi.org/10.1016/j.biortech.2021.126433.

Arthur, P.M.A. *et al.* (2022) "Performance evaluation of a full-scale upflow anaerobic sludge blanket reactor coupled with trickling filters for municipal wastewater treatment in a developing country," *Heliyon*, 8(8), 10129. Available at: https://doi.org/10.1016/j.heliyon.2022.e10129.

Asante-Sackey, D. *et al.* (2022) "Membrane bioreactors for produced water treatment: A mini-review," *Membranes*, 12(3), 275. Available at: https://doi.org/10.3390/membranes12030275.

Ayyoub, H. *et al.* (2022) "Membrane bioreactor (MBR) performance in fish canning industrial wastewater treatment," *Water Practice & Technology*, 17(6), pp. 1358–1368. Available at: https://doi.org/10.2166/wpt.2022.059.

Ayyoub, H. *et al.* (2023) "Aerobic treatment of fish canning wastewater using a pilot-scale external membrane bioreactor," *Results in Engineering*, 17, 101019. Available at: https://doi.org/10.1016/j.rineng.2023.101019.

Bishop, C.M. (2006) *Pattern recoginiton and machine learning, information science and statistics*. New York, NY: Springer Science+Business Media, LLC Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf (Accessed: December 10, 2025).

Breida, M. *et al.* (2018) "Nitrate removal from aqueous solutions by γ-Al2O3 ultrafiltration membranes," *Heliyon*, 4(1), e00498. Available at: https://doi.org/10.1016/j.heliyon.2017.e00498.

Burman, I. and Sinha, A. (2020) "Anaerobic hybrid membrane bioreactor for treatment of synthetic leachate: Impact of organic loading rate and sludge fractions on membrane fouling," *Waste Management*, 108, pp. 41–50. Available at: https://doi.org/10.1016/j.wasman.2020.04.031.

Chandrasekhar, S.S. *et al.* (2022) "Performance assessment of a side-stream membrane bioreactor for the treatment of kitchen wastewater," *Biochemical Engineering Journal*, 180, 108366. Available at: https://doi.org/10.1016/j.bej.2022.108366.

Chang, H.M. *et al.* (2022) "Enhanced understanding of osmotic membrane bioreactors through machine learning modeling of water flux and salinity," *Science of the Total Environment*, 838, 156009. Available at: https://doi.org/10.1016/j.scitotenv.2022.156009.

Checa Fernández, A C. *et al.* (2021) "Toxic / hazardous substances and environmental engineering influence of activated sludge dissolved oxygen concentration on a membrane bioreactor performance with intermittent aeration," *Journal of Environmental

Science and Health, Part A, 56(9), pp. 953-962. Available at: https://doi.org/10.1080/10934529.2021.1944834.

Correa-Galeote, D. *et al.* (2021) "Salinity is the major driver of the global eukaryotic community structure in fish-canning wastewater treatment plants," *Journal of Environmental Management*, 290, 112623. Available at: https://doi.org/10.1016/j.jenvman.2021.112623.

Eaton, A.D. *et al.* (2005) *Standard methods for the examination of water and wastewater*. Washington DC: American Public Health Association.

Elmoutez, S. *et al.* (2024) "Evaluating water quality and fouling propensity in a pilot-scale ceramic membrane bioreactor treating municipal wastewater subjected to increasing salinity levels," *Water Science and Technology*, 89(5), pp. 1297–1311. Available at: https://doi.org/10.2166/wst.2024.063.

Guo, C. and Cui, Y. (2022) "Machine learning exhibited excellent advantages in the performance simulation and prediction of free water surface constructed wetlands," *Journal of Environmental Management*, 309, 114694. Available at: https://doi.org/10.1016/j.jenvman.2022.114694.

Hao, X.D. *et al.* (2018) "A sustainability-based evaluation of membrane bioreactors over conventional activated sludge processes," *Journal of Environmental Chemical Engineering*, 6, pp. 2597–2605. Available at: https://doi.org/10.1016/j.jece.2018.03.050.

Hoover, N.L. *et al.* (2016) "Woodchip denitrification bioreactors: Impact of temperature and hydraulic retention time on nitrate removal," *Journal of Environmental Quality*, 45(3), pp. 803–812. Available at: https://doi.org/10.2134/jeq2015.03.0161.

Kamali, M. *et al.* (2021) "Artificial intelligence as a sustainable tool in wastewater treatment using membrane bioreactors," *Chemical Engineering Journal*, 417, 128070. Available at: https://doi.org/10.1016/j.cej.2020.128070.

Kazemi, P. *et al.* (2020) "Robust data-driven soft sensors for online monitoring of volatile fatty acids in anaerobic digestion processes," *Processes*, 8(1), 67. Available at: https://doi.org/10.3390/pr8010067.

Kitanou, S. *et al.* (2021a) "A comparative examination of MBR and SBR performance for municipal wastewater treatment," 16(2), pp. 582–591. Available at: https://doi.org/10.2166/wpt.2021.016.

Kitanou, S. *et al.* (2021b) "Membrane bioreactor for domestic wastewater treatment: Energetic assessment," *Desalination and Water Treatment*, 240, pp. 55–62. Available at: https://doi.org/10.5004/dwt.2021.27675.

Kovacs, D.J. *et al.* (2022) "Membrane fouling prediction and uncertainty analysis using machine learning: A wastewater treatment plant case study," *Journal of Membrane Science*, 660, 120817. Available at: https://doi.org/10.1016/j.memsci.2022.120817.

Kumar, R. *et al.* (2023) "Science of the total environment artificial intelligence and machine learning tools for high-performance microalgal wastewater treatment and algal biorefinery: A critical review," *Science of the Total Environment*, 876, 162797. Available at: https://doi.org/10.1016/j.scitotenv.2023.162797.

Kumar, V., Kalita, K. and Ramachandran, M. (2021) "A comparative study of linear, random forest and AdaBoost regressions for modeling non-traditional machining," *Processes*, 9(11), 2015. Available at: https://doi.org/10.3390/pr9112015.

Li, L. *et al.* (2021) "Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review," *Chemical Engineering Journal*, 405, 126673. Available at: https://doi.org/10.1016/j.cej.2020.126673.

Li, W., Li, C. and Wang, T. (2020) "Application of machine learning algorithms in MBR simulation under big data platform," *Water Practice and Technology*, 15(4), pp. 1238–1247. Available at: https://doi.org/10.2166/wpt.2020.095.

Mavani, N.R. *et al.* (2022) "Application of artificial intelligence in food industry – A guideline," *Food Engineering Reviews*, 14(1), pp. 134–175. Available at: https://doi.org/10.1007/s12393-021-09290-z.

Mekaoussi, H. *et al.* (2023) "Predicting biochemical oxygen demand in wastewater treatment plant using advance extreme learning machine optimized by Bat algorithm," *Heliyon*, 9(11), e21351. Available at: https://doi.org/10.1016/j.heliyon.2023.e21351.

Nguyen, L.V. and Seidu R. (2022) "Application of regression-based machine learning algorithms," *Water*, 14(24), 3993. Available at: https://doi.org/10.3390/w14243993.

Paul, A. *et al.* (2023) "Membrane bioreactor: A potential stratagem for wastewater treatment," in A.K. Nadda *et al.* (eds.) *Membranes for water treatment and remediation*. Singapore: Springer Nature Singapore, pp. 133–155. Available at: https://doi.org/10.1007/978-981-19-9176-9_6.

Qambar, A.S. and Al Khalidy, M.M. (2022) "Prediction of municipal wastewater biochemical oxygen demand using machine learning techniques: A sustainable approach," *Process Safety and Environmental Protection*, 168, pp. 833–845. Available at: https://doi.org/10.1016/j.psep.2022.10.033.

Rahman, T.U. *et al.* (2023) "The advancement in Membrane Bioreactor (MBR) technology toward sustainable industrial wastewater management," *Membranes*, 13(2), 181. Available at: https://doi.org/10.3390/membranes13020181.

Reza, A. *et al.* (2011) "Modeling of membrane bioreactor treating hypersaline oily wastewater by artificial neural network," *Journal of Hazardous Materials*, 192(2), pp. 568–575. Available at: https://doi.org/10.1016/j.jhazmat.2011.05.052.

Rodier, J. (2009) *L'analyse de l'eau: Eaux naturelles, eaux résiduaires et eaux de mer [Water analysis: Natural water, waste water and sea water]*. 9th edn. Paris: Dunod.

Salem, M. *et al.* (2022) "Random forest modelling and evaluation of the performance of a full-scale subsurface constructed wetland plant in Egypt," *Ain Shams Engineering Journal*, 13(6), 101778. Available at: https://doi.org/10.1016/j.asej.2022.101778.

Salman, H.A., Kalakech, A. and Steiti, A. (2024) "Random forest algorithm overview," *Babylonian Journal of Machine Learning*, 2024, pp. 69–79. Available at: https://doi.org/10.58496/BJML/2024/007.

Saunders, A., Drew, D.M. and Brink, W. (2021) "Machine learning models perform better than traditional empirical models for stomatal conductance when applied to multiple tree species across different forest biomes," *Trees, Forests and People*, 6, 100139. Available at: https://doi.org/10.1016/j.tfp.2021.100139.

Schmitt, F. *et al.* (2018) "Development of artificial neural networks to predict membrane fouling in an anoxic-aerobic membrane bioreactor treating domestic wastewater," *Biochemical Engineering Journal*, 133, pp. 47–58. Available at: https://doi.org/10.1016/j.bej.2018.02.001.

Sundui, B. *et al.* (2021) "Applications of machine learning algorithms for biological wastewater treatment: Updates and perspectives," *Clean Technologies and Environmental Policy*, 23(1), pp. 127–143. Available at: https://doi.org/10.1007/s10098-020-01993-x.

Terna, I.O., Ahmed, H.R. and Joo Hwa, T. (2016) "Membrane bioreactor (MBR) technology for wastewater treatment and reclamation: Membrane fouling," *Membranes*, 6(2), pp. 13–16. Available at: https://doi.org/10.3390/membranes6020033.

Umoh, U.A. *et al.* (2022) "Chapter 14 – Fuzzy-machine learning models for the prediction of fire outbreaks: A comparative analysis," in R. Pandey *et al.* (eds.) *Artificial intelligence and machine learning for EDGE computing*. Academic Press, pp. 207–233. Available at: https://doi.org/10.1016/B978-0-12-824054-0.00025-3.

VELP Scientifica (2013) *Kjeldahl method*. Available at: https://www.velp.com/public/file/valuable-information-on-kjeldahl-method-200857-241909.pdf (Accessed: December 10, 2024).

Vigneau, E. *et al.* (2018) "Random forests: A machine learning methodology to highlight the volatile organic compounds involved in olfactory perception," *Food Quality and Preference*, 68, pp. 135–145. Available at: https://doi.org/10.1016/j.foodqual.2018.02.008.

Vo, T.-D.-H. *et al.* (2021) "Influence of organic loading rates on treatment performance of membrane bioreactor treating tannery wastewater," *Environmental Technology & Innovation*, 24, 101810. Available at: https://doi.org/10.1016/j.eti.2021.101810.

Wang, Y. *et al.* (2021) "Aeration rate improves the compost quality of food waste and promotes the decomposition of toxic materials in leachate by changing the bacterial community," *Bioresource Technology*, 340, 125716. Available at: https://doi.org/10.1016/j.biortech.2021.125716.

Yusuf, Z., Wahab, N.A. and Sudin, S. (2019) "Soft computing techniques in modelling of membrane filtration system: A review," *Desalination and Water Treatment*, 161, pp. 144–155. Available at: https://doi.org/10.5004/dwt.2019.24294.

Zhong, H. *et al.* (2022a) "Water quality prediction of MBR based on machine learning: A novel dataset contribution analysis method," *Journal of Water Process Engineering*, 50, 103296. Available at: https://doi.org/10.1016/j.jwpe.2022.103296.

Zhong, H. *et al.* (2022b) "Water quality prediction of MBR based on machine learning: A novel dataset contribution analysis method," *Journal of Water Process Engineering*, 50, 103296. Available at: https://doi.org/10.1016/j.jwpe.2022.103296.

Zhu, C., Huang, H. and Chen, Y. (2022) "Recent advances in biological removal of nitroaromatics from wastewater," *Environmental Pollution*, 307, 119570. Available at: https://doi.org/10.1016/j.envpol.2022.119570.

Zhuang, L. *et al.* (2021) "Performance prediction of an internal-circulation membrane bioreactor based on models comparison and data features analysis," *Biochemical Engineering Journal*, 166, 107850. Available at: https://doi.org/10.1016/j.bej.2020.107850.