Algae and Environment
https://www.sciltp.com/journals/AlgaeEnviron

Review

A Mini Review on Fundamentals and Practical Applications of Machine Learning in Algae-Based Wastewater Treatment

Jiawen Zhang [1], Wenshan Guo [1], Huu Hao Ngo [1,*], Xuan Thanh Bui [2], Tra Van Tung [3] and Huiying Zhang [4,*]

[1] Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NWS 2007, Australia
[2] Key Laboratory of Advanced Waste Treatment Technology & Faculty of Environment and Natural Resources, Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh 700000, Vietnam
[3] Institute of Applied Technology and Sustainable Development, Nguyen Tat Thanh University, Ho Chi Minh 70000, Vietnam
[4] College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China
* Correspondence: ngohuuhao121@gmail.com (H.H.N.); 15880401686@163.com (H.Z.)

How To Cite: Zhang, J.; Guo, W.; Ngo, H.H.; et al. A Mini Review on Fundamentals and Practical Applications of Machine Learning in Algae-Based Wastewater Treatment. *Algae and Environment* **2025**, *1*(1), 2.

Received: 25 February 2025
Revised: 14 April 2025
Accepted: 12 May 2025
Published: 14 May 2025

Abstract: With the growing demand for sustainable wastewater treatment, algae-based technologies have gained increasing attention as a promising solution, owing to their capacity to efficiently remove pollutants and recover valuable nutrients from various wastewater sources. Microalgae offer a cost-effective and environmentally friendly approach by combining biological treatment with resource recovery. Despite their potential, large-scale deployment is often constrained by environmental variability and the physiological complexity of microalgal systems. Machine learning (ML), a key branch of artificial intelligence (AI), has emerged as a powerful tool for predicting pollutant levels and water quality parameters, due to its ability to model complex, non-linear relationships between input variables and system responses. Recent advancements in ML present new opportunities to enhance process control, operational stability, and treatment efficiency. This review explores the application of ML techniques, including neural networks, support vector machines, decision trees, and genetic algorithms in the modelling, prediction, and optimisation of microalgae-based wastewater treatment processes. It further discusses the potential of intelligent algorithms to manage large, complex datasets and address operational uncertainties, while also identifying current limitations and future directions for integrating AI in algae-based treatment systems.

Keywords: wastewater treatment; microalgae; machine learning; prediction and optimization

1. Introduction

Global water scarcity is becoming increasingly severe due to improper water management, population growth, and climate change [1]. Algae-based wastewater treatment harnesses algae's natural biological processes to offer a sustainable, cost-effective solution. Through photosynthesis, algae absorb carbon dioxide and remove pollutants such as nitrogen and phosphorus from wastewater, improving water quality. This approach is gaining attention for its dual benefits: purifying water while producing algal biomass that can be converted into biofuels or used as animal feed.

As an advanced sustainable technology, microalgal wastewater treatment can potentially eliminate emerging pollutants. Microalgae use both organic and inorganic compounds in wastewater as nutrients, proliferating to remove contaminants while simultaneously fixing carbon dioxide, producing oxygen, and reducing the overall carbon footprint of the treatment process [2]. However, their growth and pollutant removal efficiency depend on several factors, including wastewater characteristics, cultivation methods, light intensity, nutrient availability, and carbon dioxide levels [3]. In environments with high pollutant concentrations—such as elevated levels of total nitrogen, total phosphorus, and chemical oxygen demand—certain heavy metals can exert toxic effects on microalgal stability [4]. Despite these advantages, practical challenges remain, including optimizing growth conditions, system control, and large-scale implementation. Machine learning (ML) has gained significant traction in environmental fields such as air pollution, wastewater, and solid waste management, serving as a powerful tool for knowledge acquisition and integration. Its core principle is inductive reasoning, which enables the identification of patterns between inputs and outputs to support data-driven decision-making [5].

The advent of machine learning (ML) has revolutionized wastewater treatment, particularly in addressing the complex, dynamic nature of these systems where traditional methods falls short. ML offers predictive modelling, real-time monitoring, and optimization tools that enhance algae-based treatment systems' operational efficiency and management. By analyzing historical data, ML models can forecast algal growth, pollutant removal efficiency, and environmental impact, thereby supporting more precise decision-making. Integrating artificial intelligence (AI) and ML algorithms further improves the efficiency of microalgal wastewater treatment and resource recovery by optimizing system parameters and identifying key factors affecting microalgal cultivation [6]. Moreover, AI and ML can optimize critical parameters for algal growth—including nutrient availability, temperature, pH, and light intensity [7]—while enhancing the conversion of harvested biomass into biofuels, biohydrogen, and biofertilizers [8]. These algorithms optimize conversion processes, reducing costs and time while increasing overall productivity [6].

Researchers provided valuable insights into the application of ML in microalgae research, particularly for optimizing cultivation practices. They observed that traditional microalgal cultivation methods face limitations due to the complexity of growth dynamics [9]. Conventional approaches—affected by variables such as light intensity, temperature, pH, and nutrient availability—often lack real-time monitoring, making it challenging to respond swiftly to changes that influence algal growth. These limitations underscore the significant potential of machine learning in advancing microalgal research.

Despite significant progress, a critical knowledge gap remains in the literature, particularly concerning the effective integration of machine learning models with the practical requirements of microalgal wastewater treatment. While certain models have demonstrated promising results under controlled laboratory conditions, their application in real-world environments presents challenges related to generalizability, stability, and the ability to manage complex dynamic systems. Addressing these challenges is crucial for advancing the technology in this field.

This review provides an in-depth exploration of the application of machine learning (ML) models in microalgal wastewater treatment, systematically summarizing recent advancements and offering insights into future development trends. It begins by introducing the core concepts of machine learning, outlining fundamental methodologies, and analysing their potential value in microalgal wastewater treatment. Additionally, it elaborates on the key mechanisms involved in the treatment process. The review highlights both the advantages and limitations of machine learning models in this field, particularly their contributions to enhancing treatment efficiency, optimizing system design, and enabling intelligent monitoring. Furthermore, it examines the multiple benefits of integrating machine learning algorithms into microalgal wastewater treatment systems and envisions the future trajectory of this research area. By providing new perspectives, this review underscores the transformative potential of artificial intelligence and machine learning in enhancing monitoring efficiency and optimizing wastewater treatment system design.

2. Theoretical Basis and Applications

2.1. Algae Cultivation and Used in Wastewater Treatment

As primary producers in aquatic ecosystems, algae play a crucial role in wastewater treatment. Algae absorb carbon dioxide from water through photosynthesis, converting it into organic matter while releasing oxygen. They also utilize nutrients such as nitrogen and phosphorus from the water for growth, effectively removing these pollutants and reducing the risk of eutrophication.

The mechanisms by which algae remove nutrients from wastewater include assimilation and adsorption. Assimilation refers to absorbing nutrients like nitrogen and phosphorus into algal cells, where they form cellular components. At the same time, adsorption involves binding nutrients to the surface of algal cells, which can be removed through algal harvesting or filtration. Nutrient removal efficiency is influenced by factors such as algal species, water temperature, pH, light intensity, and nutrient concentration [6].

The efficiency of algae-based wastewater treatment is not only affected by nutrient concentration, light intensity, and water temperature but also by the interference of other pollutants. Algal systems require specific growth conditions, such as optimal light intensity and temperature ranges, to ensure efficient operation. The balance of nitrogen to phosphorus is critical for the algae's nutrient uptake efficiency and to prevent excessive growth. Despite the advantages of algae-based wastewater treatment, such as low-cost, high-energy efficiency, and environmental friendliness, along with the production of biomass for biofuels or animal feed that supports the circular economy, the system's efficiency is highly dependent on ecological conditions and faces management and technical challenges in large-scale industrial applications.

Algal cultivation is a highly complex process influenced by the characteristics of the cultivation system (e.g., reactor geometry and power input), photosynthetic efficiency, and microbial growth dynamics. These dynamics are primarily governed by environmental factors such as light intensity, pH, salinity, nutrient concentration, temperature, carbon dioxide, and dissolved oxygen levels [10]. Optimizing these conditions is crucial for enhancing biomass productivity, and conventional optimization approaches, ranging from physical models to data-driven and hybrid methods—are commonly employed. However, traditional dynamic models often fall short when applied to algae-based wastewater treatment systems. While these models can fit experimental data, they tend to oversimplify complex biological interactions and typically fail to capture subtle dynamics [11].

The Monod model, a classical kinetic model, is frequently used to describe algal growth and nutrient uptake in wastewater treatment. It assumes a saturable relationship between nutrient concentration and growth rate but overlooks critical factors, such as environmental influences and interactions between algae and other microorganisms [12]. To address these limitations, the Mathematical Model of the Algal-Bacterial Symbiotic System (MBRM) was developed to better simulate the microbial processes in integrated algal-bacterial systems, particularly for livestock and poultry wastewater. Although this model improves accuracy in representing microbial interactions, its performance declines under highly complex or variable water quality conditions [13].

Machine learning (ML) offers a powerful alternative to traditional modelling approaches in optimizing microalgae-based wastewater treatment. Unlike conventional models that rely on fixed assumptions and simplified kinetic equations, ML algorithms can learn directly from large datasets, capturing complex nonlinear relationships without requiring predefined growth functions. ML is particularly advantageous for accommodating diverse microalgal species, varying culture conditions, and fluctuating wastewater compositions. Additionally, ML models can integrate both historical and real-time monitoring data to enhance prediction accuracy and system adaptability [14].

Wei et al. (2025) [15] proposed an advanced algal-bacterial symbiotic system combining anaerobic acidification and microalgal bioaugmentation. By leveraging metabolic synergies among anaerobic bacteria, aerobic bacteria, and microalgae, this system established an efficient metabolic network that significantly enhanced wastewater treatment performance. Using response surface methodology (RSM) to optimize environmental parameters, the system achieved 98.56% chemical oxygen demand (COD) removal and biomass production of 3.43 g/L. This dual-function system improved organic load reduction, minimized aeration requirements, and facilitated resource recovery. Moreover, ML can be integrated to analyze multi-variable processes, offering a more comprehensive and accurate understanding of microalgae–wastewater interactions.

2.2. Machine Learning model Basic

Machine learning (ML), a pivotal subset of artificial intelligence (AI), is crucial in transforming data into actionable insights across various fields. Artificial Intelligence (AI) refers to the capability of machines to exhibit human-like intelligence, including learning, reasoning, planning, perceiving the environment, and making decisions. While AI represents the overarching goal, Machine Learning (ML) is a key method for achieving it—by automatically learning patterns from data and making informed judgments [16]. By utilizing algorithms, ML utilizes computers to autonomously learn from vast datasets and make decisions based on patterns identified within the data by using algorithms. ML has proven invaluable in wastewater treatment by analysing large-scale data, analysing hidden patterns, and predicting system behaviours. This capability enhances the efficiency and sustainability of wastewater treatment processes by optimizing operations and usage.

ML algorithms can be broadly classified into three categories: supervised learning, unsupervised learning, and deep learning. Supervised learning uses labelled data to train models, making it suitable for classification tasks. In the context of wastewater treatment, supervised learning can predict critical parameters, such as nutrient removal efficiency or the removal of contaminants from the water, based on historical data. Using labelled datasets in supervised learning ensures high accuracy and reliability in predicting known outcomes, essential for optimizing wastewater management practices [14].

Figure 1 presents a basic overview of supervised learning algorithms. A typical end-to-end machine learning (ML) workflow comprises three main stages: training, cross-validation, and testing. During the training phase, the model learns by adjusting its parameters based on the input data. Cross-validation follows, where hyperparameters are fine-tuned using a validation set to enhance model performance. In the final stage, testing assesses the model's generalization capability using an independent dataset. Once optimized, the trained model can be applied to predictive tasks [17].

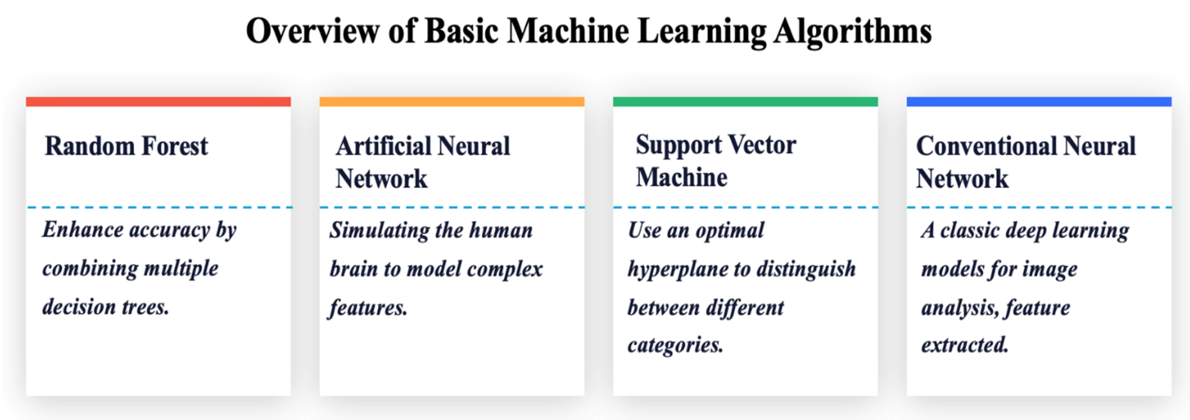

Figure 1. Machine learning models algorithms.

By leveraging ML algorithms, researchers can efficiently analyze large datasets to identify key predictive variables and discover meaningful patterns [18]. Commonly used supervised learning algorithms include Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), and Random Forests (RF). ANN mimics the structure of the human brain by adjusting connection weights through backpropagation. SVMs perform classification and regression by finding the optimal hyperplane that maximizes the margin between classes. Decision Trees use hierarchical feature splitting for classification and regression, while Random Forests combine multiple decision trees using ensemble learning to improve predictive accuracy and reduce overfitting [14].

Unsupervised learning, on the other hand, does not rely on labelled data. Instead, it identifies inherent structures or patterns in the data that may not be immediately obvious. This approach is helpful in scenarios where data labelling is not feasible or when the goal is to uncover hidden relationships within large, complex datasets, such as detecting previously unknown trends in wastewater quality or identifying system failures early on. Unsupervised learning methods can also be applied to classify or group wastewater treatment systems into different categories based on their operational conditions. This helps in decision-making regarding system improvements or upgrades.

Deep learning plays a pivotal role in real-time system optimization. Among its approaches, reinforcement learning (RL) is particularly distinguished from supervised and unsupervised learning by its interactive framework: an agent learns by interacting with an environment, continuously adjusting operational parameters to maximize cumulative rewards based on feedback. In the context of wastewater treatment, this could involve dynamically regulating variables such as chemical dosages, temperature, or pH levels to optimize plant performance. The real-time decision-making capability of RL enables continuous adaptation and fine-tuning of treatment processes, ensuring optimal efficiency even under fluctuating environmental conditions [14].

Several machine learning algorithms are commonly applied in wastewater treatment optimization. Decision trees, for instance, are frequently used for both classification and regression tasks. They work by splitting data into distinct branches based on input features, which makes them suitable for identifying patterns in treatment processes. Support vector machines (SVM), known for their effectiveness in high-dimensional spaces, are handy for classification tasks in wastewater treatment, where multiple variables might be at play. Artificial neural networks (ANN), intense learning models, are increasingly used to model complex nonlinear relationships in wastewater treatment systems. These models can account for the interactions among numerous factors, such as flow rates, nutrient concentrations, and chemical treatments, which are challenging to model using traditional approaches. Additionally, K-means clustering can help categorize different operators or system conditions in wastewater treatment processes, aiding in system diagnosis and maintenance scheduling [9]. Long short-term memory (LSTM) networks were used for real-time prediction of water quality and algal concentration [19], while autoregressive models (ARM) were used to analyze data and optimize wastewater treatment conditions [20].

Evaluating the performance of ML models is a critical step in determining their usefulness in real-world applications. Several metrics are employed for this purpose, including accuracy, precision, recall, F1 score, and mean squared error (MSE). MSE is particularly useful in regression tasks, as it quantifies the difference between predicted values and the actual outcomes, thus helping to assess the reliability of a model's predictions. For classification problems, precision, recall, and F1 scores are more informative. Precision measures the accuracy of optimistic predictions, recall gauges the ability of the model to capture all positive instances, and the F1 score is a harmonic mean of precision and recall, providing a balanced evaluation of model performance. In wastewater treatment, these metrics are essential for determining how well a model can predict critical parameters such as nutrient removal efficiency or the behaviour of specific contaminants [9].

Machine learning (ML) methods play a pivotal role in optimizing various aspects of microalgae production, with different models tailored to specific tasks. For strain selection, classification algorithms such as decision trees and support vector machines (SVM) are frequently employed to identify the most suitable strains for cultivation [21]. In growth rate prediction, regression models, including linear regression, random forests, and long short-term memory (LSTM) networks are used to estimate algal biomass development under varying environmental conditions [22].

For pollution detection, anomaly detection techniques like autoencoders and one-class SVM help identify system irregularities, thereby maintaining cultivation quality [23]. In environmental monitoring, time series analysis models such as ARIMA, LSTM, and gated recurrent units (GRU) are applied to track and forecast ecological variables, ensuring optimal growing conditions [24].

These diverse ML techniques form a robust toolkit for addressing the complex challenges of microalgae production, significantly enhancing both efficiency and sustainability. For instance, Hossain et al. (2022) [25] utilized SVM to predict nitrogen and phosphorus removal efficiencies of microalgae in municipal wastewater treatment. SVM outperformed multilayer perceptron artificial neural networks (MLP-ANN) and response surface methodology (RSM). However, SVM models can be computationally intensive with large datasets and are sensitive to missing data, requiring careful tuning of kernel functions, scaling, and data preprocessing.

Moreover, nutrient optimization, a critical component of microalgae-based wastewater treatment, has benefited from ML approaches. Artificial neural networks (ANN) have been effectively applied to maximize power density and chemical oxygen demand (COD) removal, both of which are key indicators of treatment efficiency.

Using ML algorithms in wastewater treatment transforms how treatment plants manage and optimize processes—optimizely predicting outcomes and optimizing operational parameters. These models improve treatment efficiency and promote sustainability by reducing resource usage and operational costs. However, while ML holds excellent potential, challenges remain, particularly in data quality, model interpretability, and real-time application. For instance, the performance of ML models can be affected by the quality and completeness of the data used for training, necessitating further research into data preprocessing and enhancement techniques.

Additionally, the black-box nature of some ML models and intense learning approaches make it challenging to interpret their decision-making processes, which could be a barrier to adoption in highly regulated industries such as wastewater treatment. Therefore, future research should improve model transparency and develop hybrid models that combine traditional engineering knowledge with ML to achieve more reliable and interpretable outcomes. Machine learning holds great promise for enhancing the efficiency and sustainability of wastewater treatment processes. By leveraging advanced algorithms, treatment plants can optimize their optimizer, optimize environmental impacts, and contribute to more sustainable water resource management.

3. Role of Machine Learning in Algae-Driven Wastewater Treatment

3.1. Microalgal Detection and Classification by ML

Detection and classification of microalgae is an important part of wastewater treatment, and traditional methods rely on microscope observation and manual classification, which is a cumbersome and inefficient process. Machine Learning technology is able to quickly and accurately identify the species and state of microalgae through image recognition and data analysis. Recently, researchers have begun exploring the implementation of intelligent control systems to optimize microalgae cultivation. For instance, researchers [26] proposed a precise control strategy based on light intensity and temperature feedback, which successfully optimized mixing in open pond cultivation of Spirulina and reduced energy input by 30% compared to the control group. Imamoglu found that feature extraction and classification of microalgae images using Convolutional Neural Networks (CNNs) can achieve highly accurate species identification. In addition, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have been widely used for microalgae classification, and the morphological features of microalgae are recognised by training models, which significantly improves the accuracy and efficiency of classification [14]. For another study, microalgae images of cyanobacterial and green algal taxa were classified using the AlexNet-SVM model [21]. Chong et al. reported accuracy of 96.93% and 97.63% for k-NN classifier and SVM classifier respectively in combination with optimised image preprocessing technique [27].

3.2. Prediction of Microalgal Growth and Pollution Removal by ML

A critical application of machine learning in algae-based wastewater treatment is predictive modelling. By leveraging historical data, machine learning algorithms can forecast key parameters such as algal growth rates and nutrient removal efficiencies, while accounting for variables like temperature, light intensity, and nutrient concentrations. These predictions are valuable for guiding system design, operational strategies, and process optimization. As illustrated in Figure 2, machine learning models have been applied across various aspects of microalgae-based wastewater treatment, including growth prediction, species identification, biomass estimation, process optimization, resource recovery, and reuse. Additionally, they play a vital role in real-time monitoring and feedback control. Collectively, these applications enhance the efficiency, reliability, and automation of microalgae-based treatment systems, supporting the development of intelligent and sustainable wastewater management solutions.

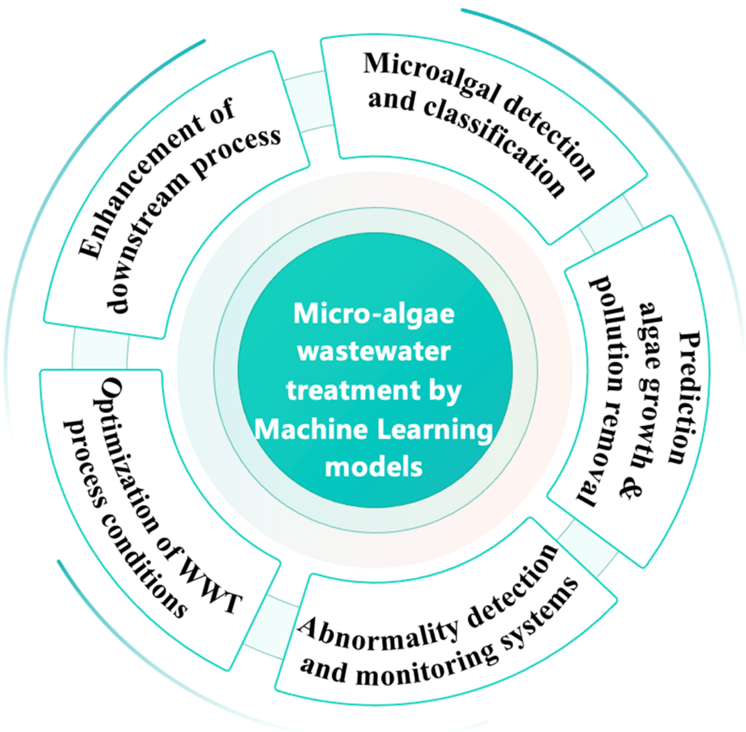

Figure 2. Machine learning model used in microalgae-based wastewater treatment.

Microalgae are highly effective in removing a wide range of pollutants from wastewater, including conventional contaminants such as chemical oxygen demand (COD), biochemical oxygen demand (BOD), ammonia nitrogen (NH_3-N), and total phosphorus (TP), as well as trace pollutants like antibiotics, pharmaceuticals, and endocrine disruptors. These pollutants are removed through several mechanisms, including adsorption, biodegradation, photodegradation, and volatilization. The functional groups on microalgal cell walls can adsorb cationic pollutants, while their metabolic activities decompose organic compounds. Light energy from photosynthesis can facilitate photodegradation, and certain volatile compounds may be released into the atmosphere through algal processes. In algal-bacterial symbiosis systems, the synergistic interactions between microalgae and bacteria further enhance pollutant removal. Organic matter excreted by microalgae stimulates bacterial heterotrophic metabolism, accelerating degradation processes. In return, bacteria regenerate nutrients that support algal growth. For instance, *Scenedesmus obliquus* was reported to remove 91.43% of BOD, 83.11% of COD, 83.74% of total nitrogen, and 54.69% of total phosphorus from pig farm wastewater [28].

Machine learning has become a valuable tool in optimizing microalgae-based wastewater treatment. By constructing models such as random forests, researchers can analyze the influence of multiple environmental parameters on algal growth and predict future trends. Real-time monitoring and control are essential to maintaining optimal system performance. Machine learning algorithms, trained on historical data, support predictive modeling of ideal growth conditions. For example, Haro et al. (2022) employed random forest models to analyze algal behavior, predict growth patterns, and optimize production strategies. Other machine learning approaches—such as long short-term memory (LSTM), extreme gradient boosting (XGBoost), and random vector functional link (RVFL)—have also been used to predict growth rates [29]. A recent study by Meenatchi Sundaram et al. (2025) demonstrated that the RVFL model achieved exceptionally low prediction errors, underscoring its potential for accurately forecasting microalgae growth—an essential factor in optimizing cultivation and harvesting processes [30]. These predictive tools support better resource allocation and decision-making in large-scale algae wastewater treatment systems. Table 1 summarizes key machine learning methods applied in this field, showcasing their diverse applications and performance outcomes.

Table 1. Representative machine learning methods for microalgae-based wastewater treatment.

Microalgae Process	Model Input	Machine Learning Model Used and Activity	Outcome	References
Microalgae feature selection	Spectral data	DT, RF, SVM; Classification	The RF model accurately classified algal communities into 13 major classes and effectively modelled total biomass	[22]
Wastewater treatment efficiency	Spectral data	LSTM, CNN; Time series analysis	LSTM consistently outperforms other methods with minimal prediction error	[24]
Model optimization	Spectral data	SVR	Effect of temperature, light-dark cycle and nitrogen-phosphorus ratio on CO_2 bio fixation with 91.1% accuracy	[25]
Microalgae growth rate prediction	Morphologic data	LSTM, XGBoost, RVFL; Regression	The RVFL model can effectively predict the growth of microalgae with an error of less than 0.01	[30]
Distinguishing between dead and alive microalgal communities	Spectral data	RF	The accuracy rate is 94.5%	[31]
Nutrient optimization	Morphologic data	ANN: optimization	maximised power density and COD removal, and the combination of ANN and FBI resulted in a 2.24% increase in performance	[32]
Model optimization	Morphologic data	ANN	Bio fixation of carbon dioxide at 91.1%	[33]
Microalgae feature selection	Morphologic data	CNN	The model achieved 89% accuracy on the test set.	[34]
Model optimization	Morphologic data	GA-ANFIS	Carbon dioxide fixation at 98.4%	[35]

3.3. Abnormality Detection and Monitoring Systems by ML

By integrating Internet of Things (IoT) sensors, machine learning algorithms can continuously track key parameters such as algal concentration, water quality, and environmental conditions. These algorithms can automatically adjust system settings based on real-time data to maintain optimal system performance. Large-scale data collection through remote sensors, such as satellites or drones, combined with machine learning analysis, can optimize system scheduling across multiple treatment sites. Integrating microalgae with wastewater treatment enhances wastewater purification and optimizes biological materials recovery through advanced process control systems [6]. Tham et al. (2022) designed and developed an IoT-enabled upscaled photobioreactor that allows remote parameter monitoring via a smartphone [36]. In addition, Lee et al. (2022) developed a 3D-printed real-time optical density monitoring device to predict microalgae growth dynamics from real-time data accurately [37]. Recent studies highlight the significant role of AI and machine learning (ML) systems in improving microalgae cultivation by reducing resource consumption and enabling more precise decision-making in the biofuel industry. For instance, ML techniques such as Isolation Forest and Autoencoder can automatically detect anomalies by analyzing both historical and real-time monitoring data [14]. Liu et al. (2023) [22] demonstrated that the Random Forest (RF) model effectively classified microalgal species into 13 categories, thereby enhancing predictive capabilities and supporting optimization strategies. Additionally, Zambon et al. (2019) [38] reported that integrating IoT into microalgae production processes reduced input requirements by 30% and increased yields by

20%. Similarly, Giannino et al. (2018) found that systems monitored with IoT and AI achieved a 9% higher yield compared to those without such technologies [39].

3.4. Optimization of Process Conditions for Improved Wastewater Treatment by ML

In microalgal cultivation, regulating physical and chemical parameters that influence growth—such as light intensity, pH, nutrient concentration, carbon dioxide levels, and algal biomass concentration—is crucial. Various sensors to monitor these parameters ensure optimal algal bioreactors' cultivation conditions. The vast data these monitoring systems generate can serve as inputs for optimizing parameters in artificial intelligence (AI) and machine learning (ML) models. Through real-time monitoring and automated management, along with developing appropriate ML models from collected data, the biomass productivity of microalgae and wastewater treatment efficiency can be significantly enhanced. Depending on the requirements of the final product, the results from ML models can provide feedback to maintain optimal cultivation conditions. Artificial neural networks (ANN) and genetic algorithms (GA) have been used to optimize wastewater and yeast concentrations, enhancing power density and chemical oxygen demand (COD) removal efficiency in microbial fuel cells [32]. Furthermore, ANN, support vector machines (SVM), and genetic algorithms (GA) have also been applied to microalgae classification, survival characteristic prediction, and algal concentration estimation [40]. These ML tools have significantly improved microalgae production efficiency and wastewater treatment effectiveness [9]. Machine learning models are essential for optimizing microalgae cultivation and wastewater treatment processes. They can predict the growth rate, biomass composition, production efficiency, and pollutant removal performance of microalgae. By optimizing resource input, ML can enhance microalgae biomass production while ensuring optimal wastewater treatment efficiency [41]. These parameters are species-specific, thus requiring the optimization of growth conditions to achieve the highest biomass yield and treatment efficiency.

Machine learning and real-time monitoring systems can optimize microalgae cultivation, harvesting, and drying processes to improve efficiency and sustainability [42]. Machine learning can also be applied to optimize the operational parameters of algae wastewater treatment systems. For example, reinforcement learning algorithms can adjust real-time parameters such as light intensity, nutrient input, and temperature to ensure that the system operates under optimal conditions, maximizing energy efficiency and nutrient removal [6].

3.5. Enhancement of Downstream Process

Machine Learning (ML) plays a crucial role in optimizing downstream processes in microalgae-based wastewater treatment. For instance, Support Vector Regression (SVR) models are employed to predict drying efficiency and fine-tune process parameters for improved performance. Additionally, ML models can analyze the influence of extraction parameters on product yield, enabling the optimization of extraction processes to enhance resource recovery. These applications not only increase the efficiency of downstream operations but also reduce production costs and strengthen the overall competitiveness of microalgae wastewater treatment technologies [14]. Sayed et al. (2024) [32] found that combining ANN with feedback inhibition (FBI) led to a 2.24% increase in performance, illustrating how machine learning can help fine-tune system parameters for optimal results. Regarding wastewater treatment efficiency, LSTM and convolutional neural networks (CNN) were applied for time series analysis regarding wastewater treatment efficiency. LSTM models consistently outperformed other methods with minimal prediction error, as demonstrated by Rostam et al. (2023) [24]. This further emphasizes the ability of machine learning to provide accurate, real-time predictions, enabling better control of the treatment process and ensuring optimal system performance. In the microalgae industry, IoT-based systems play a crucial role in streamlining and optimising production, promoting more sustainable and efficient practices [43]. Complementing these technologies, artificial intelligence (AI) and machine learning are increasingly integrated into intelligent control systems to minimise resource consumption and support informed decision-making in microalgae biorefineries [14]. Together, these advancements offer substantial opportunities to enhance both productivity and sustainability in the sector. Notably, recent studies highlight the transformative potential of machine learning in microalgae-based wastewater treatment. The ability to accurately predict key parameters, optimise nutrient conditions, and improve treatment efficiency is essential for advancing the sustainability of these systems. As research progresses, the integration of more sophisticated machine learning models is expected to drive further innovation in microalgae wastewater treatment, contributing to more effective and sustainable water management solutions [44].

4. Challenges and Future Perspectives

4.1. Challenges

During the data processing procedure, data often faces issues such as incompleteness, noise, or scarcity, which can limit the performance and generalizability of models. Additionally, problems related to data standardization and the reliability of sensors may also affect the accuracy and stability of the model. Many machine learning models, intense learning models, are often considered "black box" models due to their complexity and lack of interpretability. This lack of transparency can make it difficult for stakeholders to accept the system, especially in environments that require strict regulation and decision support in algae wastewater treatment [32]. Typically, the accuracy of machine learning models improves with increased data availability. However, acquiring large amounts of data in practical applications is often costly and time-consuming. Therefore, the limited availability of data can become a barrier to application. To address this, data preprocessing can be the overall quality of data. Data augmentation (DA) techniques can artificially expand the dataset based on known invariants, which helps the trained model generalize better. For example, Correa et al. (2017) proposed that data augmentation significantly improves the accuracy of deep learning models in microalgae classification tasks compared to situations without data augmentation [16]. However, it is essential to note that improper data augmentation may lead to inaccurate predictions.

Overfitting is a common issue in machine learning, where the model performs well on training data but poorly on new, unseen data. Regularization techniques and cross-validation methods are typically applied to prevent overfitting. Moreover, some machine learning models, such as deep learning, have high computational costs, which could hinder their application in practical algae wastewater treatment systems, particularly in resource-limited environments.

4.2. Future Perspective

In the future, advancements in machine learning and artificial intelligence will drive the intelligent development of algae wastewater treatment systems. Emerging technologies, such as transfer learning, are expected to address the issue of data scarcity and enhance the generalization capability of models. By integrating laboratory data, field observation data, and remote sensing data, the accuracy and robustness of predictive modelling will be significantly improved. Multi-scale data fusion will help machine learning models more accurately reflect the actual conditions of wastewater treatment systems, providing more substantial support for decision-making. Most existing studies use a single machine-learning model for prediction. However, combining different algorithms could achieve better results and should be the focus of future research. In most cases, ensemble models (e.g., combining ANN/SVM/RF with GA) outperform standalone models in terms of prediction performance, risk of overfitting, and robustness [45].

With the development of smart sensors, blockchain, and Internet of Things (IoT) technologies, the performance and operability of algae wastewater treatment systems will be further improved [43]. Machine learning will be key in data processing, analysis optimization, and system control. In response to the growing global water scarcity and pollution challenges, the integration of machine learning and sustainable water resource management will become the core focus of future research. This will help improve algae wastewater treatment efficiency, meet the increasing demand for clean water, and promote the sustainable development of environmental protection and resource recovery. Future research should prioritize several key areas: (1) the development of flexible and adaptable machine learning models capable of managing the diverse and complex characteristics of various wastewater types; (2) the design of integrated frameworks that closely couple machine learning with microalgal biological processes to enhance algorithm accuracy and real-time operational reliability; and (3) the promotion of interdisciplinary collaboration by leveraging advancements in biology, chemistry, and computational sciences to optimize the overall performance of microalgal wastewater treatment systems.

5. Conclusions

Machine learning (ML) techniques are demonstrating substantial potential and broad applicability in the field of microalgae cultivation and its diverse applications. These data-driven technologies are increasingly vital across key processes such as growth regulation, species identification, harvesting, extraction, and purification. By leveraging artificial intelligence (AI), the microalgae industry chain benefits from enhanced process efficiency, output optimization, and improved controllability.

In the context of wastewater treatment using microalgae, ML interventions have yielded notable progress. Through predictive modeling, real-time data monitoring, and the optimization of operational strategies, ML algorithms enhance treatment efficiency and environmental performance. These approaches enable precise estimation of key variables, support optimal resource allocation, and allow for timely operational adjustments. Collectively, these capabilities contribute to improved system performance and provide a scientific basis for shaping and implementing effective environmental policies. Furthermore, machine learning is accelerating the advancement of smart water systems, promoting more efficient and sustainable water resource management. Nevertheless, several challenges continue to hinder widespread adoption. These include issues related to data quality control, limited model interpretability, and the lack of transparency in complex real-world environments. Future research should prioritize addressing these limitations—for example, by refining data acquisition techniques to improve training dataset quality or by developing more interpretable models to build trust in practical applications.

The integration of AI with IoT sensors and UAV-based remote sensing also presents promising opportunities for comprehensive, full-cycle data acquisition and large-scale monitoring in microalgae cultivation. This technological synergy not only facilitates optimal environmental regulation for algal growth but also enables responsive adaptation to market dynamics. As these technologies mature and are more deeply integrated, the field is poised for significant breakthroughs and expanded development opportunities.

Author Contributions

J.Z.: Conceptualization, data curation, formal analysis, writing—original draft; W.G.: Project administration, supervision, writing—review & editing; H.H.N.: Conceptualization, validation, project administration, supervision, writing—review & editing; X.T.B.: Resources, writing—review & editing; T.V.T.: Visualization, writing—review & editing; H.Z.: Conceptualization, resources, project administration, writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research was mainly supported by University of Technology Sydney, Australia (UTS, RIA NGO and UTS NM GUO) and the Outstanding Research Talents Program of Fujian Agriculture and Forestry University (xjq202008).

Data Availability Statement

The data will be provided as requested.

Conflicts of Interest

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

References

1. Leal Filho, W.; Totin, E.; Franke, J.A.; et al. Understanding responses to climate-related water scarcity in Africa. *Sci. Total Environ.* **2022**, *806*, 150420. https://doi.org/10.1016/j.scitotenv.2021.150420.
2. Razzak, S.A.; Ali, S.A.M.; Hossain, M.M.; et al. Biological CO_2 fixation with production of microalgae in wastewater—A review. *Renew. Sustain. Energy Rev.* **2017**, *76*, 379–390. https://doi.org/10.1016/j.rser.2017.02.038.
3. Goswami, R.K.; Agrawal, K.; Verma, P. An exploration of natural synergy using microalgae for the remediation of pharmaceuticals and xenobiotics in wastewater. *Algal Res.* **2022**, *64*, 102703. https://doi.org/10.1016/j.algal.2022.102703.
4. Hena, S.; Gutierrez, L.; Croué, J.-P. Removal of pharmaceutical and personal care products (PPCPs) from wastewater using microalgae: A review. *J. Hazard. Mater.* **2021**, *403*, 124041. https://doi.org/10.1016/j.jhazmat.2020.124041.
5. Khan, M.; Chuenchart, W.; Surendra, K.C.; et al. Applications of artificial intelligence in anaerobic co-digestion: Recent advances and prospects. *Bioresour. Technol.* **2023**, *370*, 128501. https://doi.org/10.1016/j.biortech.2022.128501.
6. Oruganti, R.K.; Biji, A.P.; Lanuyanger, T.; et al. Artificial intelligence and machine learning tools for high-performance microalgal wastewater treatment and algal biorefinery: A critical review. *Sci. Total Environ.* **2023**, *876*, 162797. https://doi.org/10.1016/j.scitotenv.2023.162797.
7. Sahu, S.; Kaur, A.; Singh, G.; et al. Harnessing the potential of microalgae-bacteria interaction for eco-friendly wastewater treatment: A review on new strategies involving machine learning and artificial intelligence. *J. Environ. Manag.* **2023**, *346*, 119004. https://doi.org/10.1016/j.jenvman.2023.119004.
8. Anand, U.; Dey, S.; Parial, D.; et al. Algae and bacteria consortia for wastewater decontamination and transformation into biodiesel, bioethanol, biohydrogen, biofertilizers and animal feed: A review. *Environ. Chem. Lett.* **2023**, *21*, 1585–1609. https://doi.org/10.1007/s10311-023-01562-w.

9. Fu, W.; Li, X.; Yang, Y.; et al. Enhanced degradation of bisphenol A: Influence of optimization of removal, kinetic model studies, application of machine learning and microalgae-bacteria consortia. *Sci. Total Environ.* **2023**, *858*, 159876. https://doi.org/10.1016/j.scitotenv.2022.159876.
10. Chowdhuri, I.; Pal, S.C.; Chakrabortty, R. Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. *Adv. Space Res.* **2020**, *65*, 1466–1489. https://doi.org/10.1016/j.asr.2019.12.003.
11. Zhang, D.; Dechatiwongse, P.; Del Rio-Chanona, E.A.; et al. Modelling of light and temperature influences on cyanobacterial growth and biohydrogen production. *Algal Res.* **2015**, *9*, 263–274. https://doi.org/10.1016/j.algal.2015.03.015.
12. Okabe, S.; Kamizono, A.; Kawasaki, S.; et al. Interspecific competition and adaptation of anammox bacteria at different salinities: Experimental validation of the Monod growth model with salinity inhibition. *Water Res.* **2025**, *271*, 122883. https://doi.org/10.1016/j.watres.2024.122883.
13. Lakshminarasimman, N.; McKnight, M.M.; Neufeld, J.D.; et al. Characterizing biofilm thickness, density, and microbial community composition in a full-scale hybrid membrane aerated biofilm reactor. *Bioresour. Technol.* **2025**, *423*, 132207. https://doi.org/10.1016/j.biortech.2025.132207.
14. Imamoglu, E. Artificial Intelligence and/or Machine Learning Algorithms in Microalgae Bioprocesses. *Bioengineering* **2024**, *11*, 1143. https://doi.org/10.3390/bioengineering11111143.
15. Wei, Q.; Wang, C.; Qi, Y.; et al. Optimizing algal-bacterial systems for efficient sugar cane wastewater treatment: Pollutant removal and biomass resource recovery. *Bioresour. Technol.* **2025**, *429*, 132497. https://doi.org/10.1016/j.biortech.2025.132497.
16. Correa, I.; Drews, P.; Botelho, S.; et al. Deep Learning for Microalgae Classification. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017; pp. 20–25. https://doi.org/10.1109/ICMLA.2017.0-183.
17. Guo, H.; Wu, S.; Tian, Y.; et al. Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. *Bioresour. Technol.* **2021**, *319*, 124114. https://doi.org/10.1016/j.biortech.2020.124114.
18. Dallora, A.L.; Eivazzadeh, S.; Mendes, E.; et al. Machine learning and microsimulation techniques on the prognosis of dementia: A systematic literature review. *PLoS ONE* **2017**, *12*, e0179804. https://doi.org/10.1371/journal.pone.0179804.
19. Saboe, D.; Ghasemi, H.; Gao, M.M.; et al. Real-time monitoring and prediction of water quality parameters and algae concentrations using microbial potentiometric sensor signals and machine learning tools. *Sci. Total Environ.* **2021**, *764*, 142876. https://doi.org/10.1016/j.scitotenv.2020.142876.
20. Singh, V.; Mishra, V. Exploring the effects of different combinations of predictor variables for the treatment of wastewater by microalgae and biomass production. *Biochem. Eng. J.* **2021**, *174*, 108129. https://doi.org/10.1016/j.bej.2021.108129.
21. Sonmez, M.E.; Eczacıoglu, N.; Gumuş, N.E.; et al. Convolutional neural network—Support vector machine based approach for classification of cyanobacteria and chlorophyta microalgae groups. *Algal Res.* **2022**, *61*, 102568. https://doi.org/10.1016/j.algal.2021.102568.
22. Liu, M.; Huang, Y.; Hu, J.; et al. Algal community structure prediction by machine learning. *Environ. Sci. Ecotechnol.* **2023**, *14*, 100233. https://doi.org/10.1016/j.ese.2022.100233.
23. Oliveira, P.; Marcondes, F.S.; Duarte, M.S.; et al. Assessment of LSTM and GRU Models to Predict the Electricity Production from Biogas in a Wastewater Treatment Plant. In *Good Practices and New Perspectives in Information Systems and Technologies*; Springer Nature: Cham, Switzerland, 2024. https://doi.org/10.1007/978-3-031-60218-4_7.
24. Rostam, N.A.P.; Ahamed Hassain Malim, N.H.; Azmee, N.A.; et al. Harmful algal blooms (HAB) open issues: A review of ecological data challenges, factor analysis and prediction approaches using data-driven method. *Comput. Artif. Intell.* **2023**, *1*, 100. https://doi.org/10.59400/cai.v1i1.100.
25. Hossain, S.M.Z.; Sultana, N.; Razzak, S.A.; et al. Modeling and multi-objective optimization of microalgae biomass production and CO_2 biofixation using hybrid intelligence approaches. *Renew. Sustain. Energy Rev.* **2022**, *157*, 112016. https://doi.org/10.1016/j.rser.2021.112016.
26. Zhu, C.; Ji, Y.; Du, X.; et al. A smart and precise mixing strategy for efficient and cost-effective microalgae production in open ponds. *Sci. Total Environ.* **2022**, *852*, 158515. https://doi.org/10.1016/j.scitotenv.2022.158515.
27. Chong, J.W.R.; Khoo, K.S.; Chew, K.W.; et al. Artificial intelligence-driven microalgae autotrophic batch cultivation: A comparative study of machine and deep learning-based image classification models. *Algal Res.* **2024**, *79*, 103400. https://doi.org/10.1016/j.algal.2024.103400.
28. Abdelfattah, A.; Ali, S.S.; Ramadan, H.; et al. Microalgae-based wastewater treatment: Mechanisms, challenges, recent advances, and future prospects. *Environ. Sci. Ecotechnol.* **2023**, *13*, 100205. https://doi.org/10.1016/j.ese.2022.100205.

29. Haro, S.; Jesus, B.; Oiry, S.; et al. Microphytobenthos spatio-temporal dynamics across an intertidal gradient using Random Forest classification and Sentinel-2 imagery. *Sci. Total Environ.* **2022**, *804*, 149983. https://doi.org/10.1016/j.scitotenv.2021.149983.
30. Meenatchi Sundaram, K.; Kumar, D.; Lee, J.; et al. Time series forecasting of microalgae cultivation for a sustainable wastewater treatment. *Process Saf. Environ. Prot.* **2025**, *196*, 106845. https://doi.org/10.1016/j.psep.2025.106845.
31. Reimann, R.; Zeng, B.; Jakopec, M.; et al. Classification of dead and living microalgae Chlorella vulgaris by bioimage informatics and machine learning. *Algal Res.* **2020**, *48*, 101908. https://doi.org/10.1016/j.algal.2020.101908.
32. Sayed, E.T.; Rezk, H.; Abdelkareem, M.A.; et al. Artificial neural network based modelling and optimization of microalgae microbial fuel cell. *Int. J. Hydrogen Energy* **2024**, *52*, 1015–1025. https://doi.org/10.1016/j.ijhydene.2022.12.081.
33. Ansari, F.A.; Nasr, M.; Rawat, I.; et al. Artificial neural network and techno-economic estimation with algae-based tertiary wastewater treatment. *J. Water Process Eng.* **2021**, *40*, 101761. https://doi.org/10.1016/j.jwpe.2020.101761.
34. Hisham, S.D., Mohamad, S.E., Shapiai, M.I.; et al. Comparison of Conventional CNN Sequential API and Functional API for Microalgae Identification. *J. Adv. Res. Micro Nano Eng.* **2024**, *17*, 96–104. https://doi.org/10.37934/armne.17.1.96104.
35. Kushwaha, O.S.; Uthayakumar, H.; Kumaresan, K. Modeling of carbon dioxide fixation by microalgae using hybrid artificial intelligence (AI) and fuzzy logic (FL) methods and optimization by genetic algorithm (GA). *Environ. Sci. Pollut. Res.* **2022**, *30*, 24927–24948. https://doi.org/10.1007/s11356-022-19683-0.
36. Tham, P.E.; Ng, Y.J.; Vadivelu, N.; et al. Sustainable smart photobioreactor for continuous cultivation of microalgae embedded with Internet of Things. *Bioresour. Technol.* **2022**, *346*, 126558. https://doi.org/10.1016/j.biortech.2021.126558.
37. Lee, J.S.; Sung, Y.J.; Sim, S.J. Kinetic analysis of microalgae cultivation utilizing 3D-printed real-time monitoring system reveals potential of biological CO_2 conversion. *Bioresour. Technol.* **2022**, *364*, 128014. https://doi.org/10.1016/j.biortech.2022.128014.
38. Zambon, I.; Cecchini, M.; Egidi, G.; et al. Revolution 4.0: Industry vs. Agriculture in a Future Development for SMEs. *Processes* **2019**, *7*, 36. https://doi.org/10.3390/pr7010036.
39. Giannino, F.; Esposito, S.; Diano, M.; et al. A predictive decision support system (DSS) for a microalgae production plant based on Internet of Things paradigm. *Concurr. Comput. Pract. Exp.* **2018**, *30*, e4476.
40. Otálora, P.; Guzmán, J.L.; Acién, F.G.; et al. Microalgae classification based on machine learning techniques. *Algal Res.* **2021**, *55*, 102256. https://doi.org/10.1016/j.algal.2021.102256.
41. Li, K.; Liu, Q.; Fang, F.; et al. Microalgae-based wastewater treatment for nutrients recovery: A review. *Bioresour. Technol.* **2019**, *291*, 121934. https://doi.org/10.1016/j.biortech.2019.121934.
42. Yang, C.-T.; Kristiani, E.; Leong, Y.K.; et al. Machine learning in microalgae biotechnology for sustainable biofuel production: Advancements, applications, and prospects. *Bioresour. Technol.* **2024**, *413*, 131549. https://doi.org/10.1016/j.biortech.2024.131549.
43. Lim, H.R.; Khoo, K.S.; Chia, W.Y.; et al. Smart microalgae farming with internet-of-things for sustainable agriculture. *Biotechnol. Adv.* **2022**, *57*, 107931. https://doi.org/10.1016/j.biotechadv.2022.107931.
44. Xu, W.-L.; Wang, Y.-J.; Wang, Y.-T.; et al. Application and innovation of artificial intelligence models in wastewater treatment. *J. Contam. Hydrol.* **2024**, *267*, 104426. https://doi.org/10.1016/j.jconhyd.2024.104426.
45. Beltramo, T.; Hitzmann, B. Evaluation of the linear and non-linear prediction models optimized with metaheuristics: Application to anaerobic digestion processes. Engineering in Agriculture*. Environ. Food* **2019**, *12*, 397–403. https://doi.org/10.1016/j.eaef.2019.06.001.