Science of the Total Environment 876 (2023) 162797

Contents lists available at ScienceDirect

Science of the Total Environment

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

Artificial intelligence and machine learning tools for high-performance microalgal wastewater treatment and algal biorefinery: A critical review

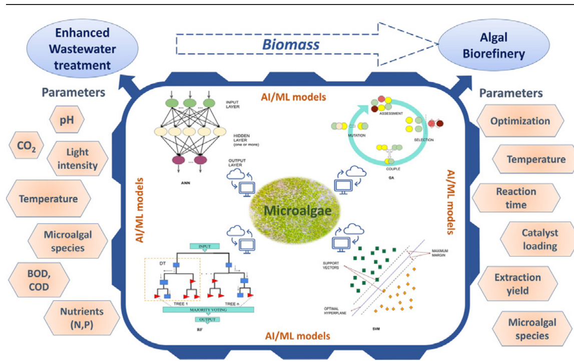
Raj Kumar Oruganti [a], Alka Pulimoottil Biji [a], Tiamenla Lanuyanger [a], Pau Loke Show [b], Malinee Sriariyanun [c], Venkata K.K. Upadhyayula [d], Venkataramana Gadhamshetty [e,f], Debraj Bhattacharyya [a,*]

[a] Department of Civil Engineering, Indian Institute of Technology Hyderabad, Kandi, Sangareddy 502284, Telangana, India
[b] Department of Chemical Engineering, Khalifa University, P.O. Box 127788, Abu Dhabi, United Arab Emirates
[c] Biorefinery and Process Automation Engineering Center, Department of Chemical and Process Engineering, The Sirindhorn Thai-German International Graduate School of Engineering, King Mongkut's University of Technology North Bangkok, Thailand
[d] Department of Chemistry, Umea University, Sweden
[e] Department of Civil and Environmental Engineering, South Dakota School of Mines and Technology, USA
[f] 2-Dimensional Materials for Biofilm Engineering Science and Technology (2D-BEST) Center, South Dakota Mines, Rapid City, SD 57701, USA

HIGHLIGHTS

- The microalgal contribution to the circular economy is substantial.
- AI/ML usage in algal cultivation assists in effective decision-making.
- Application of ML tools in algal biorefinery helps in increase in product yield.
- Novel deep-learning ML algorithms incorporating large databases are needed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Huu Hao Ngo

Keywords:
Artificial intelligence
Biorefinery
Machine learning
Microalgae
Wastewater treatment

ABSTRACT

The increased water scarcity, depletion of freshwater resources, and rising environmental awareness are stressing for the development of sustainable wastewater treatment processes. Microalgae-based wastewater treatment has resulted in a paradigm shift in our approach toward nutrient removal and simultaneous resource recovery from wastewater. Wastewater treatment and the generation of biofuels and bioproducts from microalgae can be coupled to promote the circular economy synergistically. A microalgal biorefinery transforms microalgal biomass into biofuels, bioactive chemicals, and biomaterials. The large-scale cultivation of microalgae is essential for the commercialization and industrialization of microalgae biorefinery. However, the inherent complexity of microalgal cultivation parameters regarding physiological and illumination parameters renders it challenging to facilitate a smooth and cost-effective operation. Artificial intelligence (AI)/machine learning algorithms (MLA) offer innovative strategies for assessing, predicting, and regulating uncertainties in algal wastewater treatment and biorefinery. The current study presents a critical review of the most promising AI/MLAs that demonstrate a potential to be applied in microalgal technologies. The most commonly used MLAs include artificial neural networks, support vector machine, genetic algorithms, decision tree, and random forest algorithms. Recent developments in AI have made it possible to combine cutting-edge techniques from AI research fields with microalgae for accurate analysis of large datasets. MLAs have been extensively studied for their potential in microalgae detection and classification. However, the ML application in microalgal industries, such as optimizing microalgae cultivation for increased biomass productivity, is still in its infancy. Incorporating

* Corresponding author.
E-mail address: debrajb@ce.iith.ac.in (D. Bhattacharyya).

http://dx.doi.org/10.1016/j.scitotenv.2023.162797
Received 24 December 2022; Received in revised form 23 February 2023; Accepted 7 March 2023
Available online 11 March 2023

smart AI/ML-enabled Internet of Things (IoT) based technologies can help the microalgal industries to operate effectively with minimum resources. Future research directions are also highlighted, and some of the challenges and perspectives of AI/ML are outlined. As the world is entering the digitalized industrial era, this review provides an insightful discussion about intelligent microalgal wastewater treatment and biorefinery for researchers in the field of microalgae.

Contents

1. Introduction

The current scarcity and demand for freshwater resources and the anticipated water stress due to ongoing climate changes highlight an immediate need for sustainable technologies that allow wastewater treatment and simultaneous resource recovery. Microalgae-based technologies offer a promising route for turning nutrients from wastewater into value-added biomass for subsequent resource recovery. Conventional methods of algal growing in high-rate algal ponds (HRAPs) and photobioreactors (PBRs) focus primarily on removing nutrients and organic matter from wastewater. However, in recent years, researchers have paid a lot more attention to wastewater treatment combined with resource recovery. This is because the generated biomass can be used to synthesize value-added bioproducts. Microalgae are unicellular microorganisms that leverage natural resources (e.g., carbon dioxide, sunlight, and nutrients) to meet their metabolic needs and, in turn, generate carbohydrates, lipids, and other useful products (Chen et al., 2022). Microalgae are regarded as the third generation of raw materials for biofuel production because of their carbon-neutral life cycle, non - competition with food or agricultural crops, and prospects for vertical and high-density culture system designs (Dragone et al., 2010). Microalgal biomass can be processed into fuels (biodiesel, hydrogen, and syngas) (Khoo et al., 2019), animal feedstocks, and platform chemicals (Wang et al., 2022). Recently, there has been a rise in interest in other areas of algal production, such as protein, phenolic compounds, and lutein (Sun et al., 2015), due to their repurposing potential. The major bottlenecks to the commercialization of micro-algal technologies lie in their high costs of harvesting and the technical skill needed to cultivate algae. The algal systems are highly sensitive to environmental perturbations. This necessitates constant monitoring and management of these systems to maintain optimal conditions. These recurring challenges can be addressed by embracing multidisciplinary, cutting-edge, automated, and smart technologies based on real-time monitoring.

Artificial intelligence (AI)/Machine learning algorithms (MLAs) enabled intelligent systems and dynamic models can effectively optimize the efficiency of algal systems. This is because the MLA approaches provide a more comprehensive insight into the uncertainty of biological processes than the conventional phenomenological or kinetic models (Sundui et al., 2021). AI/MLAs can be integrated seamlessly to effectively monitor, optimize, predict uncertainty, and discover faults in real-time in complex environmental systems. AI/MLAs can model the algal wastewater treatment and optimize process parameters for resource recovery. Due to their resilience and reliability, AI/MLAs have been widely employed for automating, predicting, and making decisions in complicated systems. They have been extensively used in the areas of real-time monitoring and data analysis. Artificial intelligence, or AI, is the process by which machines are given the ability to mimic human intelligence. The incorporation of AI requires a machine learning process. Machine learning is a subfield of AI. The goal of machine learning is to train a machine to solve a problem by itself using data obtained from various sources, including data collected over time and statistical analysis of that information (Jha et al., 2019). ML is a knowledge-acquisition and integration system that has been increasingly used in environmental domains like air pollution, wastewater, and solid waste treatments during the past few years. The fundamental principle of ML is to utilize inductive inference to generalize the relationships between input and output and then to use those generalizations to direct decision-making in novel contexts (Andrade Cruz et al., 2022). Fig. 1 depicts a typical workflow for an ML model, from gathering raw data to determining the optimal solution. A typical end-to-end ML process consists of three phases: training, cross-validation, and testing. During the training phase, the ML model is taught by adjusting various model parameters based on the training dataset. During the cross-validation phase, a validation dataset is used to fine-tune the model hyperparameters and identify the best model. When choosing MLAs, hyperparameter tuning is used to explore the best feasible solution in the least amount of time using fewer computational resources (Thornton et al., 2013). The selected optimum model is tested during the testing phase by measuring its results on a separate dataset. After that, the developed and optimized ML model can be utilized for prediction (Guo et al., 2021). ML has been renowned for its high prediction accuracy and its ability to save time and resources by reducing the need for repeated tests when applied to complicated non-linear domains (Singh

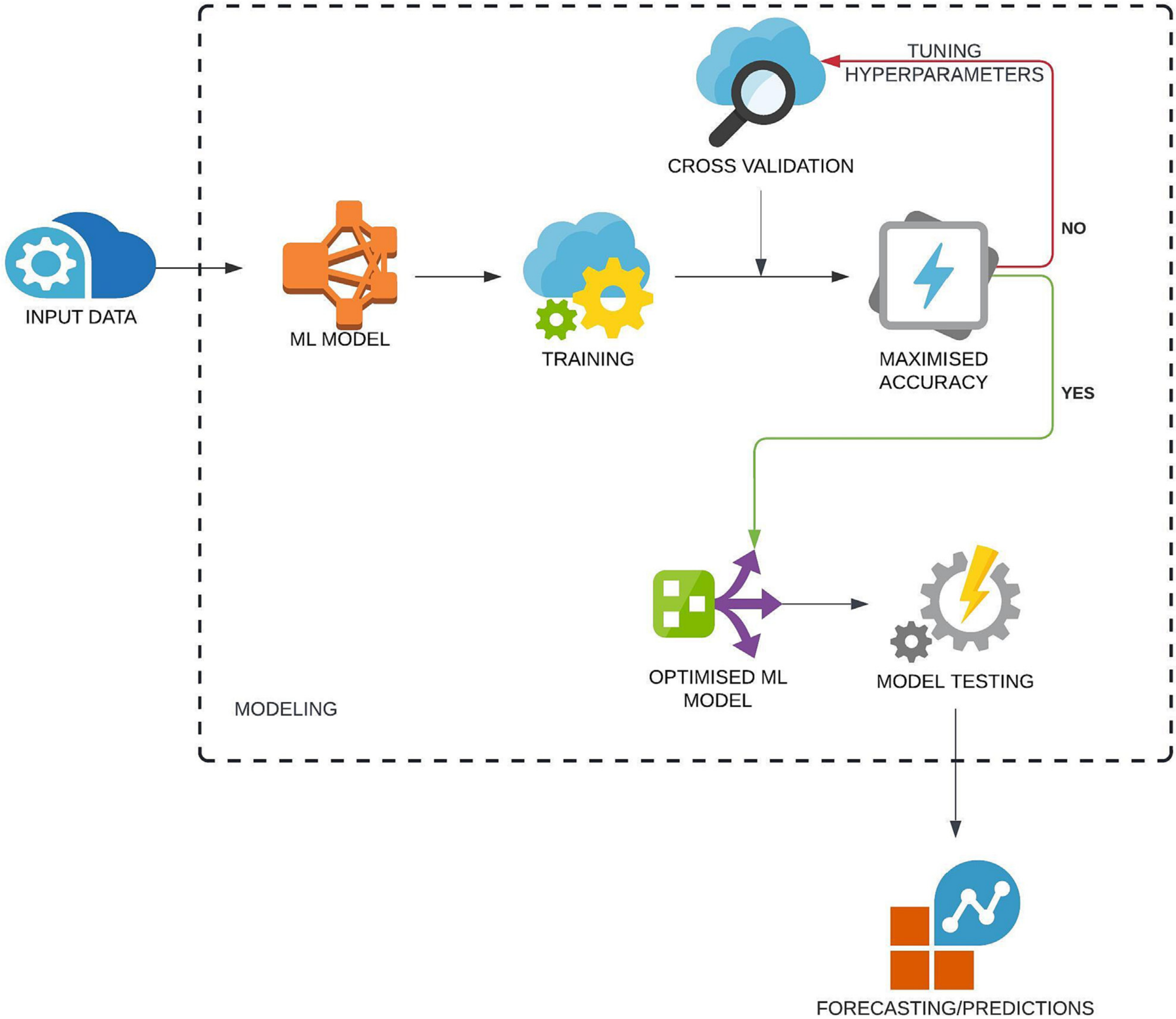


Fig. 1. Typical workflow of an ML model (Guo et al., 2021).

and Mishra, 2022). The optimal predictor variable combinations and significant patterns can be found by quickly analyzing the huge dataset using machine learning algorithms (Dallora et al., 2017).

The objective of this review is to provide an overview of the ways in which data-driven analytics, such as AI/MLAs, are being applied to microalgae research and development. Very limited studies are available that comprehensively discusses the fundamental aspects of AI/ML and their potential application in integrated microalgal wastewater treatment and resource recovery. This paper presents a critical review on the application of MLAs in microalgal wastewater treatment and optimization of the bioprocesses used in algal biorefinery based on the latest literature available. Also, we present a bibliometric analysis of the ML application in microalgal wastewater treatment. The most commonly used MLAs in the microalgal area were presented with their advantages, limitations, and applicability. The benefits of using MLAs for microalgae and the potential future prospects have been discussed. This review offers a fresh perspective for microalgal researchers and biorefinery industrialists about the potential of employing AI/ML enabled smart systems for efficient microalgal cultivation and resource recovery.

2. Bibliometric analysis through ML model

Scopus database has been used to gather the publication data on AI/MLAs application in microalgal research in the last ten years, from 2012 to 2022. The keywords used for the data collection are “microalgae” and “artificial intelligence” or “machine learning”. The Scopus database resulted in 83 research articles for the selected keywords. Fig. 2(a) & (b) shows the research trend of the articles published on this topic during the last decade and country-wise contribution. From the figure, it can be observed that there has been significant growth in the number of publications in the last couple of years. This can be attributed to the increased interest among researchers to employ AI/ML in microalgal applications. The countries at the forefront of research on this topic are China, US, India, and Malaysia. The collected database of articles is interpreted and clustered using VOSviewer software (version 1.6.18). The keywords co-occurrence tool of the software was used to visualize the network between the most recurrent keywords of AI/ML and microalgae. During the analysis, the minimum number of repetitions of keywords was limited to three. The obtained co-occurrence network map is shown in Fig. 3. The size of the circle increases with an increase in the frequency with which the keywords are featured in the titles and abstracts of the articles. Therefore, the circle size for a given keyword is proportional to its occurrence. The keywords “microalgae”, “machine learning” and “artificial intelligence” have occupied the central position on the map, having a total signal strength of 534, 481, and 168, respectively. Subsequently, the two most frequently used keywords were “wastewater treatment” and “biomass production,” with a total signal strength of 164 and 152, respectively. This is because, in most cases of the employed AI/ML models, the output variables are

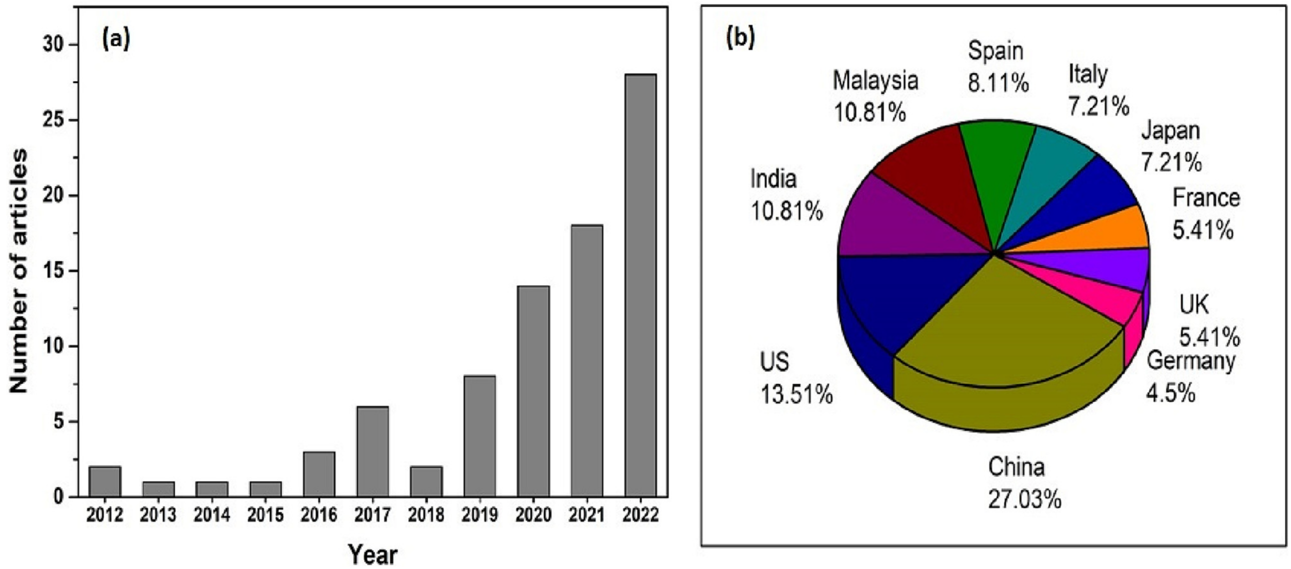


Fig. 2. (a) Number of articles published in last decade (b) Country wise list of articles (Source: Scopus database search dated 17/11/2022; keywords – "microalgae" and "machine learning" or "artificial intelligence").

biomass productivity and enhanced wastewater treatment efficiency. *Chlorella* species also appeared in the network map because it is extensively studied for its high wastewater treatment capability, biomass productivity, and high lipid content (39–42 %), which is suitable for biodiesel synthesis (Miranda et al., 2022). Some of the most commonly used phrases and topics in the gathered literature include ML algorithms such as neural networks, decision trees, random forest, support vector machines, and typical input/output variables like pH, inoculum, carbon dioxide level, biomass production, biofuels, optimization, wastewater treatment, etc.

3. Different AI/ML algorithms

3.1. Artificial neural networks (ANN)

ANNs are a set of algorithms that are part of machine learning. ANNs mimic the behavior of the human brain and how learning occurs in a human. ANNs are black-box models that employ the gradient descent backpropagation technique to predict or anticipate a target output. As illustrated in Fig. 4(a), they are composed of a number of nodes that are

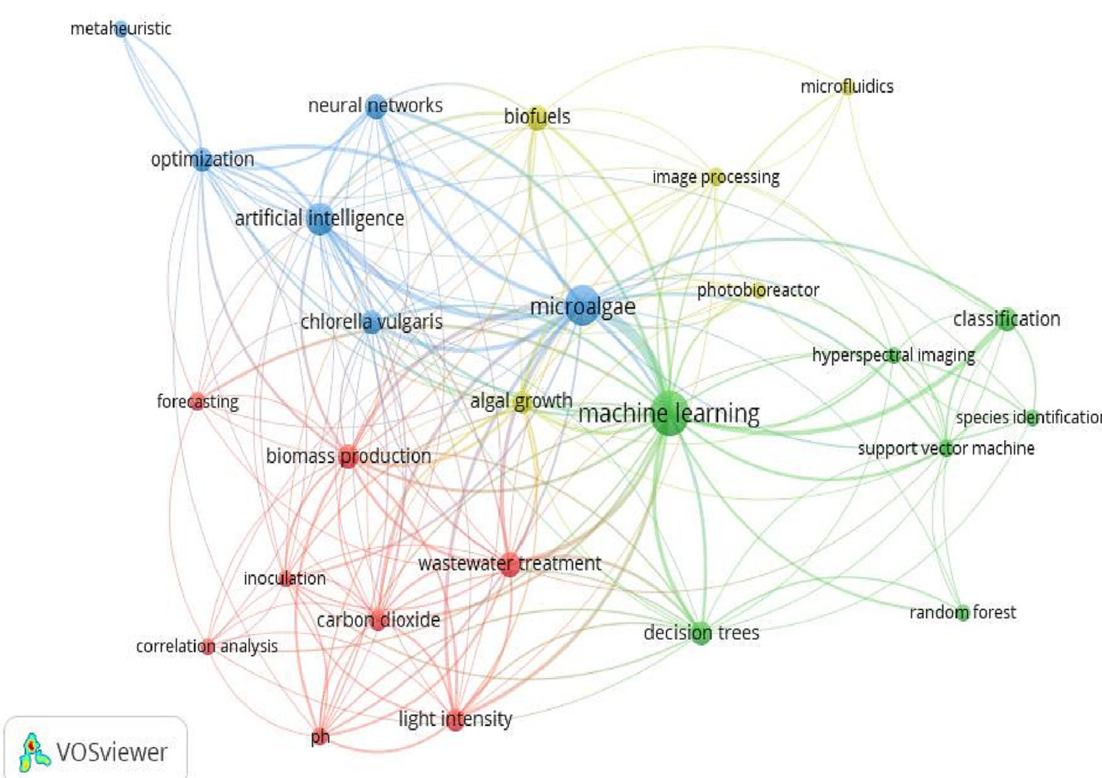


Fig. 3. Keyword co-occurrence network map of AI/ML application in microalgae.

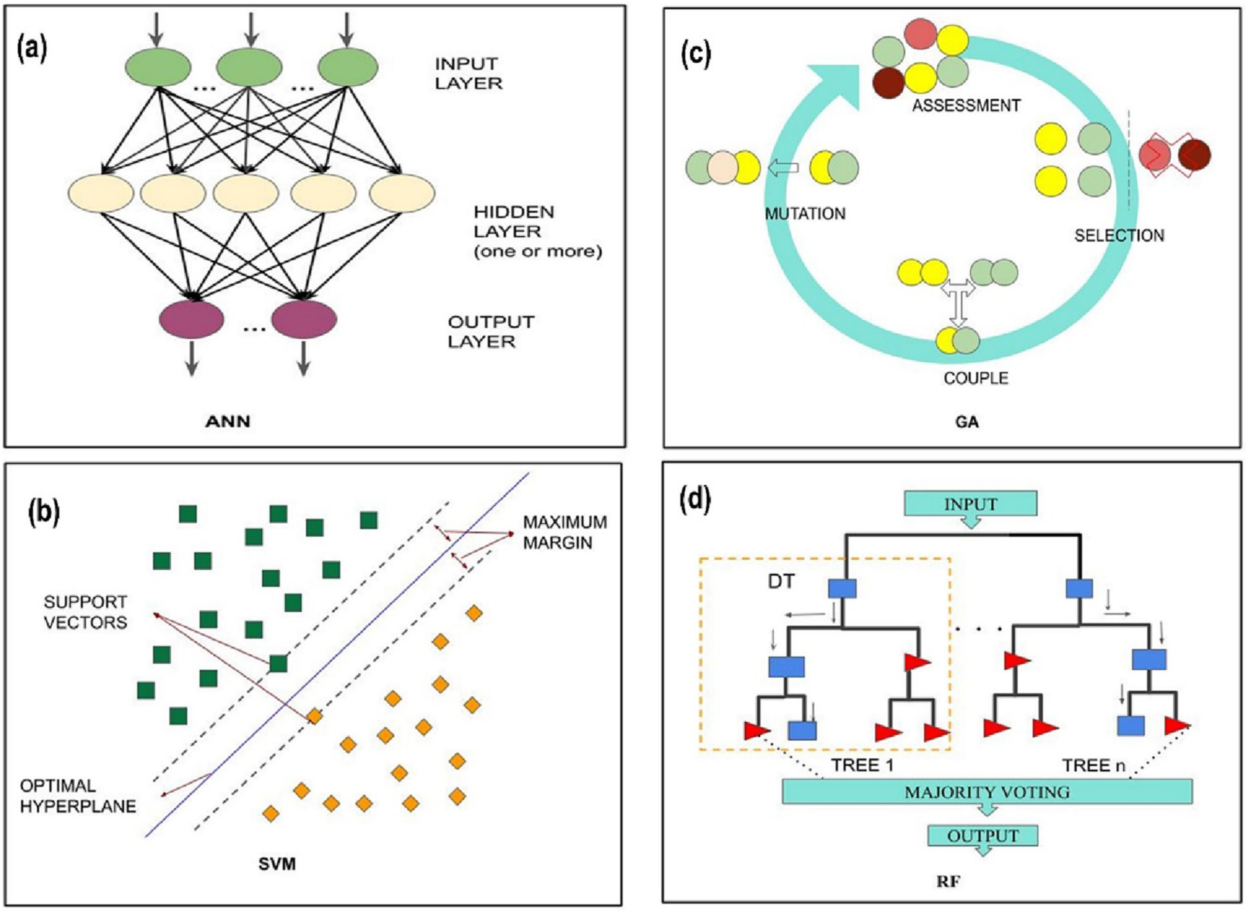


Fig. 4. Schematic representation of different AI/MLAs (a) Artificial neural networks (ANN) (b) Support vector machine (SVM) (c) Genetic algorithm (GA) (d) Decision tree (DT)/Random Forest (RF) algorithm.

structured into layers, with each node having a set of outputs and inputs. For a given layer, the input is from the output of the previous layer and provides an output function. After this, the node of the next layer gets this output function as the input and so forth until the final layer is achieved (Schmidhuber, 2015). From the information processing point of view, an ANN is an operational model based on the human brain neuron network that is non-linear, unconstrained, incredibly adaptive, and fault-tolerant (Guo et al., 2021). Fig. 4(a) depicts a simplified example of a feed-forward network, a type of ANN consisting of a fully linked multilayer network that allows signals to travel only in one direction, from input to output. Single-layer ANNs have the input layer directly connected to the output layer, while multilayer ANNs have many hidden levels between the input and the output (Andrade Cruz et al., 2022). The learning technique used to train ANN is dependent on the mode in which the neural networks were constructed. The success of ANNs depends heavily on the network size, as this determines the complexity of the data processing, network training, and optimal network architecture decisions. However, increasing the number of hidden neurons beyond a certain point is not necessarily connected with improved learning (Faris et al., 2019). Since overfitting and underfitting can occur with any size of data used for training, limiting the model generalization ability is one of the most difficult aspects of building an ANN model. If more fitting information is needed, the number of hidden layers could be raised, leading to deeper learning as the calculation process is stretched to accommodate the new data. However, as the number of neurons increases, overfitting may occur, in which the error can be smaller during the training phase but significant in the testing phase (Alrashed et al., 2018). Moreover, on the other hand, very few neurons could hinder the ANNs capability to represent the mechanism leading to overgeneralization.

ANNs have been widely employed in algal systems by several researchers. Accurate performance predictions in algal systems necessitate appropriate input-output mapping (Sundui et al., 2021). In a neural network model, the number of neurons in the input layer is directly proportional to the number of process parameters utilized during training to make predictions about the external environment. In algal systems, environmental factors such as pH, CO_2, illumination, temperature, hydraulic retention time (HRT), DO, TSS, nutrient levels, COD, etc., are considered the input parameters for an ANN model. The combination of these variables could affect the algal biomass growth and treatment efficacy. The hidden layer consists of a suitable activation function. The output layer is part of the network that communicates with the outside world, and its neuron count is typically correlated with the system's performance. For an algal system, the output layer can be represented by the neurons measuring the treatment efficacy parameters or the biomass production/growth. For example, Supriyanto et al. (2018) used a multilayer backpropagation neural network ANN model for predicting microalgal growth in an open raceway pond. The model considered the eight input parameters (initial biomass concentration, HRT, harvesting period, solar radiation, acetate and nitrate concentrations, pH, and temperature), one hidden layer, and the output layer as the algal biomass on a dry basis. Results showed that the constructed three-layer ANN achieved high prediction accuracy ($R^2 > 0.93$) across all the input configurations used.

However, since ANN is a “black box” model, it can sometimes produce unexpected results (Kannangara et al., 2018). The disadvantage of neural networks is their lack of ability to provide an explanation for their reasoning. They cannot be used for the mechanistic understanding of the process because they use only empirical data rather than comprehending the reason behind the change (Guo et al., 2021). Though ANN models are more accurate for kinetic modeling than conventional models, significant biological factors such as kinetic constants cannot be studied with ML. Since ANNs converge slowly and tend to settle for local optima during training, this makes them inappropriate for sparse data modeling (Dong and Chen, 2019). It is important to consider these constraints when working with ANN models.

3.2. Support vector machine

SVM is an advanced supervised ML technique for regression and classification problems, first reported by Cortes and Vapnik (1995), and is now considered one of the most effective ML tools available. The SVM algorithm employs a nonlinear mapping function to transform data into a high-dimensional feature space. The SVM algorithm creates a decision boundary called a hyperplane that segregates multi-dimensional space into different classes so that the unknown data point can be classified accordingly. Support vectors are used to find the hyperplane that optimizes the distance between itself and the closest data sample in a high-dimensional space. Support vectors are the data points that lie along the optimal hyperplane, as depicted in Fig. 4(b). The kernel technique used by SVM to achieve optimal data separation and its adoption of the structural risk minimization (SRM) concept, as opposed to the empirical risk minimization (ERM) principle used by ANN, make it less prone to overfitting (Andrade Cruz et al., 2022). Several researchers have used the SVM algorithm in microalgal classification and wastewater treatment. For example, in a recent study, Hossain et al. (2022) employed the SVM algorithm to predict the nitrogen and phosphorus removal in municipal wastewater treatment using microalgae. They have reported that SVM performed better compared to multilayer perceptron artificial neural network (MLP-ANN) and response surface methodology (RSM) for the same task. However, SVM has a low training efficiency when working with large volumes of data and is susceptible to missing values. Therefore, when employing SVM, it is crucial to pay close attention to factors such as kernel function, data size, and missing values.

3.3. Genetic algorithm (GA)

GA is an optimization model, which is based on the concept of natural selection and the principle of genetic evolution through the processes such as selection, crossover, and mutation (Guo et al., 2021). In the first stage, a selector operator is used on the population to pass on the best individuals to the next generation by measuring their fitness. The following step involves recombining the genes from both parents with a crossover operator to produce a new set of genes. The mutation operator is then employed to alter genes in some population members randomly. Eventually, these three steps are iterated until the optimal solution is achieved. Fig. 4 (c) shows the schematic representation of the GA process. The advantage of employing this model is that the algorithm tries to find the optimal solution by avoiding the local minimum. It was reported that GA performs better in solving combinatorial optimization problems compared to other models. GA has been used to optimize microalgal growth and resource recovery parameters. For example, Camacho-Rodríguez et al. (2015) employed GA for optimizing the composition of culture media for the microalgal species *Nannochloropsis gaditana* cultivation. This optimization helped in reducing the several nutrients requirement drastically. The main advantage of GA is its extendibility, which can be integrated with other models. Rodríguez-Miranda et al. (2021) have employed GA for modeling the effect of temperature on the microalgal culture in a large-scale raceway pond. Several studies have integrated GA with ANN and SVM for increased model accuracy. For instance, Nayak et al. (2018) have combined ANN with GA for optimizing the process parameters for treating domestic wastewater using the microalgal *Scenedesmus* sp. They have reported an enhanced biomass productivity of 57 % at the model-predicted conditions. Very recently, Kushwaha et al. (2022) have used a hybrid technique of integrating an adaptive neuro-fuzzy inference system (ANFIS) with GA for estimating microalgal CO_2 fixation. They have reported that the ANFIS-GA model showed an increased prediction capability than that of the ANFIS model alone. The main disadvantage of the GA is premature convergence, which can be caused by various factors such as selection operators, crossover, population size, and incorrect code. Also, GAs are computationally very intensive programs; thus, they require more resources.

3.4. Decision tree (DT) and random forest (RF) algorithms

DT and RF algorithms are the supervised learning MLAs, which are extensively used in both classification and regression tasks. DT process the data based on a concept tree that summarizes the given information of the training data set into a tree structure. DT uses binary discretization to iteratively separate the information into smaller sub-datasets with the goal of reducing variability within subsets before building a tree structure to carry out the regression function or classification (Guo et al., 2021). The advantage of DT is that it is quick to train, has an easy interpretation, and can model a high degree nonlinear relationship between the input and output variables. However, as the quantity of the dataset grows, DT is more likely to overfit since it builds larger, more complicated trees. RF is an ensemble learning model, which uses bagging, and random feature selection features to build multiple uncorrelated decision tree structures whose predictions are subsequently averaged (Bagherzadeh et al., 2021). RF model selects a random set of features to construct the DTs instead of training different DTs with the same dataset. RF exhibits a level of differentiation among all trees by selecting a split feature at every node of the decision tree. RF can successfully mitigate the threat of overfitting within the context of an ensemble learning approach (You et al., 2017). Fig. 4 (d) depicts the structure of both DT and RF. DT can quickly learn from continuous and discrete data with little or no preprocessing. The learning rate of RF is significantly slower than that of DT. However, it overcomes some of the shortcomings of DT, including its inability to handle nonlinear data and its high overfitting risk (Andrade Cruz et al., 2022). Several researchers have employed DT and RF algorithms for making predictions in microalgal cultivation and bioproduct extraction. For instance, in a recent study, Singh and Mishra (2022) used the DT algorithm for the various combinations of cultivation parameters for wastewater treatment using the microalgal class of *Trebouxiophyceae* and *Chlorophyceae*. They reported that the initial nitrogen concentration and the initial biomass inoculum levels have considerable effect on the biomass productivity of the class of studied microalgae. They have also suggested the suitable combination of parameters required for enhanced wastewater treatment and biomass productivity for both classes of microalgae. Similar to this, in their previous study Singh and Mishra (2021) have employed DT algorithm to identify the predictor variables such as microalgal class, factors influencing the cultivation and operating variables, which could result in high biomass productivity along with significant wastewater treatment potential. Reimann et al. (2020) have reported that among the explored MLAs, the RF algorithm performed better in classifying dead or alive microalgal populations of *Chlorella vulgaris* culture. Zhang et al. (2021) employed an RF algorithm to optimize bio-oil yield from microalgae hydrothermal liquefaction.

From these reported literature, it could be understood that the ML algorithms can swiftly interpret the large datasets associated with microalgae and identify the most efficient predictor variable combinations that could be employed while developing new experimental trials. It can help in developing more advanced microalgal wastewater treatment along with large-scale biomass production. The following sections provide an in-detail discussion regarding the usage of different AI/MLAs in different stages of microalgal cultivation, wastewater treatment, and bioproducts extraction.

4. AI/MLAs in microalgal wastewater treatment

Microalgae are a type of aquatic photosynthetic microbial organism, which uses carbon dioxide as a carbon source for photosynthesis, and in return, they generate organic molecules like proteins, lipids, and carbohydrates (Chen et al., 2022). Due to its vast potential uses in industries like food, pharmaceuticals, cosmetics, and bioenergy generation, cultivating microalgae on a large scale at a reasonable cost (Khoo et al., 2019). Optimizing resource input for producing sufficient biomass with predictable quality is a necessary first step in developing microalgae-related products and derivatives. The key to establishing a cost-effective microalgae production system is achieving a high biomass yield. Multiple parameters, like illumination, temperature, availability of nutrients, pH, aeration, and fluid mechanics modulation, contribute to microalgal productivity (Li et al., 2019). An increase in temperature, for instance, will trigger modifications to cell metabolism, including growth and respiration, and so play a critical part in microalgae growth. Additionally, photoinhibition caused by either extremely low or extremely high temperatures can impede microalgal photosynthesis. This means that it is essential to regulate the temperature and anticipate and prevent dangerous situations like a rise in the surface temperature. Direct indications that can help in monitoring the dynamics of culture include the size, morphology, cell density, community structure, pigment, and lipid content of microalgal cultivation systems. Real-time continuous control and monitoring of these factors are difficult since the methods employed to measure them are either too time-consuming, too arduous, or too destructive to microalgae. For example, low cell density cultures present difficulties for photoautotrophic microalgae cultivation, resulting in low yield. Therefore, a high-density cultivation technology is a crucial step in achieving the goal of commercializing microalgae biomass production. Researching microalgal metabolic versatility and developing effective bioreactors for growing microalgae at a high cell density are the two most common approaches.

4.1. AI/ML for microalgal cultivation

Recently microalgae have been employed as a sustainable alternative wastewater treatment due to their high nutrient removal capacity and biomass as the resource for producing valuable products. Several researchers have employed ML algorithms in predicting the best combinations of the variables for process optimization. The type of algal strain employed, and the cultivation conditions provided have a considerable impact on the microalgal wastewater treatment. The cultivation parameters such as temperature, inoculum ratio, pH, type of reactor, light intensity, CO_2, and nutrient levels are highly dependent and specific to the species of microalgae employed. Hence, it is very important to provide optimum growth conditions suitable for the particular strain to achieve maximum treatment efficiency along with high biomass productivity. MLAs can be employed for the selection of appropriate strains for wastewater treatment and optimum parameters for cultivation. Because microalgae development is intermittent, optimizing its cultivation is a non-direct process that typically necessitates dynamic optimization strategies (He et al., 2012). Despite having multifaceted applications, microalgae cultivation is a complex process to understand because of its biological nature. There are a lot of uncontrollable factors, such as culture conditions and reactor design, that could affect biomass productivity. Conventionally, CFD is used for hydrodynamics and shape optimization in photobioreactors. Physical models and correlations can be used to account for the impact of factors like temperature and light intensity. Kinetic equations, such as Monod-based models, are commonly used to model microalgal mass increase and transfer (He et al., 2012). The integration of multiple complicated biological processes is a challenge that must be met by conventional mathematical modeling. Conversely, ML has evolved as a data-driven technique that does not require the complicated relationships inherent in mathematical models (Wang et al., 2020). Indeed, all the information used in these models comes from public sources such as the internet or recorded versions of the procedure from the past. Even though ML algorithms can deal with missing data, manage multivariate data, and forecast nonlinear connections, it is still crucial to select the best algorithm for a given problem (Alzubi et al., 2018).

AI/ML models can be used to model and predict biomass productivity. Ansari et al. (2021) have employed a three-layer feed-forward back propagation ANN model for the prediction of algal dry cell weight in raceway ponds treating secondary wastewater under natural illumination supplemented with nutrients. Nayak et al. (2018) developed an ANN-Genetic Algorithm (GA) model to optimize microalgae strain *Scenedesmus* cultivation parameters such as photoperiod, light intensity, initial pH, and temperature. They have successfully reported a 57 % increase in biomass productivity at the ANN-optimized conditions. Furthermore, in a recent study, Hossain et al. (2022), have used soft computing techniques such as multilayer perception artificial neural network (MLP-ANN), RSM, and SVM for identifying the effect of operational parameters, including N:P ratio, light-dark cycle, and temperature on the municipal wastewater treatment using microalgal species *Chlorella kessleri*. They have reported that the SVR-GA hybridized model performance was better than the RSM and MLP-ANN model in simultaneous prediction of nitrogen and phosphorus removal efficiencies. Coşgun et al. (2021) employed the DT algorithm to study the effect of crucial parameters such as microalgal species, growth settings, CO_2 levels, type of reactor, nutrient conditions, and lipid extraction methods. Morowvat and Ghasemi (2016) have used ANN for optimizing growth culture medium constituents for maximizing lipid accumulation during the cultivation of *Chlorella vulgaris*. The concentrations of nutrients (nitrate and phosphate) and glucose in the growth medium served as the model input parameters. Susanna et al. (2019) were able to predict the growth of *Spirulina* and forecast the algal productivity three days in advance by using a nonlinear autoregressive multilayer perception ANN model. The input variables for the model are the initial biomass level, nitrogen, dissolved oxygen, and time while the output was the biomass concentration.

Microalgal culturing is not only affected by the macronutrients such as carbon, nitrogen, and phosphorus but also is affected by the presence and abundance of the micronutrients such as vitamins and trace elements in culture media. However, most of the existing models are developed by considering mainly macronutrients only. There are very limited studies that considered the effect of micronutrients in microalgal models. For example, López-Rosales et al. (2013) have employed the ANN algorithm Feed-forward back-propagation neural networks (FBN) to represent the nonlinear interactions between the 26 different components of the culture medium. They have used the data from a data set of 500 batch culture experiments of growing microalgal species *Protoceratium reticulatum.* Garson's algorithm has been used for assessing the importance of the components on algal growth. They reported that micronutrients and vitamins were more important (>70 %) than macronutrients (only 25 %), regardless of the fact that their concentrations in culture media are relatively low in magnitude when compared to macronutrients. Similarly, García-Camacho et al. (2016) also used the FBN algorithm for estimating the algal growth dynamics of *Karlodinium veneficumin*, in a culture medium with 25 different compositions of key nutrients. They used the data from 420 batch culture experiments as the input and considered growth profiles as the output. Similar to the earlier study, they have also reported that the combined effect of micronutrients and vitamins on microalgal growth was greater than that of the macronutrients. Banerjee et al. (2016) employed GA to optimize the environmental parameters of nutrients, light intensity, NaCl, and $NaHCO_3$ for maximizing the lipid productivity in *Nannochloropsis* sp. cultivation. In another study, heterotrophic cultivation of *Chlorella* species using glucose as the substrate was modeled using a hybrid neural network model (Wu and Shi, 2007). The glucose concentration was employed as an input variable in the model, and the predicted output was the microalgal-specific growth rate. They demonstrated that a single-input (glucose concentration) hybrid neural network model could provide a good approximation of the experimental data, implying that it might be used as a sufficient tool for optimizing heterotroph growth.

Light intensity is one of the significant factors affecting microalgal growth. Hu et al. (2008) have employed Artificial Neural Network–Model Predictive Control (ANN-MPC) for automatically controlling the green microalgae *Spirulina platensis* cultivation in a continuous algal photobioreactor. The developed ANN-MPC controller could intellectually learn the dynamics of the reactor and self-adaptively regulate the light intensity to facilitate microalgal growth. Furthermore, Del Rio-Chanona et al. (2019) investigated the dynamic effect of light intensity and nitrate content on lutein production from microalgae and developed a model for enhancing lutein production using data-driven modeling methods. They have reported that lutein yields can be increased by 40–50 % using the cultivation strategies from the data-driven models but only by 35 % using one of the physical-based model cultivation techniques.

To effectively cultivate microalgae, conducting a preliminary screening of strains is necessary before moving on to the construction of scaled-up reactors (Chen et al., 2011). This allows for the development of a microalgae-based process specifically suited to the needs of the industrial sector. The application of the AI/MLA models for screening, classifying, and estimating algal cell populations is presented in the following Section 4.2.

4.2. AI/ML-assisted screening and classification of microalgae

Biotechnological production employing microalgae relies heavily on suitable species and strain selection (Lim et al., 2022). The strains selection for cultivation is dependent on the commercial objectives demanded by the industries based on the applications such as feed, nutrition, biodiesel, etc. The parameters such as growth rate, biomass production, high-value product content, and pollutant removal capacity are employed to rank the strains. To ensure long-term sustainability and cost-effective biotechnology production, species fit for commercial use should have high biomass production (Christenson and Sims, 2011). Microalgal identification and classification can be a time-consuming and costly process due to the high diversity of microalgal species in even a small amount of water samples. As a result, researchers have recently started employing a machine-learning strategy for microalgae identification. Otálora et al. (2021) have developed two ANN-based models for the identification of the microalgal species *Chlorella vulgaris* and *Scenedesmus almeriensis* in a mixed composition. The ANN models were coupled with the FlowCAM, an instrument that captures the particles in the sample and gives a set of descriptive features for each particle. Pure samples of each species were used to train the models and validated with the mixed culture. The findings show the benefits of image analysis employing deep learning for microalgal culture classification. Drews- et al. (2013) have used the Gaussian mixed semi-supervised classification model and active learning to classify the microalgae using the data obtained from the FlowCAM. They have used microalgal features such as diameter, length, width, aspect ratio, etc., as the input data. The proposed model was able to perform better than the supervised classification algorithm SVM and achieved 92 % accuracy. Furthermore, Harmon et al. (2020) employed an SVM model to accurately classify six different microalgal species using the data obtained from frequency division multiplexed fluorescence imaging flow cytometry. ANN also can be used for microalgal classification. For instance, Liu et al. (2020) have coupled single-excitation fluorescence spectroscopy and back propagation neural network model optimized by genetic algorithm (BP-GA) for accurately monitoring algal cell concentration in *Chlamydomonas reinhardtii* culture. The input to the model was fluorescence emission spectral data, while algal concentration was considered the model output. The same study reported that the GA-optimized BP network model outperformed the conventional BP network prediction model. This is because the genetic algorithm uses selection, crossover, and mutation processes to identify the ideal values for the initial weights and thresholds of a BP neural network in order to decrease the network's prediction error (Wang et al., 2020). This implies that coupling conventional prediction models with suitable optimization algorithms results in increased model performance and accuracy. The microalgal growth is conventionally estimated by measuring the chlorophyll content. However, this approach is sometimes erroneous because of the overlapping of the spectrum of other pigments such as carotenoids with the chlorophyll. In a recent study, Ying Ying Tang et al. (2023) employed linear regression (LR) and ANN-Multilayer perceptron algorithm to predict the chlorophyll from the color models. The ANN model outperformed the conventional spectroscopy and the linear regression model in accurately estimating the chlorophyll content.

Moreover, understanding the different symbiotic and competitive interactions between the different species of microalgal is also very important in microalgal cultivation. For instance, Bi et al. (2019) have employed hyperspectral imaging technology coupled with the SVM model to study the survival traits of the three different classes of microalgae under different pH conditions. To determine which microalgal species were most prevalent, they employed an SVM classifier trained using fluorescence and transmission spectra. These methods can also help understand the symbiosis process between diverse microalgal species and microalgae-microbial communities, which is crucial for advancing wastewater treatment. Photosynthetic pigments in microalgae generate unique spectral signatures for each strain, allowing for their classification into different classes. So, the light absorption spectra of various types of diatoms, red/green/brown microalgae, and cyanobacteria can be distinguished from one another based on the relative concentration of chlorophyll, carotenoids, and other pigments (Serive et al., 2017). For example, Franco et al. (2019) have classified four different microalgal species, *Spirulina platensis, Scenedesmus almeriensis, Chlorella vulgaris,* and *Nostoc* sp., in both mixed and monoalgal cultures. They have used to classify the microalgae using the data obtained from the light absorption measurements from different microalgae species. Using the ANN and SVM as a classification approach of pattern recognition, Giraldo-Zuluaga et al. (2018) developed a methodology for the automated identification of *Scenedesmus* coenobia using microscopic image processing. Accuracy levels of 98.63 % and 97.32 % were achieved in identifying *Scenedesmus* coenobia, for SVM and ANN, respectively. Implementing ML models for the classification and identification of microalgal species would be extremely beneficial for the microalgal biorefineries searching for microalgal strains with better traits and capabilities for attaining their specific designated applications. The production of microalgae is typically done on a massive scale, and in certain circumstances, several different species of algae are cultivated, making monitoring and control challenging. For microalgal cultivation, it is necessary to establish the basic cell concentration parameters. Traditional offline techniques of determination require a lot of time and effort, and environmental factors might influence them. The incorporation of a microscopy device included with an image processing algorithm for real-time analysis of microalgae cultivation can drastically reduce the labor and time and cost of obtaining the necessary data (such as size, morphology, and cell count) in the photobioreactor. Table 1 shows the summary of the different AI/MLAs along with their potential application in microalgal cultivation, classification, and algal wastewater treatment.

5. AI/ML in microalgal biorefinery

The concept of biorefinery revolves around the sustainable conversion of biomass into a wide variety of bio-based products and bioenergy. Biomass energy (bioenergy) is an important renewable energy source with the potential to replace petroleum since it can be used to produce liquid fuel. The effective conversion of biomass into energy, fuels, and bioproducts may be attained by a number of different processes. Numerous studies have explored the microalgal utilization for several bioproducts extraction. Microalgal biorefinery entails both stages of upstream processing (USP) and downstream processing (DSP) (Chew et al., 2017). USP is associated with microalgal cultivation, where several factors could affect microalgal growth. The dominant factors that could affect microalgal growth are microalgal strain, nutrient sources, carbon dioxide (CO_2) supply, and illumination (Cheah et al., 2018). DSP includes the extraction and purification of important bioproducts from microalgal biomass and the systems and techniques employed in those processes (Lee et al., 2021). Biochemical and thermochemical conversions are the two prominent ways of biomass conversion (Tang et al., 2020). There are a lot of promising routes for converting algal biomass into bio-oil and other alternative fuels. For instance, manufacturing bio-oil from algae using hydrothermal liquefaction (HTL) is preferable as it saves energy by eliminating the need for the energy-intensive biomass drying phase. Because of its capacity to handle a variety of wet biomass feedstocks, in recent years, hydrothermal liquefaction has emerged as a central focus for biorefinery development. However, the need for time-consuming and expensive tests makes it

Table 1
AI/ML tools for Microalgal cultivation/classification and wastewater treatment.

Application	Employed AI/ML algorithms	Details	Remarks	Reference
Microalgal cultivation	Decision tree algorithm	*Trebouxiophyceae* and *Chlorophyceae* class microalgae; MATLAB; Classification and Regression Trees algorithm	The combinations of cultivation parameters were determined for enhanced wastewater treatment for two different microalgae classes using literature data. The study suggested suitable optimum cultivation parameters for each class.	(Singh and Mishra, 2022)
Real-time monitoring and predicting	Long short-term memory (LSTM) neural network	Microbial potentiometric sensor (MPS); algal cultivation pond	The developed AI model was trained from the MPS data and used to predict the real-time water quality parameters and algal concentrations	(Saboe et al., 2021)
Microalgal cultivation	Association rule mining (ARM)	Biomass productivity and nutrient removal efficiency; Data mining; MATLAB	ARM was used to determine the specific cultivation conditions for the increased wastewater treatment using the 500 data entries from the literature	(Singh and Mishra, 2022)
Optimizing photobioreactor (PBR) configuration	Convolutional Neural Networks (CNN)	Data-driven surrogate model; Computational fluid dynamics (CFD); kinetic model; hybrid stochastic optimization algorithm	Developed a surrogate modeling framework coupling CFD & kinetic model with CNN to reduce the computational resources for optimizing parameters for a pilot scale PBR	(del Rio-Chanona et al., 2019)
Predicting nutrients (N & P) removal	Multilayer perceptron artificial neural network (MLP-ANN) and Support vector regression (SVR)	*Chlorella kessleri*; municipal wastewater treatment; Genetic algorithm (GA) hybridization; Operation parameters (temperature, light-dark cycle; nitrate: phosphate N:P ratio)	Identified the effect of operational parameters on nutrient removal. Among the explored models, SVM with GA hybridization showed maximum nutrient removal efficiencies of >93 %, at 29.3 °C, 24 h:0 h light-darks cycle and N:P of 6:1	(Hossain et al., 2022)
Wastewater treatment	Artificial neural network (ANN) modeling and forensic-based investigation algorithm (FBI)	Microalgae microbial fuel cell (MMFC); Decision variables: wastewater concentration (%) and yeast concentration (%)	Employed ANN for optimizing the variables for maximum power density ($R^2 \sim 0.9783$) and COD removal ($R^2 \sim 0.9$) in MMFC.	(Sayed et al., 2023)
Optimizing process parameters for CO_2 sequestration	Artificial neural network coupled with Genetic algorithm (GA)	*Scenedesmus* sp.; domestic wastewater; coal-fired flue gas	ANN-GA model was used to optimize the process parameters, photoperiod, light intensity, initial pH, and temperature. The optimized parameters enhanced biomass productivity by 57 %	(Nayak et al., 2018)
Microalgal classification	Artificial neural networks	Scenedesmus almeriensis and *Chlorella vulgaris*; MATLAB; Deep learning toolbox	Employed two ANN models to classify the mixed microalgal sample. FlowCAM data and captured image data were used to train the ANN model	(Otálora et al., 2021)
	Support vector machine (SVM) classifier	Hyperspectral imaging technology; effect of pH; survival competition; particle swarm optimization algorithm (PSO)	Model was used to predict the effect of pH on the survival traits in a mixed culture of the three different microalgal species using hyperspectral microscopic images	(Bi et al., 2019)
	Genetic algorithm optimized Back propagation neural network model (GA-BP)	*Chlamydomonas reinhardtii*; algal cell concentration; fluorescence emission spectra; MATLAB	GA-BP model was used to assess the algal cell concentrations using the data from fluorescence emission spectra	(Liu et al., 2020)
	Linear Regression (LR) and Artificial neural network (ANN)	Predicting chlorophyll concentration from color models; *Desmodesmus* sp. and *Scenedesmus* sp.; image processing; solvent extraction	LR ($R^2 \sim 0.58$) and ANN ($R^2 \sim 0.66$) models were used to estimate the chlorophyll concentration using image processing color models	(Ying Ying Tang et al., 2023)

challenging to evaluate the quantitative and qualitative features of hydrothermal liquefaction (by)products (Shafizadeh et al., 2022). Recently, there has been an increased application of AI/MLAs in optimizing process parameters for biofuel production. For instance, Zhang et al. (2021) employed ML algorithms gradient boosting regression (GBR) and (RF) to predict and optimize the bio-oil production from the algal biomass. The model inputs were the algal composition and HTL conditions, whereas the model outputs were the bio-oil yield, nitrogen, and oxygen content of oil. The GBR model was observed to perform better for single-task and multi-task predictions, with $R^2 > 0.9$. Very recently, Sultana et al. (2022) used Bayesian optimization algorithm (BOA) assisted machine learning techniques i.e., ANN and SVM, for predicting the bio-oil yield from microalgae. It was observed that both developed ML models performed way better than the conventional optimizing statistical tool such as Response surface methodology (RSM). When comparing the two AI models developed, BOA-SVR had a greater performance at predicting biodiesel production than BOA-ANN. Li et al. (2021) have used three decision tree ML algorithms, RF, decision regression tree (DRT) and gradient boosting regression (GBR), for multi-task prediction of optimizing energy recovery, bio-oil yield, and nitrogen content from the HTL of microalgae. Among the explored ML models, RF performed better with $R^2 > 0.8$.

Hydrothermal carbonization (HTC) is a chemical process that converts high moisture content feedstock such as microalgae and sewage sludge into hydrochar, hydrogen-rich fuel gas, liquid bio-oil, and several valued added chemicals. Khoo et al. (2020) reported the conversion of algal biomass into hydrochar using HTC conversion. In a recent study, Gruber et al. (2022) employed the Bayesian regularization ANN algorithm for predicting the green hydrogen production from formic acid catalytic HTC conversion of *Chlorella vulgaris* biomass. The input variables considered for the model development are feedstock to suspension ratio and combined severity factor based on reaction conditions such as temperature, pH, and time. Whereas the output target variables are the solid, liquid, and gaseous product yield. The developed BR-ANN model predicted the HTC process accurately with an R^2 of >0.99. Chen et al. (2018) analyzed the pyrolysis kinetics of thermal degradation of carbohydrates, lipids, and proteins of three different microalgal species using the independent parallel reaction (IPR) model. They used the evolutionary algorithm particle swarm organization (PSO) for fitting the experimental kinetics data and predicted the microalgal thermal degradation curves with 97.9 % accuracy.

Other than biofuels, microalgae are also the source of several other valuable biomaterials, such as phenolics, phycobiliproteins, and nutrition compounds. Phenolic compounds as the secondary metabolites extracted from *Spirulina* species could be employed as a natural source of food preservative as it has antimicrobial properties. Very recently, Asnake Metekia et al. (2022) have used AI models, Multilayer perception (MLP) models, Adaptive-Neuro Fuzzy Inference System (ANFIS), and Step-Wise-Linear

Regression (SWLR) to predict the total phenolic compounds (TPC) extraction from the *Spirulina* algae. TPC was considered the model output, whereas the model inputs were the algal productivity, extraction yield, and percent of phenols and flavonoids. It was observed that ANFIS and SWLR models performed better compared to the MLP model. Furthermore, Saini et al. (2021) have employed multi-objective hybrid machine learning approach for simultaneously optimizing the biomass and phycobiliproteins (PBS) production from *Nostoc* sp. CCC-403. The input parameters considered for the model are the pH and three different BG-11 media compositions. They have reported a 90 % increase in biomass production and a 61.76 % increase in PBS production at the optimal conditions obtained from the model.

The drying of the microalgal biomass is the most energy-intensive stage in the downstream processes of the microalgal biorefinery. Accurate biomass parameters modeling, especially its moisture content, can enable effective regulation of the drying process to reduce operational costs. After harvesting or dewatering, the microalgae must be dried so that the water content of the cells is reduced. Approximately 75 % to 85 % of the energy consumed in algal biorefineries is spent on the drying of microalgae (Khoo et al., 2019). In essence, lowering the drying operational costs through efficient control is necessary for microalgal products to be commercially viable. Algal drying requires realistic models for process control to be effective. For example, Sonkar and Mallick (2020) have applied ML algorithm logistic regression to optimize the temperature and rotational speed of a rotary drum dryer for drying *Chlorella minutissima* biomass. Furthermore, Ching et al. (2021) demonstrated the use of AI models such as ANN networks, SVM, and extreme gradient boosting machine (XBG) for modeling vacuum drying of *Chlorococcum infusionum* for producing algal biofuel. They reported that the XBG approach performed better than competing models and showed improved ability at approximating sample points at the extremes of the dataset. In extension to this very recently, Pilario et al. (2022) have used Gaussian process autoregressive (GPAR) models for the prediction of moisture content during the vacuum drying of the biomass. It was reported that GPAR models outperformed the other studied models, such as ANN, SVM, RF, and XBG, for the same task.

Algal biomass harvesting is a critical stage in algal biorefinery, accounting for roughly 30 % of the overall expenses of algal biomass production.

Table 2
AI/ML tools in microalgal biomass conversion technologies.

Conversion technology	Feed biomass/species	Employed AI/ML algorithms	Input parameters	Output parameters	Remarks	Reference
Transesterification	Chlorella CG12	ANN and GA	reaction temperature, reaction time, and MeOH: oil molar ratio.	Percentage FAME conversion	Optimized reaction conditions for algal oil conversion to FAME. ANN performed better than RSM ($R^2 \sim 0.99$)	(Srivastava et al., 2018)
Hydrothermal liquefaction	Chlorella	GBR and RF	Biomass elemental and biochemical composition, Ratios of elementary composition	Oil yield, oxygen, and nitrogen contents of oil	Optimized conditions for increased bio-oil yield. GBR ($R^2 > 0.9$) exhibited better performance than RF	(Zhang et al., 2021)
Combustion	Spirulina	ANN	Fuel load, blending, and injection pressure	Combustion, performance, and emission characteristics	ANN was employed to characterize the microalgal biodiesel combustion properties	(Salam and Verma, 2019)
Thermal conversion	*Chlorella Vulgaris*	PDSE deep neural network algorithm	Temperature, heating rate, and the catalyst type	Microalgal biomass conversion	Employed deep neural networks for optimizing the thermal catalytic conversion of microalgae	(Teng et al., 2019)
Enzymatic Hydrolysis	Mixed microalgal culture	ANN	Biomass concentration, pH, temperature, and hydrolysis time	Reducing sugars yield	Modeled and predicted the optimum experimental conditions for enzymatic hydrolysis of mixed microalgae	(Shokrkar et al., 2017)
Catalytic Transesterification	Jatropha-algal oil blend	RSM, ANFIS	Blending molar ratio, reaction time, temperature, and catalyst dosage	Biodiesel yield	ANFIS model was more robust with R^2 of >0.99 in predicting the KOH catalyst-mediated transesterification	(Kumar et al., 2018)
Acid mediated Hydrothermal carbonization (HTC)	*Chlorella vulgaris*	ANN	Temperature, catalyst to suspension ratio, feedstock to suspension ratio	HTC products of gas, liquid, and solid phases	ANN model applied to HTC conversion of microalgal biomass to green hydrogen and obtained high accuracy ($R^2 \sim 0.99$)	(Gruber et al., 2022)
Microalgal secondary metabolites extraction	Spirulina	ANFIS, MLP, SWLR	Spirulina productivity, total flavonoids, extraction yield, percent of flavonoid, percent of phenols	Total phenolic compounds	Employed AI algorithms to predict the effect of growth mediums on total phenolic compounds. ANFIS and SWLR gave superior results than MLP	(Asnake Metekia et al., 2022)
Transesterification	Nannochloropsis oculate	BOA hybridization with ANN and SVM	Catalyst dose, algal oil to methanol ratio, reaction time, and temperature	Biodiesel yield	BOA-based ANN and SVM were used for the prediction of biodiesel yield from microalgal oil	(Sultana et al., 2022)
Acid catalytic direct transesterification	*Chlorella pyrenoidosa*	ANN and RSM	Time, temperature, solvent-to-wet biomass ratio, hydrochloric acid concentration	FAME yield	Biodiesel yield from *C. pyrenoidosa* was modeled using ANN model with high accuracy ($R^2 \sim 0.94$)	(Muhammad et al., 2022)
Catalytic transesterification	Nannochloropsis salina	ANN-MLP and RSM	Temperature, time, catalyst concentration, Oil: Methanol ratio	Biodiesel yield	CaO catalyst for biodiesel production from *Nannochloropsis salina* optimized using ANN-MLP ($R^2 \sim 0.94$) and RSM ($R^2 \sim 0.875$)	(Vinoth Arul Raj et al., 2021)
Transesterification	Explored different algal species	N2IC model	Reaction temperature, time, type of algae, and methanol-to-algal-oil ratio	Biodiesel production	Experimental conditions from the literature were modeled and optimized using N2IC model ($R^2 \sim 0.972$)	(Mahfouz et al., 2023)
Phycobiliproteins (PBS) production	cyanobacterium *Nostoc* sp. CCC-403	CNN and MOGA model	Three BG-11 media compositions and pH (6–10)	Biomass and PBS production	Employed multi-objective hybrid machine learning optimization for PBS production from *Nostoc* sp. CCC-403	(Saini et al., 2021)

ANN-Artificial Neural Networks, GA-Genetic Algorithm, FAME-Fatty Acid Methyl Ester, GBR-Gradient Boosting Regression, RF-Random Forest, PDSE- Progressive Depth Swarm Evolution, RSM-Response Surface Methodology, ANFIS- Adaptive Neuro-Fuzzy Inference System, MLP-Multilayer perception, SWLR-Step Wise Linear Regression, SVM-Support Vector Machine, BOA-Bayesian optimization algorithm, Neural-network-inspired correlation (N2IC) model, CNN-Connected neural network, MOGA-Multi objective genetic algorithm.

Pre-concentrating biomass using flocculation reduces microalgal harvesting costs and energy use. The physical features of the biomass, such as the structure and shape of the flocs, are critical for designing downstream separating or concentrating units in biorefineries at an industrial scale. The harvesting performance depends on the composition of floc structures, geometry, and settling velocity. Hence, understanding these flocs features is crucial in industrial harvesting setups. Applying ML models for enhancing algal harvesting could significantly reduce the costs associated with it. For instance, Lopez-Exposito et al. (2019) have used the ML model RF regression algorithm for estimating the *Chlorella sorokiniana* flocs dimensions by correlating the suspension chord length distribution with the average geometry of the flocs. Hence, these AI/ML models could be used to create autonomous, intelligent flocculation control systems that modify floc shape to suit the needs of following concentration operations by controlling the process stirring intensity. In converting microalgae to biodiesel, the grown microalgae must be harvested, and microalgal flocculation is an important step in harvesting and dewatering. Flocculation modeling can be used to evaluate and forecast the performance of a flocculation system under a variety of influencing conditions. Zenooz et al. (2017) have modeled the *Chlorella* sp. flocculation with ferric chloride under different using ANN models Radial Basis Function (RBF) and Multilayer perception (MLP). Based on the results, the MLP algorithm performed better in predicting microalgal flocculation.

The harvested microalgal biomass can be converted into biodiesel via transesterification. However, transesterification is affected by factors such as temperature, time, solvent ratio, etc. These process parameters can be optimized by using AI/ML models. For instance, Srivastava et al. (2018) employed an ANN model coupled with GA for optimizing the conversion of microalgal oil derived from *Chlorella* to Fatty acids methyl esters (FAME) via supercritical methanol transesterification. The input variables considered in the model are temperature, reaction time, and methanol: oil ratio, whereas the output variable was the percentage conversion. They have shown that the developed ANN model performed better than the conventional statistical optimization technique RSM. Improved biodiesel yield from the microalgal lipid has been obtained by using a fuzzy model coupled with particle swarm optimization (Nassef et al., 2019). Salam and Verma (2019) have employed the ANN model to analyze the combustion properties of microalgal biodiesel. Kumar et al. (2018) have used ANFIS and RSM models to predict the biodiesel yield during the transesterification of jatropha and algal oil blend and demonstrated that ANFIS performed better than the RSM. In similar, Muhammad et al. (2022) have used the ANN model to optimize the reaction conditions such as reaction time, temperature, acid concentration, and solid-biomass ratio in the conversion of *Chlorella pyrenoidosa* biomass into biodiesel via acid-mediated direct transesterification.

Enzymatic hydrolysis is also an approach in the algal biorefinery to convert microalgal biomass into reducing sugars for ethanol production. Shokrkar et al. (2017) employed the ANN model to predict the optimum condition for maximizing the reducing sugars yield during the enzymatic hydrolysis of mixed microalgae. The input process variables considered for optimization are the pH, biomass inoculum, temperature, and hydrolysis time, while the model output variable was the yield of reducing sugars. The application of different AI/ML models in various processes of algal biorefinery has been summarized in Table 2.

6. AI/ML integrated framework for optimized algal systems

Integrating microalgae with wastewater treatment can recover valuable resources by employing cutting-edge process control systems for optimized treatment and biorefinerys. Fig. 5 represents an integrated AI/ML-enabled smart systems framework for microalgal cultivation and resource recovery. In microalgal biorefinery industries, it is essential to incorporate microalgal productivity with the subsequent process operations. During microalgal cultivation, it is crucial to regulate the physicochemical parameters that affect its growth, such as the light intensity, pH, nutrients, CO_2, and algal biomass concentrations. There are numerous sensors available that can monitor these physiochemical parameters during algal cultivation. Different sensors and measuring tools/techniques can be employed in algal bioreactors to maintain the optimal conditions. This huge amount of datasets generated from these monitoring systems can be used as the input for the AI/ML models for parameter optimization. Biomass productivity and treatment efficiency could be enhanced by using online sensors for real-time monitoring and automation and by developing ML models from the acquired data. Based on the end product requirement, the output from these models can be used as feedback for maintaining optimum cultivation parameters. Recently, several researchers have started exploring the implementation of smart control systems for microalgal cultivation. For example, Zhu et al. (2022) have used a smart and precise control strategy for efficient paddle mixing in *Spirulina* open pond cultivation based on the feedback from light intensity and temperature and reported a decrease in 30 % energy input compared to the control. Tham et al. (2022) have designed and fabricated an Internet of Things (IoT) enabled up-scaled photobioreactor for facilitating remote monitoring of the parameters via smartphone. Furthermore, Lee et al. (2022) developed a 3D-printed real-time optical density monitoring instrument and were able to accurately predict the microalgal growth kinetics from the real-time data. These examples from the literature indicate that employing AI/ML-enabled smart systems in microalgal cultivation could significantly minimize resource consumption and aid microalgal biorefinery industries in better decision-making.

7. Challenges and future perspectives

Microalgal cultivation involves moving microalgae from their natural habitat into a laboratory or other artificial setting where their parameters have to be adjusted for optimal growth. Successful cultivation of microalgae requires optimal growth conditions. The produced algal biomass can be converted into several value-added products. Multiproduct algal biorefineries could be integrated with wastewater treatment systems to provide a sustainable method of synthesizing bioproducts in a circular bioeconomy framework while maintaining the sustainability of the water-energy environment nexus. Very recently, Malik et al. (2022) have demonstrated a novel biorefinery route by cultivating microalgae in wastewater and converting the biomass into biodiesel. Also, the biomass leftover after the biodiesel extraction was used for fermentation.

The first and foremost difficulty in implementing these AI/ML models in any area is that a lot of information is needed for the training and validation of these models. Generally, the accuracy of ML models increases with the increase in data availability. However, obtaining a huge amount of data can be expensive in terms of both time and cost in a real-world setting. In case of limited data availability, data preprocessing can be employed to increase the overall quality of the data. Data augmentation (DA) can artificially increase datasets based on known invariances, allowing the trained model to learn these invariances with better generalization. For instance, Correa et al. (2017) reported that data augmentation supported deep learning models have shown better accuracy than that without DA in microalgal classification. It must be stressed that irrational data augmentation may result in inaccurate forecasts.

Most of the existing studies employ a single ML model for making predictions. However, combinations of different algorithms have the potential to achieve better results and should be the subject of future research. In most cases, integrated models (such as ANN/SVM/RF coupled with GA) outperform standalone models in terms of prediction performance, overfitting risk, and robustness (Beltramo and Hitzmann, 2019). In addition, cutting-edge ML techniques like Reinforcement Learning and Deep Learning have garnered a lot of interest for their potential to forecast and optimize various energy-producing processes. One of the most popular types of machine learning is called deep learning, which also goes by the names "Deep Structured Learning," "Hierarchical Learning," and "Deep Machine Learning" (Lecun et al., 2015). Deep learning aims to find previously unknown features in data by extracting highly specialized low-level features and then forming more generalized high-level features (Lecun et al., 2015). Different types of DL networks that are used most frequently

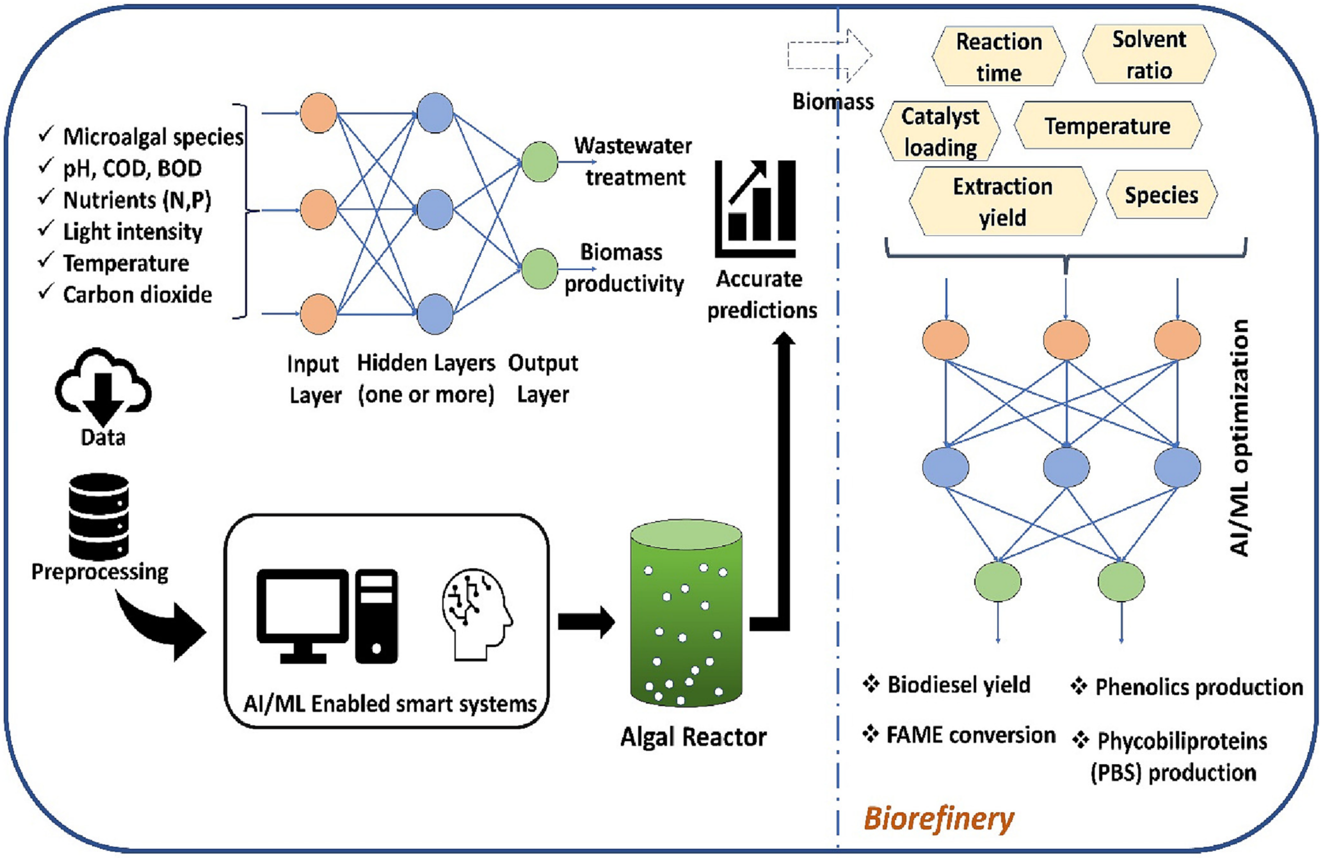


Fig. 5. Integrated framework for AI/ML enabled smart systems for microalgal cultivation and resource recovery.

are Convolutional neural networks (CNNs), Recursive neural networks (RNN), and auto-encoders (AE) (Lv and Lei, 2020). Naturalistic metaheuristic methods (such as the genetic algorithm, particle swarm optimization, differential evolution, etc.) might be used in place of exact algorithms to speed up the optimization process for microalgae reactors (Hernández-Pérez et al., 2019).

Providing ideal parameters, including pH, temperature, CO_2 supply and dissolved O_2, is essential for optimizing the reactor configuration/design, and increasing production at a reasonable cost is necessary. To achieve this, developing models and implementing complex control procedures is necessary. Conventionally, this has been handled by testing a variety of design configurations using integrated physical models that connect CFD and kinetic modeling. However, when simulating large-scale systems, this approach becomes computationally intractable and numerically unstable, necessitating time-consuming computing efforts and rendering mathematical optimization impracticable. These limitations can be overcome by coupling the physical models with the data-driven deep learning models. For instance, del Rio-Chanona et al., (2019) employed a data-driven deep learning surrogate modeling framework for optimizing the design configuration and operating conditions in a pilot scale photobioreactor. The developed framework was able to bring down the computational time from months to days. To begin, a CFD model was built to analyze the fundamental biological system behavior and produce sufficient data sets. Next, a convolutional neural network (CNN) was built to replicate the original model's nonlinearity and complexity. Further, a form of hybrid stochastic optimization was then used to determine the best possible settings for cultivation and PBR design. A recent study also has proposed employing a deep learning algorithm Mask-RCNN (Region-based Convolutional Neural Network) model, for instant segmentation in diatom algae detection from water samples (Ruiz-santaquiteria et al., 2020). On average, Mask-RCNN achieved 0.86 % sensitivity and 0.91 % specificity from microscopic images of a variety of mixed species of algae in the water.

Different reactor configurations and goals necessitate distinct approaches to automating algal bioreactor systems. However, in conjunction with online sensors, MLAs can be used to reduce costs and maximize the efficiency of algal systems for increased biomass production by using the fundamental computational concepts of intelligently controlling process parameters using algorithms. The operational costs of a microalgal biorefinery can be reduced through the automation of the algal cultivation and harvesting system. The operators may also be able to monitor the microalgal growth and productivity in real-time using a network of AI/ML-enabled plug-and-play Internet of Things (IoT) sensors (Lim et al., 2022). When applied to the training of deep learning algorithms, the data from IoT sensors can ensure that these AI/ML models can perform tasks like analysis, monitoring, and prediction that are unique to each application. The process of identifying microalgal strains and species, which now relies on microscopic pictures and spectroscopic analysis, might be greatly accelerated and automated with AI-based algorithms. The microalgal biorefinery has been challenged with forecasting microalgae biomass production. The operators of the culturing reactor were tasked with measuring the biomass of the microalgae on a daily basis. However, the implementation of IoT sensors and an optimization strategy based on machine learning has the potential to improve resource utilization. This could be accomplished by optimizing the culture conditions to produce high levels of microalgal biomass while simultaneously reducing the amount of investment required. In a very recent study, Peter et al. (2023) employed an AI-enabled IoT-based unique digital architecture framework for intelligent microalgae growth surveillance and optimizing the nutrient media recycling strategy that could result in high biomass production in semi-batch cultivation of *Chlorella vulgaris*. Furthermore, a biorefinery organization's production plan and operations can be enhanced with the use of an AI/ML predictive model that estimates microalgae productivity.

It takes a lot of time and effort on the part of researchers to conduct experimental studies in the process of microalgae conversion into valuable products. Numerous methods, including thermochemical and biochemical processes, are being explored for their potential to be used in the microalgae conversion process (Suali and Sarbatly, 2012). Conversion mechanisms in microalgae are difficult to anticipate because of their diverse chemical composition (Teng et al., 2019). AI/MLAs are being widely used for performance prediction and determining optimum experimental conditions in biomass conversion technologies. Recent advances in AI algorithm development have made it possible for optimizing experimental conditions for increased yield under uncertainty (Chen et al., 2018). Advances in neural network design currently allow for the successful simulation of temporal effects, which makes them suitable for studying dynamic microalgae conversion technologies. In addition, neuro-evolutionary meta-learning can be used to forecast the most efficient thermal conversion of microalgae. Teng et al. (2019) used the neuro-evolutionary approach to determine the optimum thermal conversion of *Chlorella vulgaris*. They have employed the neuro-evolutionary algorithm Progressive Depth Swarm-Evolution (PDSE) to model the thermogravimetric analysis (TGA) data of catalytic thermal degradation. The proposed model was able to generate accurate predictions compared to conventional approaches. At the model-predicted optimum conditions, 83 % of biomass conversion was reported.

Integrating wastewater treatment with the microalgal biorefinery significantly contributes to the United Nation's sustainable development goals (SDGs) both directly and indirectly. The microalgae directly contribute to the SDGs - SDG-2 (zero hunger), SDG-6 (clean water and sanitation), SDG-7 (affordable and clean energy), SDG-9 (industry, innovation and infrastructure), SDG-12 (responsible consumption and production), SDG-14 (life below water), and SDG-15 (life on land). Microalgal biotechnology integrates and helps in achieving different SDGs (Olabi et al., 2023; Sutherland et al., 2021). Microalgal biotechnology must achieve product optimization and cost-effective large-scale cultivation to achieve SDGs and needs significant investment to expedite technological progress. However, microalgal biotechnology offers viable alternatives that have reduced environmental impacts. Incorporating AI/ML models into microalgal biotechnology can assist in effective decision-making, significantly lowering associated process costs and in effectively contributing to the attainment of the SDGs.

8. Conclusions

The complex and diverse nature of microalgae necessitates a large amount of data and knowledge from multiple disciplines, including species selection, cultivation parameters, reactor design, and conversion technologies for integrating wastewater treatment with algal biorefinery. This article discussed the benefits of using various AI/ML models such as ANN, ANFIS, SVM, RF, and GA in microalgal applications. Several AI algorithms can cut down the time and effort needed to optimize growing conditions for microalgae and increase yields. Using AI/ML models could help the microalgae biorefinery to meet production goals in a more eco-friendly, intelligent, autonomous, and cost-effective way.

CRediT authorship contribution statement

Raj Kumar Oruganti: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Software, Investigation, Writing – original draft. **Alka Pulimoottil Biji:** Visualization, Investigation. **Tiamenla Lanuyanger:** Visualization, Investigation. **Pau Loke Show:** Writing – review & editing. **Malinee Sriariyanun:** Writing – review & editing, Visualization. **Venkata K.K. Upadhyayula:** Conceptualization, Writing – review & editing. **Venkataramana Gadhamshetty:** Conceptualization, Writing – review & editing. **Debraj Bhattacharyya:** Conceptualization, Methodology, Project administration, Supervision, Funding acquisition, Writing – review & editing.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

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

Acknowledgment

This work was supported by the Science and Engineering Research Board of India under IMPRINT 2 Program.

References

Alrashed, A.A.A.A., Karimipour, A., Bagherzadeh, S.A., Safaei, M.R., Afrand, M., 2018. Electro- and thermophysical properties of water-based nanofluids containing copper ferrite nanoparticles coated with silica: experimental data, modeling through enhanced ANN and curve fitting. Int. J. Heat Mass Transf. 127, 925–935. https://doi.org/10.1016/j.ijheatmasstransfer.2018.07.123.

Alzubi, J., Nayyar, A., Kumar, A., 2018. Machine learning from theory to algorithms : an overview machine learning from theory to algorithms : an overview. J. Phys., 012012 https://doi.org/10.1088/1742-6596/1142/1/012012.

Andrade Cruz, I., Chuenchart, W., Long, F., Surendra, K.C., Renata Santos Andrade, L., Bilal, M., Liu, H., Tavares Figueiredo, R., Khanal, S.K., Fernando Romanholo Ferreira, L., 2022. Application of machine learning in anaerobic digestion: perspectives and challenges. Bioresour. Technol. 345, 126433. https://doi.org/10.1016/j.biortech.2021.126433.

Ansari, F.A., Nasr, M., Rawat, I., Bux, F., 2021. Artificial neural network and techno-economic estimation with algae-based tertiary wastewater treatment. J. Water Process Eng. 40, 101761. https://doi.org/10.1016/j.jwpe.2020.101761.

Asnake Metekia, W., Garba Usman, A., Hatice Ulusoy, B., Isah Abba, S., Chirkena Bali, K., 2022. Artificial intelligence-based approaches for modeling the effects of spirulina growth mediums on total phenolic compounds. Saudi J. Biol. Sci. 29, 1111–1117. https://doi.org/10.1016/j.sjbs.2021.09.055.

Bagherzadeh, F., Mehrani, M.J., Basirifard, M., Roostaei, J., 2021. Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance. J. Water Process Eng. 41, 102033. https://doi.org/10.1016/j.jwpe.2021.102033.

Banerjee, A., Guria, C., Maiti, S.K., 2016. Fertilizer assisted optimal cultivation of microalgae using response surface method and genetic algorithm for biofuel feedstock. Energy 115, 1272–1290. https://doi.org/10.1016/j.energy.2016.09.066.

Beltramo, T., Hitzmann, B., 2019. Engineering in agriculture, environment and food evaluation of the linear and non-linear prediction models optimized with metaheuristics : application to anaerobic digestion processes. Eng. Agric. Environ. Food 1–7. https://doi.org/10.1016/j.eaef.2019.06.001.

Bi, X., Lin, S., Zhu, S., Yin, H., Li, Z., Chen, Z., 2019. Species identification and survival competition analysis of microalgae via hyperspectral microscopic images. 176, 191–197. https://doi.org/10.1016/j.ijleo.2018.09.077.

Camacho-Rodríguez, J., Cerón-García, M.C., Fernández-Sevilla, J.M., Molina-Grima, E., 2015. Genetic algorithm for the medium optimization of the microalga nannochloropsis gaditana cultured to aquaculture. Bioresour. Technol. 177, 102–109. https://doi.org/10.1016/j.biortech.2014.11.057.

Cheah, W.Y., Show, P.L., Juan, J.C., Chang, J.S., Ling, T.C., 2018. Enhancing biomass and lipid productions of microalgae in palm oil mill effluent using carbon and nutrient supplementation. Energy Convers. Manag. 164, 188–197. https://doi.org/10.1016/j.enconman.2018.02.094.

Chen, C., Yeh, K., Aisyah, R., Lee, D., Chang, J., 2011. Bioresource technology cultivation, photobioreactor design and harvesting of microalgae for biodiesel production : a critical review. Bioresour. Technol. 102, 71–81. https://doi.org/10.1016/j.biortech.2010.06.159.

Chen, J., Dai, L., Mataya, D., Cobb, K., Chen, P., Ruan, R., 2022. Enhanced sustainable integration of CO2 utilization and wastewater treatment using microalgae in circular economy concept. Bioresour. Technol. 366, 128188. https://doi.org/10.1016/j.biortech.2022.128188.

Chen, W.H., Chu, Y.S., Liu, J.L., Chang, J.S., 2018. Thermal degradation of carbohydrates, proteins and lipids in microalgae analyzed by evolutionary computation. Energy Convers. Manag. 160, 209–219. https://doi.org/10.1016/j.enconman.2018.01.036.

Chew, K.W., Yap, J.Y., Show, P.L., Suan, N.H., Juan, J.C., Ling, T.C., Lee, D.J., Chang, J.S., 2017. Microalgae biorefinery: high value products perspectives. Bioresour. Technol. 229, 53–62. https://doi.org/10.1016/j.biortech.2017.01.006.

Ching, P.M.L., Mayol, A.P., San Juan, J.L.G., Calapatia, A.M., So, R.H.Y., Sy, C.L., Ubando, A.T., Culaba, A.B., 2021. AI methods for modeling the vacuum drying characteristics of Chlorococcum infusionum for algal biofuel production. Process Integr. Optim. Sustain. 5, 247–256. https://doi.org/10.1007/s41660-020-00145-4.

Christenson, L., Sims, R., 2011. Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts. Biotechnol. Adv. 29, 686–702. https://doi.org/10.1016/j.biotechadv.2011.05.015.

Correa, I., Drews, P., Botelho, S., De Souza, M.S., Tavano, V.M., 2017. Deep learning for microalgae classification. Proc. - 16th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2017 2017-Decem, pp. 20–25 https://doi.org/10.1109/ICMLA.2017.0-183.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1088/1742-6596/628/1/012073.

Coşgun, A., Günay, M.E., Yıldırım, R., 2021. Exploring the critical factors of algal biomass and lipid production for renewable fuel production by machine learning. Renew. Energy 163, 1299–1317. https://doi.org/10.1016/j.renene.2020.09.034.

Dallora, A.L., Eivazzadeh, S., Mendes, E., Berglund, J., Anderberg, P., 2017. Machine learning and microsimulation techniques on the prognosis of dementia: a systematic literature review. PLoS One 12, 1–23. https://doi.org/10.1371/journal.pone.0179804.

Del Rio-Chanona, E.A., Ahmed, N.R., Wagner, J., Lu, Y., Zhang, D., Jing, K., 2019. Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses. Biotechnol. Bioeng. https://doi.org/10.1002/bit.27131.

del Rio-Chanona, E.A., Wagner, J.L., Ali, H., Fiorelli, F., Zhang, D., Hellgardt, K., 2019. Deep learning-based surrogate modeling and optimization for microalgal biofuel production and photobioreactor design. AIChE J. 65, 915–923. https://doi.org/10.1002/aic.16473.

Dong, C., Chen, J., 2019. Optimization of process parameters for anaerobic fermentation of corn stalk based on least squares support vector machine. Bioresour. Technol. 271, 174–181. https://doi.org/10.1016/j.biortech.2018.09.085.

Dragone, G., Fernandes, B., Vicente, A., Teixeira, J.A., 2010. Third generation biofuels from microalgae. Curr. Res. Technol. Educ. Top. Appl. Microbiol. Microb. Biotechnol., 1355–1366 https://doi.org/10.1055/s-0028-1134195.

Drews-, P., Colares, R.G., Machado, P., de Faria, M., Detoni, A., Tavano, V., 2013. Microalgae classification using semi-supervised and active learning based on gaussian mixture models. J. Braz. Comput. Soc. 19, 411–422. https://doi.org/10.1007/s13173-013-0121-y.

Faris, H., Mirjalili, S., Aljarah, I., 2019. Automatic selection of hidden neurons and weights in neural networks using grey wolf optimizer based on a hybrid encoding scheme. Int. J. Mach. Learn. Cybern. 10, 2901–2920. https://doi.org/10.1007/s13042-018-00913-2.

Franco, B.M., Navas, L.M., Gómez, C., Sepúlveda, C., Acién, F.G., 2019. Monoalgal and mixed algal cultures discrimination by using an artificial neural network. Algal Res. 38, 101419. https://doi.org/10.1016/j.algal.2019.101419.

García-Camacho, F., López-Rosales, L., Sánchez-Mirón, A., Belarbi, E.H., Chisti, Y., Molina-Grima, E., 2016. Artificial neural network modeling for predicting the growth of the microalga karlodinium veneficum. Algal Res. 14, 58–64. https://doi.org/10.1016/j.algal.2016.01.002.

Giraldo-Zuluaga, J.H., Salazar, A., Diez, G., Gomez, A., Martínez, T., Vargas, J.F., Peñuela, M., 2018. Automatic identification of scenedesmus polymorphic microalgae from microscopic images. Pattern Anal. Appl. 21, 601–612. https://doi.org/10.1007/s10044-017-0662-3.

Gruber, Z., Toth, A.J., Menyhárd, A., Mizsey, P., Owsianiak, M., Fozer, D., 2022. Improving green hydrogen production from Chlorella vulgaris via formic acid-mediated hydrothermal carbonisation and neural network modelling. Bioresour. Technol. 365. https://doi.org/10.1016/j.biortech.2022.128071.

Guo, H.Nan, Wu, S.Biao, Tian, Y.Jie, Zhang, J., Liu, H.Tao, 2021. Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: a review. Bioresour. Technol. 319, 124114. https://doi.org/10.1016/j.biortech.2020.124114.

Harmon, J., Mikami, H., Kanno, H., Ito, T., Goda, K., 2020. Accurate classification of microalgae by intelligent frequency-division-multiplexed fluorescence imaging flow cytometry. OSA Contin. 3, 430. https://doi.org/10.1364/osac.387523.

He, L., Subramanian, V.R., Tang, Y.J., 2012. Experimental analysis and model-based optimization of microalgae growth in photo-bioreactors using flue gas. Biomass Bioenergy 41, 131–138. https://doi.org/10.1016/j.biombioe.2012.02.025.

Hernández-Pérez, L.G., Sánchez-Tuirán, E., Ojeda, K.A., Germa, L., El-halwagi, M.M., Ponce-ortega, J.M., 2019. Optimization of microalgae-to-biodiesel production process using a metaheuristic technique´. ACS Sustain. Chem. Eng. Eng. 7, 8490–8498. https://doi.org/10.1021/acssuschemeng.9b00274.

Hossain, S.M.Z., Sultana, N., Jassim, M.S., Coskuner, G., Hazin, L.M., Razzak, S.A., Hossain, M.M., 2022. Soft-computing modeling and multiresponse optimization for nutrient removal process from municipal wastewater using microalgae. J. Water Process Eng. 45, 102490. https://doi.org/10.1016/j.jwpe.2021.102490.

Hu, D., Liu, H., Yang, C., Hu, E., 2008. The design and optimization for light-algae bioreactor controller based on artificial neural network-model predictive control. Acta Astronaut. 63, 1067–1075. https://doi.org/10.1016/j.actaastro.2008.02.008.

Jha, K., Doshi, A., Patel, P., Shah, M., 2019. A comprehensive review on automation in agriculture using artificial intelligence. Artif. Intell. Agric. 2, 1–12. https://doi.org/10.1016/j.aiia.2019.05.004.

Kannangara, M., Dua, R., Ahmadi, L., Bensebaa, F., 2018. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. Waste Manag. 74, 3–15. https://doi.org/10.1016/j.wasman.2017.11.057.

Khoo, C.G., Dasan, Y.K., Lam, M.K., Lee, K.T., 2019. Algae biorefinery: review on a broad spectrum of downstream processes and products. Bioresour. Technol. 292, 121964. https://doi.org/10.1016/j.biortech.2019.121964.

Khoo, C.G., Lam, M.K., Mohamed, A.R., Lee, K.T., 2020. Hydrochar production from high-ash low-lipid microalgal biomass via hydrothermal carbonization: effects of operational parameters and products characterization. Environ. Res. 188, 109828. https://doi.org/10.1016/j.envres.2020.109828.

Kumar, S., Jain, S., Kumar, H., 2018. Performance evaluation of adaptive neuro-fuzzy inference system and response surface methodology in modeling biodiesel synthesis from jatropha–algae oil. Energy Sources Part A 40, 3000–3008. https://doi.org/10.1080/15567036.2018.1515277.

Kushwaha, O.S., Uthayakumar, H., Kumaresan, K., 2022. 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. https://doi.org/10.1007/s11356-022-19683-0.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lee, J.S., Sung, Y.J., Sim, S.J., 2022. Kinetic analysis of microalgae cultivation utilizing 3D-printed real-time monitoring system reveals potential of biological CO2 conversion. Bioresour. Technol. 364, 128014. https://doi.org/10.1016/j.biortech.2022.128014.

Lee, S.Y., Khoiroh, I., Vo, D.V.N., Senthil Kumar, P., Show, P.L., 2021. Techniques of lipid extraction from microalgae for biofuel production: a review. Environ. Chem. Lett. 19, 231–251. https://doi.org/10.1007/s10311-020-01088-5.

Li, J., Zhang, W., Liu, T., Yang, L., Li, H., Peng, H., Jiang, S., Wang, X., Leng, L., 2021. Machine learning aided bio-oil production with high energy recovery and low nitrogen content from hydrothermal liquefaction of biomass with experiment verification. Chem. Eng. J. 425, 130649. https://doi.org/10.1016/j.cej.2021.130649.

Li, K., Liu, Q., Fang, F., Luo, R., Lu, Q., Zhou, W., Huo, S., Cheng, P., Liu, J., Addy, M., Chen, P., Chen, D., Ruan, R., 2019. Microalgae-based wastewater treatment for nutrients recovery: a review. Bioresour. Technol. 291, 121934. https://doi.org/10.1016/j.biortech.2019.121934.

Lim, H.R., Khoo, K.S., Chia, W.Y., Chew, K.W., Ho, S.H., Show, P.L., 2022. Smart microalgae farming with internet-of-things for sustainable agriculture. Biotechnol. Adv. 57, 107931. https://doi.org/10.1016/j.biotechadv.2022.107931.

Liu, J.Y., Zeng, L.H., Ren, Z.H., Du, T.M., Liu, X., 2020. Rapid in situ measurements of algal cell concentrations using an artificial neural network and single-excitation fluorescence spectrometry. Algal Res. 45, 101739. https://doi.org/10.1016/j.algal.2019.101739.

Lopez-Exposito, P., Negro, C., Blanco, A., 2019. Direct estimation of microalgal flocs fractal dimension through laser reflectance and machine learning. Algal Res. 37, 240–247. https://doi.org/10.1016/j.algal.2018.12.007.

López-Rosales, L., Gallardo-Rodríguez, J.J., Sánchez-Mirón, A., Contreras-Gómez, A., García-Camacho, F., Molina-Grima, E., 2013. Modelling of multi-nutrient interactions in growth of the dinoflagellate microalga Protoceratium reticulatum using artificial neural networks. Bioresour. Technol. 146, 682–688. https://doi.org/10.1016/j.biortech.2013.07.141.

Lv, W.H., Lei, J.Y., 2020. Deep learning development review. Proc. - 2020 3rd Int. Conf. Adv. Electron. Mater. Comput. Softw. Eng. AEMCSE 2020, pp. 171–174 https://doi.org/10.1109/AEMCSE50948.2020.00043.

Mahfouz, A.Bin, Ali, A., Crocker, M., Ahmed, A., Nasir, R., Show, P.L., 2023. Neural-network-inspired correlation (N2IC) model for estimating biodiesel conversion in algal biodiesel units. Fermentation 9, 47. https://doi.org/10.3390/fermentation9010047.

Malik, S., Shahid, A., Betenbaugh, M.J., Liu, C.G., Mehmood, M.A., 2022. A novel wastewater-derived cascading algal biorefinery route for complete valorization of the biomass to biodiesel and value-added bioproducts. Energy Convers. Manag. 256, 115360. https://doi.org/10.1016/j.enconman.2022.115360.

Miranda, A.M., Hernandez-Tenorio, F., Ocampo, D., Vargas, G.J., Sáez, A.A., 2022. Trends on CO2 capture with microalgae: a bibliometric analysis. Molecules 27, 1–14. https://doi.org/10.3390/molecules27154669.

Morowvat, M.H., Ghasemi, Y., 2016. Medium optimization by artificial neural networks for maximizing the triglycerides-rich lipids from biomass of chlorella vulgaris. Int. J. Pharm. Clin. Res. 8, 1414–1417.

Muhammad, G., Potchamyou Ngatcha, A.D., Lv, Y., Xiong, W., El-Badry, Y.A., Asmatulu, E., Xu, J., Alam, M.A., 2022. Enhanced biodiesel production from wet microalgae biomass optimized via response surface methodology and artificial neural network. Renew. Energy 184, 753–764. https://doi.org/10.1016/j.renene.2021.11.091.

Nassef, A.M., Sayed, E.T., Rezk, H., Abdelkareem, M.A., Rodriguez, C., Olabi, A.G., 2019. Fuzzy-modeling with particle swarm optimization for enhancing the production of biodiesel from Microalga. Energy Sources Part A 41, 2094–2103. https://doi.org/10.1080/15567036.2018.1549171.

Nayak, M., Dhanarajan, G., Dineshkumar, R., Sen, R., 2018. Artificial intelligence driven process optimization for cleaner production of biomass with co-valorization of wastewater and flue gas in an algal biorefinery. J. Clean. Prod. 201, 1092–1100. https://doi.org/10.1016/j.jclepro.2018.08.048.

Olabi, A.G., Shehata, N., Sayed, E.T., Rodriguez, C., Anyanwu, R.C., Russell, C., Abdelkareem, M.A., 2023. Role of microalgae in achieving sustainable development goals and circular economy. Sci. Total Environ. 854, 158689. https://doi.org/10.1016/j.scitotenv.2022.158689.

Otálora, P., Guzmán, J.L., Acién, F.G., Berenguel, M., Reul, A., 2021. Microalgae classification based on machine learning techniques. Algal Res. 55. https://doi.org/10.1016/j.algal.2021.102256.

Peter, A.P., Chew, K.W., Pandey, A., Lau, S.Y., Rajendran, S., Ting, H.Y., Munawaroh, H.S.H., Phuong, N.Van, Show, P.L., 2023. Artificial intelligence model for monitoring biomass growth in semi-batch Chlorella vulgaris cultivation. Fuel 333, 126438. https://doi.org/10.1016/j.fuel.2022.126438.

Pilario, K.E.S., Ching, P.M.L., Calapatia, A.M.A., Culaba, A.B., 2022. Predicting drying curves in algal biorefineries using gaussian process autoregressive models. Digit. Chem. Eng. 4, 100036. https://doi.org/10.1016/j.dche.2022.100036.

Reimann, R., Zeng, B., Jakopec, M., Burdukiewicz, M., Petrick, I., Schierack, P., Rödiger, S., 2020. Classification of dead and living microalgae Chlorella vulgaris by bioimage informatics and machine learning. Algal Res. 48, 101908. https://doi.org/10.1016/j.algal.2020.101908.

Rodríguez-Miranda, E., Acién, F.G., Guzmán, J.L., Berenguel, M., Visioli, A., 2021. A new model to analyze the temperature effect on the microalgae performance at large scale raceway reactors. Biotechnol. Bioeng. 118, 877–889. https://doi.org/10.1002/bit.27617.

Ruiz-santaquiteria, J., Bueno, G., Deniz, O., Vallez, N., Cristobal, G., 2020. Engineering applications of artificial intelligence semantic versus instance segmentation in microscopic algae detection ☆. Eng. Appl. Artif. Intell. 87, 103271. https://doi.org/10.1016/j.engappai.2019.103271.

Saboe, D., Ghasemi, H., Gao, M.M., Samardzic, M., Hristovski, K.D., Boscovic, D., Burge, S.R., Burge, R.G., Hoffman, D.A., 2021. Real-time monitoring and prediction of water quality parameters and algae concentrations using microbial potentiometric sensor signals and machine learning tools. Sci. Total Environ. 764, 142876. https://doi.org/10.1016/j.scitotenv.2020.142876.

Saini, D.K., Rai, A., Devi, A., Pabbi, S., Chhabra, D., Chang, J.S., Shukla, P., 2021. A multi-objective hybrid machine learning approach-based optimization for enhanced biomass and bioactive phycobiliproteins production in nostoc sp. CCC-403. Bioresour. Technol. 329. https://doi.org/10.1016/j.biortech.2021.124908.

Salam, S., Verma, T.N., 2019. Appending empirical modelling to numerical solution for behaviour characterisation of microalgae biodiesel. Energy Convers. Manag. 180, 496–510. https://doi.org/10.1016/j.enconman.2018.11.014.

Sayed, E.T., Rezk, H., Abdelkareem, M.A., Olabi, A.G., 2023. Artificial neural network based modelling and optimization of microalgae microbial fuel cell. Int. J. Hydrog. Energy https://doi.org/10.1016/j.ijhydene.2022.12.081.

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Serive, B., Nicolau, E., Bérard, J.B., Kaas, R., Pasquet, V., Picot, L., Cadoret, J.P., 2017. Community analysis of pigment patterns from 37 microalgae strains reveals new carotenoids and porphyrins characteristic of distinct strains and taxonomic groups. PLoS ONE https://doi.org/10.1371/journal.pone.0171872.

Shafizadeh, A., Shahbeig, H., Nadian, M.H., Mobli, H., Dowlati, M., Gupta, V.K., Peng, W., Lam, S.S., Tabatabaei, M., Aghbashlo, M., 2022. Machine learning predicts and optimizes hydrothermal liquefaction of biomass. Chem. Eng. J. 445, 136579. https://doi.org/10.1016/j.cej.2022.136579.

Shokrkar, H., Ebrahimi, S., Zamani, M., 2017. Extraction of sugars from mixed microalgae culture using enzymatic hydrolysis: experimental study and modeling. Chem. Eng. Commun. 204, 1246–1257. https://doi.org/10.1080/00986445.2017.1356291.

Singh, V., Mishra, V., 2022. Evaluation of the effects of input variables on the growth of two microalgae classes during wastewater treatment. Water Res. 213, 118165. https://doi.org/10.1016/j.watres.2022.118165.

Singh, V., Mishra, V., 2021. Exploring the effects of different combinations of predictor variables for the treatment of wastewater by microalgae and biomass production. Biochem. Eng. J. 174, 108129. https://doi.org/10.1016/j.bej.2021.108129.

Sonkar, S.S., Mallick, N., 2020. Application of machine learning for development of a drying protocol for microalga Chlorella minutissima in a single rotary drum dryer for biodiesel production. Authorea Prepr, pp. 1–10.

Srivastava, G., Paul, A.K., Goud, V.V., 2018. Optimization of non-catalytic transesterification of microalgae oil to biodiesel under supercritical methanol condition. Energy Convers. Manag. 156, 269–278. https://doi.org/10.1016/j.enconman.2017.10.093.

Suali, E., Sarbatly, R., 2012. Conversion of microalgae to biofuel. Renew. Sust. Energ. Rev. 16, 4316–4342. https://doi.org/10.1016/j.rser.2012.03.047.

Sultana, N., Hossain, S.M.Z., Abusaad, M., Alanbar, N., Senan, Y., Razzak, S.A., 2022. Prediction of biodiesel production from microalgal oil using bayesian optimization algorithm-based machine learning approaches. Fuel 309, 122184. https://doi.org/10.1016/j.fuel.2021.122184.

Sun, Z., Li, T., Zhou, Z., Jiang, Y., 2015. Microalgae as a Source of Lutein : Chemistry , Biosynthesis , and Carotenogenesis. https://doi.org/10.1007/10.

Sundui, B., Ramirez Calderon, O.A., Abdeldayem, O.M., Lázaro-Gil, J., Rene, E.R., Sambuu, U., 2021. Applications of machine learning algorithms for biological wastewater treatment: updates and perspectives. Clean Techn. Environ. Policy 23, 127–143. https://doi.org/10.1007/s10098-020-01993-x.

Supriyanto, Noguchi, R., Ahamed, T., Mikihide, D., Watanabe, M.M., 2018. A decision tree approach to estimate the microalgae production in open raceway pond. IOP Conf. Ser. Earth Environ. Sci., 209 https://doi.org/10.1088/1755-1315/209/1/012050.

Susanna, D., Dhanapal, R., Mahalingam, R., Ramamurthy, V., 2019. Increasing productivity of Spirulina platensis in photobioreactors using artificial neural network modeling. Biotechnol. Bioeng. 116, 2960–2970. https://doi.org/10.1002/bit.27128.

Sutherland, D.L., McCauley, J., Labeeuw, L., Ray, P., Kuzhiumparambil, U., Hall, C., Doblin, M., Nguyen, L.N., Ralph, P.J., 2021. How microalgal biotechnology can assist with the UN sustainable development goals for natural resource management. Curr. Res. Environ. Sustain. 3, 100050. https://doi.org/10.1016/j.crsust.2021.100050.

Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wu, Z., Chen, H., Naqvi, S.R., 2020. Prediction of bio-oil yield and hydrogen contents based on machine learning method: effect of biomass compositions and pyrolysis conditions. Energy Fuels 34, 11050–11060. https://doi.org/10.1021/acs.energyfuels.0c01893.

Teng, S.Y., Loy, A.C.M., Leong, W.D., How, B.S., Chin, B.L.F., Máša, V., 2019. Catalytic thermal degradation of Chlorella vulgaris: evolving deep neural networks for optimization. Bioresour. Technol. 292, 121971. https://doi.org/10.1016/j.biortech.2019.121971.

Tham, P.E., Ng, Y.J., Vadivelu, N., Lim, H.R., Khoo, K.S., Chew, K.W., Show, P.L., 2022. Sustainable smart photobioreactor for continuous cultivation of microalgae embedded with internet of things. Bioresour. Technol. 346, 126558. https://doi.org/10.1016/j.biortech.2021.126558.

Thornton, C., Frank, H., Leyton-brown, K., 2013. Auto-WEKA : Combined Selection and Hyperparameter Optimization of Classification Algorithms. Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min, pp. 847–855.

Vinoth Arul Raj, J., Praveen Kumar, R., Vijayakumar, B., Gnansounou, E., Bharathiraja, B., 2021. Modelling and process optimization for biodiesel production from Nannochloropsis salina using artificial neural network. Bioresour. Technol. 329, 2–9. https://doi.org/10.1016/j.biortech.2021.124872.

Wang, S., Wang, J., Shang, F., Wang, Y., Cheng, Q., Liu, N., 2020. A GA-BP method of detecting carbamate pesticide mixture based on three-dimensional fluorescence spectroscopy. Spectrochim. Acta A Mol. Biomol. Spectrosc. 224, 117396. https://doi.org/10.1016/j.saa.2019.117396.

Wang, Z., Peng, X., Xia, A., Shah, A.A., Huang, Y., Zhu, Xianqing, Zhu, Xun, Liao, Q., 2022. The role of machine learning to boost the bioenergy and biofuels conversion. Bioresour. Technol. 343, 126099. https://doi.org/10.1016/j.biortech.2021.126099.

Wu, Z., Shi, X., 2007. Optimization for high-density cultivation of heterotrophic chlorella based on a hybrid neural network model. Lett. Appl. Microbiol. 44, 13–18. https://doi.org/10.1111/j.1472-765X.2006.02038.x.

Ying Ying Tang, D., Wayne Chew, K., Ting, H.Y., Sia, Y.H., Gentili, F.G., Park, Y.K., Banat, F., Culaba, A.B., Ma, Z., Loke Show, P., 2023. Application of regression and artificial neural network analysis of Red-Green-Blue image components in prediction of chlorophyll content in microalgae. Bioresour. Technol. 370, 128503. https://doi.org/10.1016/j.biortech.2022.128503.

You, H., Ma, Z., Tang, Y., Wang, Y., Yan, J., Ni, M., Cen, K., Huang, Q., 2017. Comparison of ANN (MLP), ANFIS, SVM, and RF models for the online classification of heating value of burning municipal solid waste in circulating fluidized bed incinerators. Waste Manag. 68, 186–197. https://doi.org/10.1016/j.wasman.2017.03.044.

Zenooz, A.M., Ashtiani, F.Z., Ranjbar, R., Nikbakht, F., Bolouri, O., 2017. Comparison of different artificial neural network architectures in modeling of chlorella sp. Flocculation. Prep. Biochem. Biotechnol. 47, 570–577. https://doi.org/10.1080/10826068.2016.1275013.

Zhang, W., Li, J., Liu, T., Leng, S., Yang, L., Peng, H., Jiang, S., Zhou, W., Leng, L., Li, H., 2021. Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae. Bioresour. Technol. 342, 126011. https://doi.org/10.1016/j.biortech.2021.126011.

Zhu, C., Ji, Y., Du, X., Kong, F., Chi, Z., Zhao, Y., 2022. A smart and precise mixing strategy for efficient and cost-effective microalgae production in open ponds. Sci. Total Environ. 852. https://doi.org/10.1016/j.scitotenv.2022.158515.