Review

Machine Learning Implementation in Membrane Bioreactor Systems: Progress, Challenges, and Future Perspectives: A Review

Zacharias Frontistis [1,*] **, Grigoris Lykogiannis** [2] **and Anastasios Sarmpanis** [2]

[1] Department of Chemical Engineering, University of Western Macedonia, 50132 Kozani, Greece
[2] ECOTECH LTD., 17122 Athens, Greece
*   Correspondence: zfrontistis@uowm.gr; Tel.: +30-24-6105-6659

Abstract: This study offers a review of machine learning (ML) applications in membrane bioreactor (MBR) systems, an emerging technology in advanced wastewater treatment. The review focuses on implementing ML algorithms to enhance the prediction of membrane fouling, control and optimize the system, and predict faults early, thereby enabling the development of novel cleaning strategies. Key ML algorithms such as artificial neural networks (ANNs), support vector machines (SVMs), random forest, and reinforcement learning (RL) are briefly introduced, with an emphasis on their potential and limitations in advanced wastewater applications. The main challenges obstructing the implementation, namely data quality, interpretability, and transferability of ML, are identified. Finally, future research trends are proposed, including ML integration with big data, the Internet of Things (IoT), and hybrid model development. The review also underscores the need for interdisciplinary collaboration and investment in data management, along with the implementation of new policies addressing data privacy and security. By addressing these challenges, the integration of ML into MBRs has the potential to significantly enhance performance and reduce the energy footprint, providing a sustainable solution for advanced wastewater treatment.

1. Introduction

Since ancient times, the management and treatment of water and wastewater has been recognized as an important factor towards the sustainability of water resources and environmental protection. Taking a step beyond conventional wastewater treatment technologies such as activated sludge, membrane bioreactors (MBRs) have already demonstrated their superiority for wastewater treatment, derived from the combination of enhanced biological activity with membrane separation [1–6]. MBR systems are characterized by numerous benefits, such as low sludge production and enhanced efficiency, reduced footprint, and the production of high-quality effluent, facilitating water reuse [1,2,7–10]. Nevertheless, despite these advantages, some significant challenges related to MBR technology remain, including membrane fouling, increased energy consumption, and process control, prevailing the broader implementation of MBR systems [2,5,8–10].

At the same time, artificial intelligence has rapidly developed. In particular, machine learning (ML) has piqued the interest of several researchers towards the application of these advanced technologies to optimize and control various aspects of conventional or novel wastewater treatment processes [11–14]. As with several industrial processes, ML could enhance MBR efficiency using data-driven approaches for modeling and predicting the behavior of similar systems [15–18]. Therefore, ML could reduce operational costs by optimizing efficiency and controlling the MBR system, while simultaneously minimizing the environmental footprint [14,15]. This critical review aims to provide an overview

of the current state-of-the-art of ML applications in MBR wastewater treatment. It will discuss the main ML techniques, their possible MBR implementation and optimization, the main challenges and limitations, and the future perspectives and opportunities in this interesting field.

2. Principles of Membrane Bioreactors

The superior efficiency of MBRs is derived from their hybrid operation, which combines the biological decomposition of pollutants with the physical separation provided by the membrane. Below is an overview of the fundamental MBR principles and their configuration, including the challenges related to their operation.

2.1. MBR Configurations and Components

The most common MBR classification is into two distinct configurations: submerged and side-stream (sequential). Submerged systems are those where the membranes are immersed in the bioreactor. Therefore, the membrane directly filtrates the mixed-liquor suspended solids (MLSS) that exist inside the bioreactor [3–6]. On the other hand, in side-stream MBR, the filtration system is external; thus, the MLSS are circulated through the membrane outside of the bioreactor. Most applications involve the use of submerged systems due to their simpler design and maintenance and lower energy requirements [5,7]. The most important MBR components are the membrane module, the bioreactor, the system providing aeration, and the system for monitoring and control [2,3,6]. Inside the bioreactor, the microorganisms necessary for the biological decomposition grow due to the suitable environment provided by the aeration system and the existing food—organic matter and nutrients [2,7]. In addition to providing oxygen for aerobic microorganisms, the aeration system helps minimize the fouling of the membrane through the promotion of particle dispersion and the cross-flow condition [2–6]. One or more membrane filtration units can form a membrane module. Membranes can be composed of various materials, such as polymeric (e.g., polyvinylidene fluoride) or ceramic, with different pore sizes, ranging from microfiltration to ultrafiltration [3–8]. The monitoring–control system measures the main performance indicators, including transmembrane pressure (TMP), dissolved oxygen, pH, temperature, and MLSS, and enables the MBR automation and optimization [14,15].

2.2. Primary Challenges

Although MBRs present significant advantages and superior efficiency, some operational problems and limitations hinder their widespread use and industrial application [3]. The primary challenge is membrane fouling. Fouling occurs from the deposition and accumulation of particles, colloids, and soluble organic matter on the membrane surface and within the membrane pores [9]. Due to fouling, filtration efficiency declines over time, while the transmembrane pressure significantly increases. Consequently, the system needs more frequent membrane cleaning or replacement, thus reducing the performance and increasing the cost [9]. As with most aerobic processes, energy consumption is a significant drawback in the MBR process due to the high energy demands associated with the aeration systems to provide the necessary oxygen to maintain optimal levels inside the reactor and to prevent fouling [5–9]. Aeration optimization and reducing the MBR energy footprint are crucial to ensuring the feasibility of the MBR technology in the competitive sector of water treatment technologies. In addition, as in many industrial processes, control and optimization are vital to achieving stable and optimal performance. Monitoring and control include the measurement and adjustment of several process parameters, including the hydraulic retention time, sludge retention time, and nutrient dosing, to maintain optimal conditions for the removal of organic matter, simultaneously allowing for the accommodation of possible process disturbances and fluctuations in influent quality, while keeping TMP low to prevent fouling and increase the system's running time [8,13,16,17].

2.3. Key Performance Indicators in MBR Processes

The evaluation of MBR performance is typically based on several parameters, indicating the efficiency of pollutant removal, the membrane fouling rate, energy consumption, and the production of sludge [1–3,8,9]. These indicators can quantify significant results regarding efficiency and effectiveness, allowing for system control and optimization.

(i)   The efficiency of pollutant removal: The primary objective of the MBR system is the removal of organic matter, nutrients, and suspended solids. Efficiency is usually estimated by measuring the chemical oxygen demand (COD), biochemical oxygen demand (BOD), total suspended solids (TSS), nitrogen, and phosphorus in the influent and effluent. High efficiency indicates an effective operation of both biological decomposition and physical separation, resulting in high-quality effluent suitable for water reuse or discharge to water bodies.

(ii)   Membrane fouling rate: A critical indicator is the rate of membrane fouling, since it directly impacts the membrane lifetime, efficiency, and the cleaning procedure chosen. The fouling rate can be estimated by monitoring the increase in transmembrane pressure (TMP) or the decrease in permeate flux over time [3]. A low fouling rate indicates a stable MBR system with a reduced cleaning frequency. Since MBRs are dynamical systems where the occurrence of unpredicted events of TMP due to fouling is difficult to predict, the use of ML to predict membrane fouling is a very promising strategy.

(iii)   Energy consumption: Energy consumption is also a key indicator as it directly affects the viability and environmental footprint of the process. Energy requirements are primarily affected by the aeration system, pumping, and membrane fouling. Different strategies have been proposed to reduce the footprint, including optimization of the aeration rate, the use of more efficient equipment, and the use of more complex and efficient control algorithms to adjust the system parameters according to the observed conditions and the quality of the effluent and influent.

(iv)   Sludge production: As in most biological processes, sludge production is a drawback of the biological activity of the microorganisms, and sludge management remains an economic and environmental challenge [2]. In MBRs, sludge production can be quantified through the measurement of MLSS and the estimation of sludge removal. The goal is to achieve lower production rates, as they reduce the associated costs with sludge handling, dewatering, and disposal. MBRs are characterized by longer sludge retention times, and thus, they produce less sludge with higher biomass concentrations than conventional activated sludge systems.

In summary, MBR systems offer significant advantages for wastewater management, such as high efficiency and reduced sludge production in a compact design. Nevertheless, some operational challenges remain, including membrane fouling, high energy consumption, and process control. Through the monitoring and optimization of key indicators such as the pollutant removal efficiency, membrane fouling rate, energy consumption, and sludge production, MBR systems can achieve optimal performance and overcome operational problems.

3. Fundamentals of Machine Learning

Machine learning (ML) is a subset of artificial intelligence that has already demonstrated superior performance in several different applications, including systems control and optimization, even in complex systems such as those described in MBR systems [14–18]. This section provides an overview of ML fundamentals, including the main techniques used, algorithms, and learning paradigms. The key concepts related to selection, reduction, and model evaluation are presented.

3.1. Machine Learning Techniques and Algorithms

One common classification of ML techniques divides them into three main categories: (i) unsupervised learning, (ii) reinforcement learning, and (iii) supervised learning. Each category consists of several algorithms tailored to different problem types and data characteristics [19–24].

Reinforcement learning algorithms use trial and error procedures through interaction with an environment to learn the optimal action or tasks. These algorithms are often used in the control of dynamic systems and optimization. Indicative reinforcement learning algorithms include Q-learning, deep Q-networks (DQNs), and proximal policy optimization (PPO). On the other hand, unsupervised learning algorithms rely on the discovery of patterns or structures and the estimation of relationships within unlabeled data. These algorithms are often used for tasks such as clustering and dimensionality reduction. Some well-known unsupervised learning algorithms are K-means clustering, hierarchical clustering, Gaussian mixture models, and principal component analysis (PCA).

Supervised learning algorithms impose the learning mapping between the input and the output labels based on a provided set of training examples. Supervised learning algorithms can be divided into: (a) regression (predicting continuous outputs) and (b) classification (predicting discrete outputs). Common supervised learning algorithms include, among others, regression (linear, logistic), support vector machines (SVMs), k-nearest neighbors (KNN), random forest decision trees, and artificial neural networks (ANNs).

Table 1 includes a concise description of the ML algorithms used in MBR for wastewater treatment.

Table 1. ML algorithms used in MBR wastewater treatment [19–24].

Algorithm	Description
Artificial neural network (ANN)	Artificial neural network (ANN) is a popular algorithm for predicting, optimizing, and controlling MBRs. ANNs can analyze complex datasets and identify patterns and relationships, thus enabling effective decision-making and control strategies.
Support vector machines (SVMs)	Support vector machines (SVMs) are mainly used in MBR systems for classification, regression, and future prediction. SVM models can identify and map nonlinear relationships between variables, enhancing the accuracy and efficiency of MBR control and optimization.
Random forest (RF)	Random forest (RF) employs decision trees to improve MBR systems' accuracy. RF models can handle complex and large datasets and identify and quantify relationships between system input and output variables, leading to precise and effective control strategies.
Adaptive network-based fuzzy inference system (ANFIS)	Adaptive network-based fuzzy inference system (ANFIS) is a hybrid ML algorithm that integrates fuzzy logic and neural networks. It is characterized by enhanced prediction and control of MBRs. ANFIS models can capture numerical and linguistic information, thus facilitating effective decision-making and control.
Support vector regression (SVR)	Support vector regression (SVR) is a machine learning algorithm used in MBR systems for regression analysis and prediction. SVR models can identify and map nonlinear relationships between different variables, thereby improving the accuracy and effectiveness of MBR control and optimization.
Partial least squares regression (PLSR)	PLSR is an algorithm used in MBR systems combining principal component analysis and multiple regression. PLSR can deal with multivariate data that are collinear and reduce the dimensionality of the data, leading to more accurate and effective MBR optimization and control.
Deep learning (DL)	Deep learning (DL) is a subfield of machine learning characterized by using ANNs with multiple layers for improved accuracy and effectiveness. DL models can analyze large and complex datasets and identify patterns and relationships between parameters, thereby enabling precise and adaptive control strategies.

Partial least squares regression (PLSR) is a multivariate method used for both classification and regression. The PLSR algorithm searches for the linear combination of input parameters that correlates most strongly with the target variable. First, PLSR maximizes the covariance between the input and the target variables, creating new orthogonal latent variables. Then, the algorithm performs regression between the target and these latent variables, resulting in a linear model that predicts the output from the inputs. PLSR is effective as it can handle multicollinearity (correlated inputs) and reduce data dimensionality using latent variables. Additionally, it can handle noisy data by creating a new set of latent variables.

Support vector machine (SVM) is a common supervised algorithm used in both classification and regression tasks. The algorithm works by finding the optimal hyperplane that separates data of different classes in a high-dimensional space, with the aim of maximizing the margin between the closest points of different classes. The SVM uses a kernel function to transform the data into a high-dimensional space, and then optimizes the separation of data points from different classes by choosing a hyperplane that maximizes the margin. The data points closest to the hyperplane are known as 'support vectors', as they determine the position of the hyperplane. SVMs are effective because they can learn complex relationships between inputs and outputs, and their reliance on support vectors to define the hyperplane can reduce overfitting. They are also capable of ignoring noisy data and outliers.

Random forest is an algorithm that consists of many decision trees and can be used for both classification and regression tasks. It creates a large number of decision trees, each trained on a different subset of data, and combines all the predictions to produce the final result. More specifically, the random forest algorithm begins by choosing a random subset of the training data and using it to train a decision tree. This process is repeated several times, with the result being a 'forest' of decision trees. The final prediction is based on majority voting from individual trees; in classification tasks, the most frequent result predicted by majority of trees is chosen, while for regression, the average of the individual predictions is used. Random forest is a highly effective algorithm, capable of understanding complex relationships between inputs and outputs. It also avoids overfitting by training on different data subsets, thereby preventing over-specialization. Moreover, it can effectively handle noisy data and outliers. However, random forest algorithms can have high computational complexity and can lead to difficulties in interpreting the results.

Artificial neural network (ANN) is perhaps the most well-known machine learning algorithm, inspired by the functioning of the human brain. ANNs consist of interconnected nodes (neurons) arranged in layers, each performing a specific function. The training of an ANN is based on adjusting the weights and biases of the neurons comprising the ANN until the network can accurately predict or classify the training data. Backpropagation is the most commonly used algorithm to estimate the ANN's error (the difference between the simulated and actual values) and update the neurons' biases and weights to minimize the estimated error. Similar to other machine learning techniques, ANNs are quite effective at learning complex relationships between inputs and outputs, with different neurons learning different aspects of these relationships. ANNs are also versatile, capable of handling noisy data and ignoring outliers.

Deep learning is a more complex version of ANN, which typically has 1–2 layers, and is based on multiple layers of neurons. This complexity allows deep learning algorithms to learn and simulate more intricate correlations between input parameters and target values. However, this increased capability comes with the need for a large amount of training data and greater computational power compared to conventional ANNs, making deep learning the go-to choice for applications that require high accuracy and have abundant data available.

ANFIS, or the adaptive network-based fuzzy inference system, is a hybrid algorithm that combines artificial neural networks (ANNs) and fuzzy logic, a well-known method used to handle data uncertainty and imprecision. ANFIS algorithms are suitable for data that vary over time as they are self-learning and can thus be updated with new data. The

process begins with the definition of fuzzy sets for the input and target variables, followed by the establishment of fuzzy rules (i.e., the correlation between inputs and outputs). Finally, the trained model is used to make predictions by applying these fuzzy rules to the input variables. The ANFIS algorithm is robust and accurate, capable of achieving precision and filtering out noise and outliers. In contrast to other techniques, the results obtained through ANFIS are also interpretable.

For different ML applications, preprocessing is needed. Among the different techniques, feature selection and dimensionality reduction are well-established since they can significantly impact model performance and interpretability. Feature selection is based on the selection of a subset of the most relevant features from the original dataset, and then eliminates irrelevant features that introduce noise or overfitting. Dimensionality reduction tries to transform the original high-dimensional data onto a lower-dimensional space, while preserving the essential structure and relationships between data points. It is worth noting that feature selection and dimensionality reduction are often incorporated in the first step of several algorithms. Some standard dimensionality reduction techniques include PCA, linear discriminant analysis (LDA), and t-distributed stochastic neighbor embedding (t-SNE).

3.2. Model Evaluation and Validation

As in most models, the evaluation and validation of ML model performance is vital to ensure the generalization and reliability needed in real-world applications. For model evaluations, usually, the dataset is randomly divided into training, validation, and test sets [20,22–24]. The training set is used to fit the model, the validation set is used to tune the hyperparameters, and the test set is used to assess the model's performance on unseen data. Performance is evaluated using metrics such as the mean squared error (MSE) for regression tasks and accuracy, precision, recall, and the F score for classification tasks. In addition, cross-validation is used, a technique where the dataset is repeatedly partitioned into different training and test subsets to overcome the common problem of overfitting in systems with high degrees of freedom [22–24].

3.3. Challenges in Applying Machine Learning to MBR Systems

Although ML can potentially improve the system performance and control, some challenges must still be solved before the wide implementation of these systems in MBRs. Among other challenges, these include: (i) data quality and availability, (ii) model interpretability, and (iii) adaptability to changes in process conditions [10,12–14].

(i)   Data quality and availability: As in most artificial intelligence approaches, the presence of high-quality data is crucial for the accuracy and reliability of the produced ML model. Unfortunately, in the specific case of MBR systems, the quality of data can be significantly affected by issues such as sensor noise, missing values, and biases in data collection. Some well-known techniques, such as preprocessing, data cleaning, and normalization, can improve the data quality to some degree.

(ii)   Model interpretability: Although ML models can yield very accurate predictions due to their complex nature, the interpretability of the results and the correlation with the physical systems remain problematic and can limit their adoption by practitioners. The implementation of explainable AI techniques (such as local interpretable model-agnostic explanation (LIME), for example) could bridge the gap between the predictions and the human observer.

(iii) Adaptability to changes in process conditions: As with most systems of environmental engineering and wastewater treatment, MBRs could show significant disturbances and fluctuations in the influent quality. Deviations from the steady state can affect the process efficiency and the quality of the effluent. With this in mind, the ML models must be designed to be able to adapt to these changes and provide reliable simulation results and predictions under a wide range of conditions. Different technologies,

such as online learning, can be used to enhance the adaptation of the models in MBR applications.

In a nutshell, machine learning can revolutionize the control and optimization of MBRs using data-driven approaches and advanced algorithms. ML can be implemented into MBR, leading to enhanced efficiency and stability by addressing challenges related to low data quality, model interpretability, and adaptability.

3.4. Applications of Machine Learning in Membrane Bioreactor Systems

Various studies have demonstrated the potential of machine learning in the simulation, optimization, and control of MBRs. Existing applications encompass the most critical parts of the system, including the prediction of membrane fouling, the optimization of operating parameters, automatic control, and early fault detection.

As previously stated, membrane fouling is the main disadvantage associated with using physical processes such as separation. Increased fouling enhances energy consumption, decreases permeability, and reduces the lifespan of membrane modules [25–28]. Consequently, several studies have implemented ML algorithms to predict membrane fouling, aiming to design better maintenance and optimize cleaning strategies. A common approach involves the use of supervised learning to predict fouling indicators (transmembrane pressure (TMP) or fouling rate) in relation to input parameters and wastewater properties [29–31]. Common supervised algorithms such as artificial neural networks (ANNs), support vector machines (SVMs), and random forest (RF) have shown superior efficiency and generalization in predicting the transmembrane pressure [30–35].

Jiang et al. [36] used two partial least squares (PLSR) models to simulate membrane fouling in a novel rotating tubular membrane bioreactor (RTMBR). The chosen inputs included the rotary speed (RS), aeration rate (AR), mixed-liquor suspended solids (MLSS), bound extracellular polymeric substances (bEPS), and mean particle size (MPS). According to the authors, the evaluated models demonstrated satisfactory accuracy ($R^2 = 0.71$), with MLSS emerging as the most critical factor, followed by bEPS, RS, MPS, and AR. Based on these observations, the authors conducted an energy analysis and concluded that increasing the rotary speed is a more effective method to mitigate membrane fouling compared to enhancing the aeration rate.

Kulesha et al. [37] examined the application of PLSR in predicting the fouling of a biofilm-membrane bioreactor system treating municipal wastewaters. The authors utilized input parameters such as mixed-liquor suspended solids (MLSS), the diluted sludge volume index (DSVI), chemical oxygen demand (COD), and sludge relative hydrophobicity (RH), including their slopes. In contrast, the transmembrane pressure, permeability, and their respective slopes were the output parameters. The researchers partitioned the observation period of 114 days into 3 different regions and established 3 distinct models contingent on the operating conditions. These models successfully predicted the system fouling intensity. Cross-validation conducted by the researchers demonstrated low uncertainty, and the models were subsequently applied to adjust the operating variables according to the measured biomass.

In a study conducted by Dalmau et al. [29], the researchers utilized data from a pilot plant, the BR UNIT, operated for 462 days, to compare 2 different strategies for the prediction of TMP: (i) a conventional deterministic model based on the activated sludge model, ASM2d, frequently used for the modeling of biological treatment plants in conjunction with a model for the filtration stage, and (ii) a multivariable regression data-driven model. According to the authors, each model predicted better under different conditions. The data-driven model showed superior performance for higher pH changes and low pH (<7). The authors proposed the combination of the two approaches to create a unified model capable of predicting the TMP across a wide range of operating conditions.

Recently, Zhong et al. [30] investigated the modeling of water quality effluent from a membrane bioreactor used to treat high-salt ammonia nitrogen influent using various methodologies, such as: linear regression (LR), regularized linear regression (RR), kernel

ridge regression (KRR), polynomial regression (PR), k-nearest neighbor (KNN), support vector machine (SVR), gradient boosting (GB), and random forest (RF). NH_4^+-N_{out}, NO_3^--N_{out}, NO_2^--N_{out}, COD_{out}, and TN_{out} were selected as the outputs, while salinity, DO, HRT, pH, water temperature, CODin, NH_4^+-N_{in}, C/N, and NH_4^+-N_{out} were chosen as the input parameters after an initial screening. According to the results, the examined algorithms were able to simulate the use of an MBR in high-salinity wastewater. The integrated learning algorithms (RF, GB) provided the best fit for effluent quality data. However, it is worth noting that RF required the most computational power compared to the other algorithms. The authors emphasized the need for a combination of different datasets, including long-term data, to improve the model accuracy.

In their intriguing work, Li et al. [31] employed principal component analysis (PCA) to select only three input parameters—mixed-liquor suspended solids (MLSS), resistance, and pressure—for the prediction of the membrane's flux. The authors evaluated different algorithms (random forest (RF), backpropagation neural networks, and support vector machine) using the Hadoop big data platform. They concluded that the random forest-based model demonstrated the lowest root mean square error (RMSE) and the highest value.

Zhao et al. [33] applied an ANN based on a radial basis function for the prediction of the interaction on a randomly rough membrane surface. The researchers quantified the interactions using the Derjaguin–Landau–Verwey–Overbeek (XDLVO) methodology. Interestingly, the constructed RBF ANN model showed satisfactory accuracy using only 2% of the computational time compared to the advanced XDLVO methodology. According to the researchers, RBF ANN appears an interesting strategy toward the study of membrane fouling and interface behavior.

Schmitt and coworkers [34] simulated the TMP in a lab-scale anoxic–aerobic membrane bioreactor (AO-MBR), treating municipal effluent using a backpropagation ANN trained using the Levenberg–Marquardt (LM) algorithm. The authors examined ten input parameters, namely, COD, MLSS, MLVSS, pH, DO, alkalinity, TN, TP, NO_3-N, and NH_4-N, while they tried to reduce the degrees of freedom of the systems through classification in different groups. According to the results presented, conventional indicators such as MLSS, COD, pH, and DO did not yield satisfactory results toward ANN modeling. In contrast, a satisfactory TMP prediction was achieved ($R^2 = 0.85$) when TN_{in}–TN_{eff}, TP_{in}–TP_{an}, and $Nitrate_{mbr}$–$Nitrate_{eff}$ were used as the input parameters of the ANN.

Giwa et al. [38] developed an artificial neural network (ANN) model to simulate the operation of a hybrid electro-assisted membrane bioreactor in Masdar City, Abu Dhabi (UAE). They utilized parameters such as dissolved oxygen (DO), mixed-liquor volatile suspended solids (MLVSS), pH, and electrical conductivity as inputs to predict the removal rates of COD, PO_4^{3-}-P, and NH_4^+-N. The optimized ANN incorporated seven neurons in the hidden layer. Among various algorithms examined for training the ANN, the Levenberg–Marquardt algorithm exhibited the highest efficiency. The ANN model demonstrated a superior modeling efficiency for all output variables, namely COD (r = 0.9942), PO_4^{3-}-P (r = 0.9998), and NH_4^+-N (r = 0.9955).

In the work of Li et al. [39], data-driven deep learning methods were applied to model and predict the treatment of real municipal wastewater using anaerobic membrane bioreactors (AnMBRs). Six parameters related to the experimental conditions (reactor temperature, environmental temperature, influent temperature, influent pH, influent COD, and flux), and eight parameters for wastewater treatment evaluation (effluent pH, effluent COD, COD removal efficiency, biogas composition (CH_4, N_2, and CO_2), biogas production rate, and oxidation-reduction potential), were selected based on one-year operating data from two AnMBRs to establish the datasets. The authors proposed three deep learning network structures to analyze and reproduce the relationship between the input parameters and the output evaluation parameters. Statistical analysis indicated that the deep learning results closely matched the AnMBR experimental results. The prediction accuracy of the proposed densely connected convolutional network (DenseNet) reached as high as 97.44%,

with a single calculation time reduced to under 1 s, suggesting the high performance of the AnMBR treatment prediction using the deep learning method.

In their groundbreaking study, Kocavs et al. [40] employed various data-driven algorithms, specifically random forest (RF), artificial neural network (ANN), and long short-term memory (LSTM) network, to predict the transmembrane pressure (TMP). The authors used high-quality data comprising 80,000 data points gathered over 4 years of operation from a full-scale treatment plant. The results indicated that while all models delivered satisfactory simulation outcomes, the best performance—gauged using indicators such as root mean squared error and the coefficient of determination—was accomplished by the RF models. Moreover, the researchers concluded that the LSTM failed to identify extreme values, and the ANN also exhibited inconsistencies when handling extreme TMP values.

Recently, Hosseinzadeh et al. [41] explored the simulation of water flux in an osmotic membrane bioreactor (OMBR) using both the adaptive network-based fuzzy inference system (ANFIS) and the artificial neural network (ANN). The selected input parameters were dissolved oxygen, conductivity, and mixed-liquor suspended solids (MLSS). The researchers evaluated the algorithms for two distinct membranes, specifically, thin-film composite (TFC) and cellulose triacetate (CTA), employing four separate datasets. Conductivity emerged as the most critical parameter for all models, except for the MLSS model of the CTA membrane when ANFIS was the employed algorithm. The root mean square error for TFC (0.2527) and CTA (0.1230) in the ANFIS models was lower than that in the ANN models, which were 0.4049 and 0.1449, respectively. Sensitivity analysis revealed that conductivity was the most influential factor for both TFC and CTA membranes in ANN models, while in ANFIS models, conductivity (for TFC) and MLSS (for CTA) were key parameters. Evaluated by the RMSE, SSE, Adj-R^2, and R^2, ANFIS displayed a higher modeling accuracy compared to ANN.

On the other hand, there are fewer applications of unsupervised learning for the simulation of MBR systems, including applying techniques such as principal component analysis (PCA) and clustering to identify trends and potential patterns in fouling pressure data [36].

Maere et al. [42] implemented a data-driven approach based on principal component analysis (PCA) and fuzzy clustering (FC), with the aim of assessing information from a single routine measurement for any membrane bioreactor (MBR), the online transmembrane pressure (TMP) measurements. Three distinct algorithms were created to infer the membrane state from the TMP data from a lab-scale reactor. These algorithms were evaluated for their potential use in designing a real-time fouling control technique. All algorithms examined demonstrated the capability to correlate patterns and data trends. However, only the two functional methods addressed outliers and noisy data. The use of B-splines did not justify the increased complexity with better results, while applying fuzzy clustering after PCA failed to classify all the available data. On the other hand, the use of factor analysis successfully exploited the observed linearity and divided the fouling effects into reversible and irreversible categories. According to the authors, each technique had advantages and drawbacks, and although they showed high potential, they needed to be implemented under different operating conditions, rather than under well-controlled laboratory experiments, to evaluate their efficiency in a wider range.

Despite unsupervised learning not being as common in relevant systems, it can often reveal hidden relationships between parameters and fouling, assisting in system optimization and the redesign of new strategies for membrane cleaning.

As with most industrial applications, the control and optimization of the system is vital to maximizing the system performance and efficiency. From this perspective, different machine learning algorithms have been implemented to develop an advanced control strategy that can dynamically adjust operational parameters such as the aeration rate, sludge retention time (SRT), and waste sludge removal rate, in response to fluctuations in process conditions [43,44].

Reinforcement learning, which is starting to gain ground in several applications, has shown promise through algorithms such as Q-learning and deep Q-networks (DQN). These have yielded promising MBR optimization and control results, learning optimal actions based on the reward signals of indicators such as energy efficiency or nutrient removal. This strategy can reduce energy consumption and chemical use while achieving cleaner effluent under varying conditions [45]. For example, Nam et al. [45] investigated a different aeration system for MBRs. The researchers utilized a deep reinforcement learning (DRL)-based optimal operating system, with the goal of minimizing energy consumption and enhancing the effluent quality. Indeed, the application of the deep Q-network (DQN) algorithm successfully reduced energy consumption for aeration by approximately 34%, without compromising the quality of the effluent.

Early fault detection in MBRs, including equipment malfunction and sensor failure or process disturbances, can significantly improve the reliability of the proposed systems and minimize or even eliminate the downtime and the production of effluents that do not meet the strict regulations [46,47].

In a recent study [47], the authors proposed an innovative system based on a combination of explainable AI (XAI) and a new multi-sensor fusion-based automated data reconciliation and imputation (MSF-ARI) for monitoring membrane fouling in MBRs, featuring autonomous handling of sensor malfunctions. The researchers validated the MSF-ARI technique using missing or faulty data and then employed an integrated biological–physical MBR to assess the effect on energy consumption and membrane fouling. According to the results, MSF-ARI demonstrated superior efficiency, diagnosing fault groups with a 100% detection rate. The authors concluded that the application of MSF-ARI can prevent membrane fouling in the early stages due to cake formation derived from the sludge. Furthermore, they found that the proposed strategy could achieve an optimal balance between energy consumption and operation.

Different algorithms have shown promising results, including supervised algorithms such as ANNs and SVMs, which have been used for the classification between normal and faulty operations based on data from the sensors and process parameters. Unsupervised algorithms, including anomaly detection algorithms, can also identify unusual patterns or deviations from the expected results as a possible indication of failure.

3.5. Challenges and Limitations of ML in MBR Wastewater Treatment

Integrating machine learning algorithms into advanced wastewater treatment plants, such as membrane bioreactors, presents a promising approach to enhance the system efficiency and sustainability through minimizing energy and environmental footprints. However, several limitations need to be addressed to fully leverage ML in these systems.

One of the biggest obstacles for the application of advanced algorithms in wastewater treatment units, including MBRs, is the need for high-quality and substantial amounts of data, required to train and validate the relevant models. Data quality is crucial, as an ML model trained with poor-quality data could produce misleading predictions or propose control strategies far from the optimal region [21–24]. Regrettably, data issues are common in relevant plants, including missing or noisy, inconsistent data resulting from sensor malfunction, measurement errors, or even data handling problems. In addition to quality, data abundance is also a critical issue for implementing advanced processes. ML techniques require large datasets to train in different generalized patterns of the system and across a variety of conditions. High-quality and abundant data collection in advanced wastewater treatment plants is often limited by resources and the availability of equipment and existing infrastructure for data management. On the other hand, this increased need for both quantitative and qualitative data, as well as computational power, does not always result in a proportional increase in simulation activity. Furthermore, the increase in complexity also comes with certain disadvantages.

As already mentioned, the interpretation and transparency of these models should always be taken into consideration, particularly for ML algorithms used to support decisions

and control the treatment plant. More complex algorithms, including deep neural networks, are often considered as 'black box' models by wastewater engineers since their workings and decisions are difficult to analyze [22,24,29]. This lack of interpretability can hinder trust in these algorithms among both engineers and regulatory bodies, who often hesitate to trust models they do not fully understand. Developing a new generation of explainable or hybrid models based on clever combinations and advancement of the existing algorithms can assist in this regard, in conjunction with the broader spread of artificial intelligence in other fields [24,29,30].

Another crucial aspect is the generalizability and transferability of these algorithms. The models developed should be capable of handling a variety of inputs, including data derived from different operating conditions or different MBR systems. However, ML algorithms can have limited transferability due to the diversity of the MBR design (including membranes used) and configuration (submerged and side-stream), the variability of the wastewater composition, and the different conditions such as hydraulic retention time, MLVSS loading, and operating pressure, which establish unique relationships between the input and output parameters in each system [20,21,32,34]. Adaptation of technologies such as transfer learning, domain adaptation, and meta-learning, already utilized in different applications, could enhance the generalizability and transferability of ML models in MBR applications.

3.6. Integration of ML Models into Existing Control Systems

Currently, most membrane bioreactor (MBR) systems rely on conventional control strategies, such as the use of PID controllers, which may not be directly compatible with the selected ML algorithms [21]. Implementing ML models in the control system of an MBR could necessitate modifications in the software, hardware, and data management infrastructure, along with alterations in the procedures and workforce training. Therefore, overcoming these obstacles is crucial for facilitating the seamless adoption of ML models in MBRs, ultimately achieving optimal performance, and maximizing their impact. Since machine learning algorithms are continuously evolving and rapidly improving, their potential applications in MBR systems will likely expand.

In this context, Table 2 provides a summary of various strategies employed to manage the operation of advanced wastewater treatment systems. It begins with conventional strategies, then progresses to rule-based control systems, and eventually to more complex algorithms, such as machine learning (ML)-based control, optimization, and hybrid-combined systems. Table 2 also includes the advantages and disadvantages of each technique. Additionally, Figure 1 illustrates the indicated challenges and opportunities associated with the implementation of machine learning in membrane bioreactors.

Table 2. Control strategies used in advanced wastewater treatment.

Technology	Description	Advantages	Limitations	Examples of Applications
Conventional control strategies	Control strategies based on fixed rules, heuristics, or manual adjustments by operators.	Simple and familiar for operators. Low cost and minimal equipment requirements.	Limited ability to adapt to changing conditions. Reduced efficiency and effectiveness compared to ML-based control.	Fixed setpoints for flow rates, dissolved oxygen, and other process variables.
Rule-based control systems	Control systems utilize logical rules to determine control actions based on sensors data and variables of the system.	Account for complex interrelationships between variables, allowing high customization for specific applications.	Limited ability for learning and adaptation with time. High cost. Complex implementation.	MBR aeration is controlled by fuzzy logic. Sludge removal.

Table 2. *Cont.*

Technology	Description	Advantages	Limitations	Examples of Applications
ML-based control and optimization	Control and optimization strategies based on machine learning algorithms that learn from data to make decisions and control system variables.	Improved system performance and efficiency. Ability to adapt to changing conditions and learn over time. Decreased energy consumption and reduced chemical use.	High initial investment and equipment requirements. Complex implementation and difficult maintenance.	ML-based control for nutrient removal and MBR fouling control.
Hybrid systems combining conventional and ML-based control	Combine the benefits of both conventional and ML-based control strategies, improving system performance.	Very efficient. High customization for specific applications.	Requirement of additional equipment. Complex implementation.	Hybrid rule-based and ML-based, controlling membrane fouling and nutrient removal.

Machine learning implementation in MBR

Limitation/Challenge

- Interpretability and transparency
- Data quality and quantity
- Generalizability and transferability
- Integration into existing control systems
- Policy and regulatory considerations

Opportunity/Application

- Process optimization and control
- Improved data analysis and prediction
- Decision-making support
- Novel MBR designs and hybrid systems
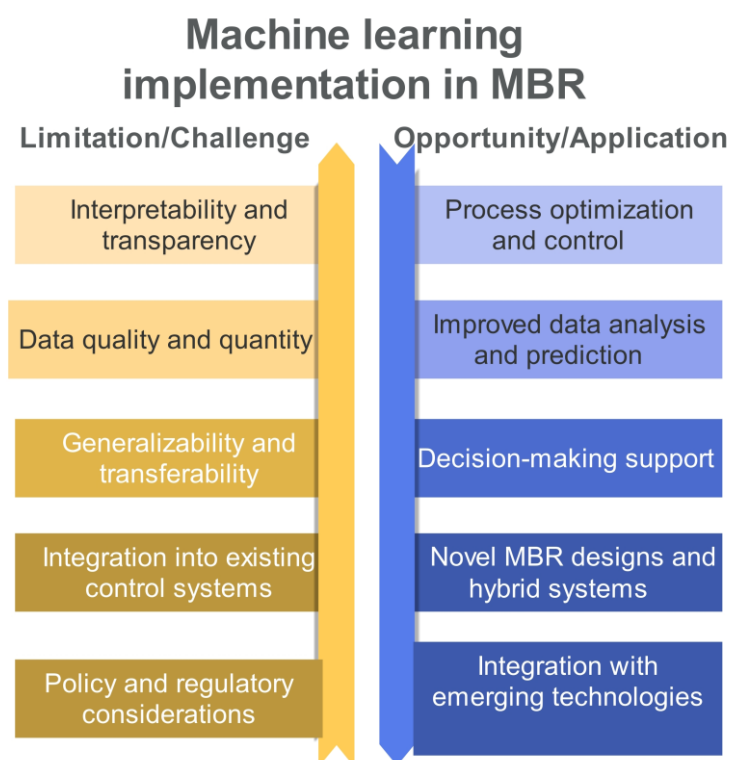- Integration with emerging technologies

Figure 1. Machine learning implementation in MBR: challenges and opportunities.

The integration of ML with other emerging technologies, such as the new generation of advanced sensors and the Internet of Things (IoT), has the potential to implement MBR monitoring, control, and optimization in real time [48,49].

Combining ML algorithms with emerging technologies boasts significant advantages, such as enhanced data collection and analysis using IoT-enabled sensors and new-generation devices. This facilitates high-resolution and quality data collection, addressing these critical issues. Advanced ML algorithms, designed with the capability to analyze big data generated from IoT devices, can extract valuable insights and patterns, useful for optimizing MBRs [14,48]. The integration of ML with the new generation of sensors and IoT devices will enable real-time MBR control, and hence, the continuous optimization of process efficiency. Leveraging these advanced ML algorithms to predict possible system

changes allows the operators to devise more efficient control strategies with higher accuracy, thus enhancing MBR performance [48–50].

3.7. Enhancing Membrane Bioreactor Design through Data-Driven Machine Learning for Sustainable Wastewater Treatment and Resource Recovery

On the other hand, beyond optimizing existing MBRs, as shown in Figure 1 ML can be used to drive the design of novel MBR configurations, including hybrid systems. More sophisticated ML algorithms can be implemented to analyze and model the relationships between the MBR design parameters, operating conditions, and outcomes. This strategy facilitates the identification of optimal configurations for specific requirements. The proposed data-driven approach for MBR design could lead to highly efficient systems, tailored to the specific needs of various environmental problems. Meanwhile, the novel design can take into account the environmental and energy footprints using unified approaches, such as lifecycle analysis [50,51], implementing, for example, the need for less use of chemicals. At the same time, integrating ML with other wastewater treatment techniques, including anaerobic digestion or physicochemical processes, such as advanced oxidation processes (AOPs) and nutrient recovery under the perspective of the circular economy, could result in innovative hybrid systems with enhanced performance and sustainability. In this approach, ML models can optimize both the operation and control of these systems and estimate the optimal combination of different processes in terms of energy efficiency, resource recovery, and the environmental footprint.

3.8. Policy and Regulatory Considerations for ML Implementation in Wastewater Treatment

An additional aspect to consider regarding the broad applications of machine learning in similar systems is relevant policy and legislation. Among the issues that could arise are data privacy and security. In fact, data collection, storage, and analysis may raise concerns related to privacy and the security of the data. Therefore, there is a need to develop policies and regulations that ensure data protection and safeguard sensitive information derived from the efficient use of ML for MBR control and optimization [43,44]. Another critical issue is model validation and certification. A process must be established to ensure the reliability and safety of relevant ML algorithms for use in advanced wastewater treatment plants and to verify that such models meet performance standards while demonstrating the necessary interpretability and generalizability. Simultaneously, a skilled workforce is required to properly implement ML in relevant systems. Both industry and policymakers must invest in training and educational projects to develop the necessary expertise for engineers and operators working on MBRs. Finally, to further promote such applications, engaging with relevant stakeholders, such as treatment plant operators, engineers, legislative bodies, and the general public, is vital to communicate the advantages of implementing ML in environmental protection in similar systems.

4. Conclusions and Recommendations for Future Research

Integrating machine learning with membrane bioreactor (MBR) systems has already demonstrated great potential in improving the efficiency and sustainability of advanced wastewater treatment, providing high-quality effluent for several applications, including water reuse. This review has provided an overview of various ML technologies and their applications in relevant systems, specifically focusing on predicting membrane fouling, process control, optimization, early fault detection, and designing new cleaning strategies. Additionally, future research directions have been proposed, such as integrating big data and the Internet of Things (IoT), and the development of hybrid models. Considering the current state-of-the-art, the following recommendations are proposed to enhance the implementation.

To fully exploit ML's potential in MBR systems, it is crucial to strengthen interdisciplinary collaboration between chemical engineers, environmental scientists, and computer scientists. This collaboration will allow for the development of novel ML models that,

in addition to accuracy and efficiency, will be relevant and interpretable for the personnel involved in wastewater engineering. Investing in data collection and management is a prerequisite, followed by developing standardized data formats and protocols that can handle high-quality big data.

The development of interpretable models, coupled with relevant visualization tools, will facilitate their use by MBR operators and incorporate techniques such as LIME or Shapley Additive exPlanations (SHAP) into existing models. In particular, the development of user-friendly tools capable of visualizing the data will bridge the gap between the predictions derived from the algorithms and human understanding, thus enabling practitioners to leverage ML in their daily operations. The adaptability and robustness of ML algorithms must be enhanced to account for the different disturbances and fluctuations in influent observed in similar systems that could potentially affect the performance and, therefore, the effluent quality. Different approaches, such as transfer and online learning, must be implemented in ML algorithms to ensure reliable predictions and MBR simulation under varying conditions. Similar to most AI tools, open-source development and knowledge sharing within the research community will accelerate the progress and implementation of these technologies in advanced wastewater treatment plants. This acceleration can be achieved through open-access publication, sharing datasets and code in relevant repositories, and organizing related workshops. In conclusion, a machine learning-driven MBR system appears to be a promising approach for improving advanced wastewater treatment, enabling more efficient and sustainable operations. The relevant technology is rapidly advancing, and by addressing the challenges mentioned above, such systems have great potential to prevail in a highly competitive industrial sector.

Author Contributions: Conceptualization, Z.F., G.L. and A.S.; methodology, Z.F., G.L. and A.S.; formal analysis, Z.F., G.L. and A.S.; investigation, Z.F., G.L. and A.S.; writing—original draft preparation, Z.F., G.L. and A.S.; writing—review and editing, Z.F., G.L. and A.S.; visualization, Z.F.; project administration, G.L. and A.S.; funding acquisition, G.L. and A.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Al-Asheh, S.; Bagheri, M.; Aidan, A. Membrane bioreactor for wastewater treatment: A review. *Case Stud. Chem. Environ. Eng.* **2021**, *4*, 100109. [CrossRef]
2. Asante-Sackey, D.; Rathilal, S.; Tetteh, E.K.; Armah, E.K. Membrane Bioreactors for Produced Water Treatment: A Mini-Review. *Membranes* **2022**, *12*, 275. [CrossRef] [PubMed]
3. Krzeminski, P.; Leverette, L.; Malamis, S.; Katsou, E. Membrane bioreactors—A review on recent developments in energy reduction, fouling control, novel configurations, LCA and market prospects. *J. Memb. Sci.* **2017**, *527*, 207–227. [CrossRef]
4. Luo, W.; Hai, F.I.; Price, W.E.; Guo, W.; Ngo, H.H.; Yamamoto, K.; Nghiem, L.D. High retention membrane bioreactors: Challenges and opportunities. *Bioresour. Technol.* **2014**, *167*, 539–546. [CrossRef]
5. Santos, A.; Ma, W.; Judd, S.J. Membrane bioreactors: Two decades of research and implementation. *Desalination* **2011**, *273*, 148–154. [CrossRef]
6. Meng, F.; Chae, S.R.; Drews, A.; Kraume, M.; Shin, H.S.; Yang, F. Recent advances in membrane bioreactors (MBRs): Membrane fouling and membrane material. *Water Res.* **2009**, *43*, 1489–1512. [CrossRef]
7. Goswami, L.; Vinoth Kumar, R.; Borah, S.N.; Arul Manikandan, N.; Pakshirajan, K.; Pugazhenthi, G. Membrane bioreactor and integrated membrane bioreactor systems for micropollutant removal from wastewater: A review. *J. Water Process Eng.* **2018**, *26*, 314–328. [CrossRef]
8. Judd, S. The status of membrane bioreactor technology. *Trends Biotechnol.* **2008**, *26*, 109–116. [CrossRef]
9. Xiao, K.; Liang, S.; Wang, X.; Chen, C.; Huang, X. Current state and challenges of full-scale membrane bioreactor applications: A critical review. *Bioresour. Technol.* **2019**, *271*, 473–481. [CrossRef]

10. Bhattacharyya, A.; Liu, L.; Lee, K.; Miao, J. Review of Biological Processes in a Membrane Bioreactor (MBR): Effects of Wastewater Characteristics and Operational Parameters on Biodegradation Efficiency When Treating Industrial Oily Wastewater. *J. Mar. Sci. Eng.* **2022**, *10*, 1229. [CrossRef]

11. Sahith, J.K.; Lal, B. Artificial Intelligence in Water Treatment Process Optimization. *Gas Hydrate Water Treat. Technol. Econ. Ind. Asp.* **2022**, 139–153.

12. El-Rawy, M.; Abd-Ellah, M.K.; Fathi, H.; Ahmed, A.K.A. Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. *J. Water Process Eng.* **2021**, *44*, 102380. [CrossRef]

13. Ramesh, P.; Suganya, K.; Maheswari, T.U.; Sebastian, S.P.; Banu, K.S.P. Relevance of Artificial Intelligence in Wastewater Management. *Digit. Agric. Revolut. Innov. Chall. Agric. Through Technol. Disrupt.* **2022**, 311–332.

14. Kamali, M.; Appels, L.; Yu, X.; Aminabhavi, T.M.; Dewil, R. Artificial intelligence as a sustainable tool in wastewater treatment using membrane bioreactors. *Chem. Eng. J.* **2021**, *417*, 128070. [CrossRef]

15. Nourani, V.; Asghari, P.; Sharghi, E. Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data. *J. Clean. Prod.* **2021**, *291*, 125772. [CrossRef]

16. Zhao, L.; Dai, T.; Qiao, Z.; Sun, P.; Hao, J.; Yang, Y. Application of artificial intelligence to wastewater treatment: A bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse. *Process Saf. Environ. Prot.* **2020**, *133*, 169–182. [CrossRef]

17. Malviya, A.; Jaspal, D. Artificial intelligence as an upcoming technology in wastewater treatment: A comprehensive review. *Environ. Technol. Rev.* **2021**, *10*, 177–187. [CrossRef]

18. Nourani, V.; Elkiran, G.; Abba, S.I. Wastewater treatment plant performance analysis using artificial intelligence—An ensemble approach. *Water Sci. Technol.* **2018**, *78*, 2064–2076. [CrossRef]

19. Mamandipoor, B.; Majd, M.; Sheikhalishahi, S.; Modena, C.; Osmani, V. Monitoring and detecting faults in wastewater treatment plants using deep learning. *Environ. Monit. Assess.* **2020**, *192*, 148. [CrossRef]

20. Ray, S. A Quick Review of Machine Learning Algorithms. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 35–39.

21. Sundui, B.; Ramirez Calderon, O.A.; Abdeldayem, O.M.; Lázaro-Gil, J.; Rene, E.R.; Sambuu, U. Applications of machine learning algorithms for biological wastewater treatment: Updates and perspectives. *Clean Technol. Environ. Policy* **2021**, *23*, 127–143. [CrossRef]

22. Sun, A.Y.; Scanlon, B.R. How can Big Data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. *Environ. Res. Lett.* **2019**, *14*, 073001. [CrossRef]

23. Azrour, M.; Mabrouki, J.; Fattah, G.; Guezzaz, A.; Aziz, F. Machine learning algorithms for efficient water quality prediction. *Model. Earth Syst. Environ.* **2022**, *8*, 2793–2801. [CrossRef]

24. Hino, M.; Benami, E.; Brooks, N. Machine learning for environmental monitoring. *Nat. Sustain.* **2018**, *1*, 583–588. [CrossRef]

25. Zuthi, M.F.R.; Ngo, H.H.; Guo, W.S. Modelling bioprocesses and membrane fouling in membrane bioreactor (MBR): A review towards finding an integrated model framework. *Bioresour. Technol.* **2012**, *122*, 119–129. [CrossRef] [PubMed]

26. Iorhemen, O.T.; Hamza, R.A.; Tay, J.H. Membrane Bioreactor (MBR) Technology for Wastewater Treatment and Reclamation: Membrane Fouling. *Membranes* **2016**, *6*, 33. [CrossRef]

27. Du, X.; Shi, Y.; Jegatheesan, V.; Ul Haq, I. A Review on the Mechanism, Impacts and Control Methods of Membrane Fouling in MBR System. *Membranes* **2020**, *10*, 24. [CrossRef]

28. Meng, F.; Zhang, S.; Oh, Y.; Zhou, Z.; Shin, H.S.; Chae, S.R. Fouling in membrane bioreactors: An updated review. *Water Res.* **2017**, *114*, 151–180. [CrossRef]

29. Dalmau, M.; Atanasova, N.; Gabarrón, S.; Rodriguez-Roda, I.; Comas, J. Comparison of a deterministic and a data driven model to describe MBR fouling. *Chem. Eng. J.* **2015**, *260*, 300–308. [CrossRef]

30. Zhong, H.; Yuan, Y.; Luo, L.; Ye, J.; Chen, M.; Zhong, C. Water quality prediction of MBR based on machine learning: A novel dataset contribution analysis method. *J. Water Process Eng.* **2022**, *50*, 103296. [CrossRef]

31. Li, W.; Li, C.; Wang, T. Application of machine learning algorithms in MBR simulation under big data platform. *Water Pract. Technol.* **2020**, *15*, 1238–1247. [CrossRef]

32. Niu, C.; Li, X.; Dai, R.; Wang, Z. Artificial intelligence-incorporated membrane fouling prediction for membrane-based processes in the past 20 years: A critical review. *Water Res.* **2022**, *216*, 118299. [CrossRef]

33. Zhao, Z.; Lou, Y.; Chen, Y.; Lin, H.; Li, R.; Yu, G. Prediction of interfacial interactions related with membrane fouling in a membrane bioreactor based on radial basis function artificial neural network (ANN). *Bioresour. Technol.* **2019**, *282*, 262–268. [CrossRef]

34. Schmitt, F.; Banu, R.; Yeom, I.T.; Do, K.U. Development of artificial neural networks to predict membrane fouling in an anoxic-aerobic membrane bioreactor treating domestic wastewater. *Biochem. Eng. J.* **2018**, *133*, 47–58. [CrossRef]

35. Wang, Z.; Zeng, J.; Shi, Y.; Ling, G. MBR membrane fouling diagnosis based on improved residual neural network. *J. Environ. Chem. Eng.* **2023**, *11*, 109742. [CrossRef]

36. Jiang, T.; Zhang, H.; Gao, D.; Dong, F.; Gao, J.; Yang, F. Fouling characteristics of a novel rotating tubular membrane bioreactor. *Chem. Eng. Process. Process Intensif.* **2012**, *62*, 39–46. [CrossRef]

37. Kulesha, O.; Maletskyi, Z.; Ratnaweera, H. Multivariate Chemometric Analysis of Membrane Fouling Patterns in Biofilm Ceramic Membrane Bioreactor. *Water* **2018**, *10*, 982. [CrossRef]
38. Giwa, A.; Daer, S.; Ahmed, I.; Marpu, P.R.; Hasan, S.W. Experimental investigation and artificial neural networks ANNs modeling of electrically-enhanced membrane bioreactor for wastewater treatment. *J. Water Process Eng.* **2016**, *11*, 88–97. [CrossRef]
39. Li, G.; Ji, J.; Ni, J.; Wang, S.; Guo, Y.; Hu, Y.; Liu, S.; Huang, S.F.; Li, Y.Y. Application of deep learning for predicting the treatment performance of real municipal wastewater based on one-year operation of two anaerobic membrane bioreactors. *Sci. Total Environ.* **2022**, *813*, 151920. [CrossRef]
40. Kovacs, D.J.; Li, Z.; Baetz, B.W.; Hong, Y.; Donnaz, S.; Zhao, X.; Zhou, P.; Ding, H.; Dong, Q. Membrane fouling prediction and uncertainty analysis using machine learning: A wastewater treatment plant case study. *J. Memb. Sci.* **2022**, *660*, 120817. [CrossRef]
41. Hosseinzadeh, A.; Zhou, J.L.; Altaee, A.; Baziar, M.; Li, X. Modeling water flux in osmotic membrane bioreactor by adaptive network-based fuzzy inference system and artificial neural network. *Bioresour. Technol.* **2020**, *310*, 123391. [CrossRef]
42. Maere, T.; Villez, K.; Marsili-Libelli, S.; Naessens, W.; Nopens, I. Membrane bioreactor fouling behaviour assessment through principal component analysis and fuzzy clustering. *Water Res.* **2012**, *46*, 6132–6142. [CrossRef]
43. Bagheri, M.; Akbari, A.; Mirbagheri, S.A. Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review. *Process Saf. Environ. Prot.* **2019**, *123*, 229–252. [CrossRef]
44. Nam, K.J.; Heo, S.K.; Rhee, G.H.; Kim, M.J.; Yoo, C.K. Dual-objective optimization for energy-saving and fouling mitigation in MBR plants using AI-based influent prediction and an integrated biological-physical model. *J. Memb. Sci.* **2021**, *626*, 119208. [CrossRef]
45. Nam, K.J.; Heo, S.K.; Loy-Benitez, J.; Ifaei, P.; Yoo, C.K. An autonomous operational trajectory searching system for an economic and environmental membrane bioreactor plant using deep reinforcement learning. *Water Sci. Technol.* **2020**, *81*, 1578–1587. [CrossRef] [PubMed]
46. Santos, A.V.; Lin, A.R.A.; Amaral, M.C.S.; Oliveira, S.M.A.C. Improving control of membrane fouling on membrane bioreactors: A data-driven approach. *Chem. Eng. J.* **2021**, *426*, 131291. [CrossRef]
47. Ba-Alawi, A.H.; Nam, K.J.; Heo, S.K.; Woo, T.Y.; Aamer, H.; Yoo, C.K. Explainable multisensor fusion-based automatic reconciliation and imputation of faulty and missing data in membrane bioreactor plants for fouling alleviation and energy saving. *Chem. Eng. J.* **2023**, *452*, 139220. [CrossRef]
48. Zhang, W.; Ma, F.; Ren, M.; Yang, F. Application with Internet of things technology in the municipal industrial wastewater treatment based on membrane bioreactor process. *Appl. Water Sci.* **2021**, *11*, 52. [CrossRef]
49. Lowe, M.; Qin, R.; Mao, X. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. *Water* **2022**, *14*, 1384. [CrossRef]
50. Tsui, T.H.; Zhang, L.; Zhang, J.; Dai, Y.; Tong, Y.W. Engineering interface between bioenergy recovery and biogas desulfurization: Sustainability interplays of biochar application. *Renew. Sustain. Energy Rev.* **2022**, *157*, 112053. [CrossRef]
51. Tsui, T.H.; van Loosdrecht, M.C.M.; Dai, Y.; Tong, Y.W. Machine learning and circular bioeconomy: Building new resource efficiency from diverse waste streams. *Bioresour. Technol.* **2023**, *369*, 128445. [CrossRef]