Biointerface Research in Applied Chemistry
Open-Access Journal (ISSN: 2069-5837)

Article
Volume 16, Issue 4, 2026, 109
https://doi.org/10.33263/BRIAC164.109

Optimization of an Artificial Neural Network for Forecasting the Operating Conditions of CO_2 Fixation Rate By *Chlorella* Genus Algae

Koh Zhi Ying [1], Nur Hidayah binti Zainan [1], Yanti Maslina Binti Mohd Jusoh [1], Zanariah binti Hashim [1], Yu Huan Hi [2], Nur Farzana binti Ahmad Sanadi [1,*]

[1] Department of Bioprocess and Polymer Engineering, Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia, 81300, Skudai, Johor, Malaysia
[2] School of Chemistry and Chemical Engineering, Linyi University, Lanshan District, Linyi, Shandong, 276012, China
* Correspondence: nurfarzana.as@utm.my;

Received: 19.01.2026; Accepted: 16.05.2026; Published: 15.08.2026

Abstract: Algae-based CO_2 capture technology, specifically from the *Chlorella* genus algae, has emerged as a highly effective and feasible natural system for CO_2 capture with minimal environmental consequences. However, optimizing the algae's cultivation conditions, such as temperature, initial pH, and inlet CO_2 concentration required for an ideal CO_2 fixation rate, remains challenging due to the parameters' complex and nonlinear interdependencies. Artificial neural networks (ANNs), combined with optimization tools such as the genetic algorithm (GA) and the Bayesian optimization algorithm (BOA), are widely used to simulate biological processes. In this study, an ANN model combined with optimization tools (GA or BOA) is employed to predict algal CO_2 fixation rates using temperature, initial CO_2 concentration, and initial pH as inputs. A total dataset of 46 experimental observations from peer-reviewed studies (2010–2025) was used to train the model. The models generated from this study, ANN standalone, ANN optimized with GA (ANN-GA), and ANN optimized with BOA (ANN-BOA), were compared in terms of the fitness performance (root mean squared error (RMSE), Mean Absolute Error (MAE), and correlation coefficient (R^2)). It is found that the ANN optimized with GA has the best fitness performance across the overall datasets (R^2 = 0.8495, RMSE = 0.1384, MAE = 0.1038). The ANN-GA model also identified the best operating conditions for a high CO_2 fixation rate (pH = 7.0-8.5, temperature = 30-35°C, CO_2 inlet concentration = 5-10 %). The findings from this study demonstrate moderate predictive performance under controlled conditions, providing proof-of-concept support for data-driven optimization and a foundation for future work before scaling up.

Keywords: CO_2 fixation rate; *Chlorella* genus; artificial neural network; machine learning; CO_2 capture.

1. Introduction

The rising carbon dioxide (CO_2) emission in the atmosphere is a major driver of climate change, with fuel combustion from transportation contributing approximately 24% of worldwide CO_2 emissions in 2024 [1]. In 2022, Malaysia ranked 10th among Asia Pacific countries in CO_2 emissions, producing 241.142 metric tons of CO_2 per year [2]. These trends underscore the need for efficient, sustainable CO_2 capture technology. Existing carbon capture

technologies, such as chemical looping combustion (CLC), post-combustion carbon capture (PCC), and pre-combustion carbon capture, are costly and require substantial energy inputs for operation and maintenance [3]. Furthermore, these technologies are primarily designed for large industrial facilities, limiting their adaptability to decentralized or small-scale applications [4]. As a result, biological CO_2 capture using microalgae has emerged as a promising complementary approach.

Microalgae have emerged as a highly effective natural system for CO_2 capture. These photosynthetic microorganisms absorb CO_2 from the environment via fixation and convert it into valuable biomass, which can be utilized for biofuels, animal feed, and bioproducts, thereby promoting a circular economy [5]. Hence, it offers significant potential to mitigate climate change and advance carbon neutrality [6]. Among various strains, species within the *Chlorella* genus are frequently reported for their relatively high CO_2 fixation efficiency and rapid growth rate [7]. Table 1 compares the CO_2 capture capability of *Chlorella* sp. with other microalgae species, which is measured by CO_2 fixation rate (g of CO_2 L^{-1} day^{-1}).

Table 1. Comparison between species of microalgae and their CO_2 fixation rate.

Species	CO_2 fixation rate (g of CO_2 L^{-1} day^{-1})	Reference
Spirulina LAMB171	0.424	[8]
Chlorella sp. (AG10133)	0.957	[9]
Anabaena variabilis (AG10064)	0.14	
Parachlorella kessleri	0.21	[10]
Dunaliella salina	0.78	[11]

To fully utilize Chlorella algae's CO_2 capture capability, the operational parameters for cultivation must be maintained at optimal levels. Parameters such as temperature, light intensity, inlet CO_2 concentration, nutrient availability, and initial pH interact in a highly coupled, nonlinear manner to govern carbon fixation efficiency. Although optimal ranges are reported (CO_2 inlet concentration: 5–15%, temperatures: 20–30°C, initial pH: 7.0–8.5), these values are context-specific and often exhibit interaction effects rather than independent responses. Such multivariable coupling complicates mechanistic modeling, particularly when biological regulation and mass transfer limitations occur simultaneously. This highlights the need for predictive tools that can capture nonlinear relationships under controlled laboratory conditions.

Mathematical models, such as mass transfer and kinetic models, are commonly used to simulate the CO_2 capture capability of algae. For example, a dynamic mass-transfer-based model can be used to predict CO_2 removal by *Chlorella vulgaris* based on microalgal biomass growth, CO_2 concentration in the gas phase, and liquid phase [12]. The Monod-type kinetic model can predict CO_2 capture by the microalgae based on biomass formation, considering limitations from initial pH, inorganic carbon, ammonium, nitrate, and phosphate [13]. While mathematical models provide mechanistic insights into microalgal CO_2 capture, their predictive capability is often constrained by simplifying assumptions, parameter sensitivity, and limited ability to represent complex multivariate interactions. Monod-type models typically treat substrate limitation independently and assume steady or quasi-steady conditions, while mass-transfer models rely on fixed coefficients that may vary with mixing intensity, biomass concentration, and gas flow dynamics. As a result, mathematical models struggle to represent the strongly coupled, nonlinear interactions among operating parameters that govern CO_2 biofixation in real photobioreactor systems.

Computational Fluid Dynamics (CFD) partially addresses these limitations by integrating hydrodynamics, gas–liquid mass transfer, light attenuation, and CO_2-dependent growth kinetics within a defined photobioreactor geometry [14]. Compared to mathematical models, CFD models offer detailed physical insight for reactor design and process understanding, particularly in identifying mass transfer limitations and optimizing aeration or mixing strategies. However, the model is computationally intensive, requires extensive parameterization (e.g., kinetic constants, bubble size, transport coefficients), and is constrained by simplifying assumptions such as fixed bubble diameter and idealized light attenuation, which may reduce predictive robustness under dynamic or scaled-up conditions. Moreover, CFD models are typically reactor-specific and require reconfiguration for different geometries or operating regimes. Hence, both mathematical and CFD models face difficulties in delivering robust, rapid, and transferable predictions under dynamic cultivation environments. This highlights a critical gap: a modeling framework capable of capturing high-dimensional nonlinear interactions without imposing rigid mechanistic assumptions. A data-driven Artificial Intelligence (AI) modeling framework can likely capture the nonlinear interactions among temperature, inlet CO_2 concentration, and pH to predict *Chlorella* CO_2 fixation rates.

Among the AI-based model frameworks, Artificial Neural Network (ANN) is widely used in algae research for biomass production [15], biogas yield [16], and lipid production [17]. ANNs offer a data-driven alternative that can learn complex, nonlinear relationships from historical and real-time data. This enables the integration of multiple variables, including light intensity, nutrient concentrations, CO_2 levels, and environmental conditions, without requiring exact mechanistic equations. By reducing reliance on explicit kinetic assumptions and intensive parameter estimation, AI models enhance predictive accuracy, adaptability, and operational flexibility, particularly in dynamic cultivation systems. Therefore, while mechanistic models remain essential for mechanistic interpretation and theoretical understanding, AI-based approaches offer superior predictive and optimization capabilities for microalgal CO_2 capture performance in real-world applications. However, the predictive power of ANN models depends critically on the optimization method. Training an ANN involves navigating the high-dimensional, non-convex loss landscape to achieve a solution that generalizes beyond the training data. The choice of optimizer is decisive; it dictates whether the model captures true underlying phenomena or merely encodes the noise and biases specific to the training sample. Consequently, optimization algorithms play a central role in determining the optimal ANN architecture (number of neurons, number of hidden layers) for robust predictive performance.

Many studies have tried to improve ANN predictions by integrating them with advanced optimization tools. For example, [18] combined ANN with Genetic Algorithm (GA) to forecast the CO_2 fixation rate of microalgae based on temperature, nitrogen content, and initial pH and achieved a high correlation coefficient (R^2) value of 0.97865. Similarly, [19] employed an ANN framework combined with GA optimization to predict the microalgae biomass cultured in domestic wastewater, reporting an R^2 value of 0.9947 for the overall dataset. Another study used a Bayesian optimization algorithm (BOA) with an ANN to predict the CO_2 fixation rate of Chlorella algae based on temperature, light-dark cycle, and nitrogen-to-phosphorus ratio, yielding an R^2 value of 0.845 [20]. While these individual results are promising, this literature often treats the choice of optimizer as secondary, provided the model achieves a satisfactory R^2 result. This represents a critical gap: the optimization algorithm itself is a variable that fundamentally influences the ANN's learning process and predictive performance.

GA and BOA use fundamentally different search mechanisms. GA performs global exploration through evolutionary operators such as mutation and crossover, promoting diversity but potentially requiring extensive computational evaluations. In contrast, BOA constructs a probabilistic surrogate model of the objective function to guide sampling toward promising regions of the search space, potentially improving sample efficiency but at the cost of sensitivity to surrogate assumptions [21]. These mechanistic differences can significantly impact convergence behavior, bias–variance balance, and generalization stability of ANN models trained. A study by [22] demonstrated that GA and BOA yield distinct ANN hyperparameter configurations, which substantially affect predictive performance. However, comparative studies examining GA and BOA for ANN optimization in the specific context of predicting algal CO_2 fixation rates remain scarce. Continuing to use these tools interchangeably undermines the reliability and reproducibility of AI applications in this field, leaving it unclear which approach is more suitable for modeling algae-based systems.

To address this gap, the present study evaluates GA and BOA as the primary variables of interest by integrating them separately with an identical ANN model. The objective is to predict algal CO_2 fixation rates using temperature, initial CO_2 concentration, and initial pH as inputs. This study directly addresses this gap by treating the optimization algorithm as the main variable of interest. It is hypothesized that selecting between GA and BOA will significantly influence the ANN's predictive performance and generalization. This study enables a critical comparison of the two optimizers by examining how their distinct search strategies affect model convergence, complexity, and predictive performance. The findings of this study will establish a clear, data-driven framework for optimizer selection, enhancing the reliability and real-world utility of AI-driven approaches in algal carbon capture technology.

2. Materials and Methods

The methodology of this study is structured into two main phases: dataset construction, followed by ANN model development and optimization. In the first phase, the research scope was defined to establish clear boundaries and objectives. Subsequently, a screening process was conducted to identify and extract data from relevant papers. The extracted data were compiled into a database to serve as the foundation for the ANN model.

In the second phase, a feed-forward backpropagation ANN was developed in MATLAB using a 70/20/10 data split for training, testing, and validation. After evaluating the baseline ANN model's performance using the coefficient of determination (R^2), Mean Absolute Error (MAE), and root mean square error (RMSE), the model was separately optimized using GA and BOA to compare their resulting accuracies directly. The overall methodology, including database construction and ANN model development, is illustrated in Figure 1.

2.1. Dataset construction.

2.1.1. Scope of review.

The papers collected in this study were limited to the following scope of study:

Cultivation method – Only papers that used closed cultivation systems (e.g., bioreactors, shake flasks) were considered, as these systems are not directly affected by external climate factors such as sunlight intensity or rainfall. Open pond systems without roofing were excluded.

Microalgal species – The study focused exclusively on *Chlorella* species (e.g., *Chlorella* sp., *Chlorella vulgaris*), as CO_2 fixation capacity varies considerably among species (see Table 1).

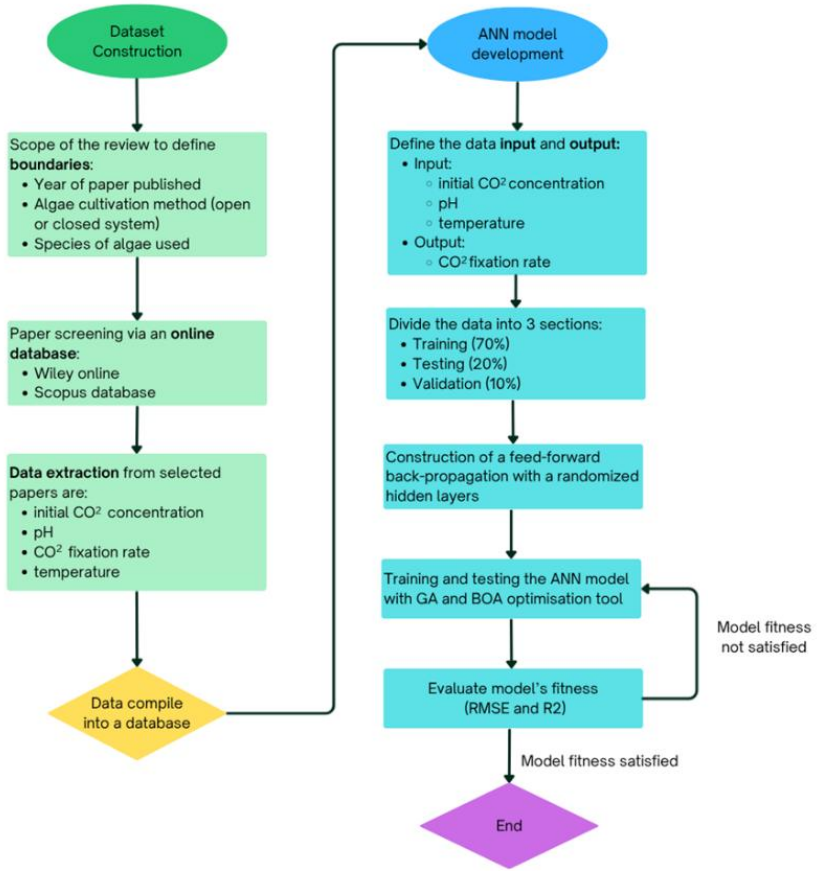

Figure 1. Overall methodology for database construction and ANN model development.

2.1.2. Paper screening.

A paper screening process was conducted to collect the data for the model training process. The papers were screened and collected on the Web of Science and Scopus databases, following established bibliographic search protocols [23]. The bibliographic search focused on research articles published between 2000 and 2025. The paper screening process criteria were as follows:

Advance search: The keywords used included: "algae biomass", "biofuel", "CO_2 fixation", "algae for CO_2 fixation", "algae for biomass production", "algae production", "algae process design", "carbon capture", "concentrated CO_2", "CO_2 utilization", "atmospheric CO_2 capture" "microalgae", and "carbon fixation"

Title and abstract: The title or abstract must contain "algae" or "CO_2 capture". Papers with titles and/or abstracts that are not related to CO_2 fixation were excluded.

2.1.3. Data extraction.

After identifying potentially eligible papers, the information from each article was individually reviewed against the inclusion and exclusion criteria. The following data were extracted: data input: (i) initial pH, (ii) temperature (°C), (iii) CO_2 inlet concentration (%), and data output: CO_2 fixation rate (g of CO_2 L^{-1} day^{-1}). The dataset was stored in the database, which was used for the ANN model construction phase.

2.2. ANN model construction and optimization.

2.2.1. Data preprocessing.

All input and output variables collected from the literature are normalized to ensure uniform scaling and improve the numerical stability and convergence performance of the ANN model. Data normalization is a crucial preprocessing step, as variables with different magnitudes may indirectly influence the learning process, jeopardizing the model's performance.

In this study, the min-max normalization method was used; this approach has been successfully applied in similar microalgal modeling studies to preserve data distribution [17]. All variables were scaled into a bounded range to improve numerical stability and prevent features with larger magnitudes from dominating the learning process [24]. Eq. (1) shows the general min-max normalization equation:

$$X_i^N = a + \frac{(x_i - x_{min})(b-a)}{x_{max} - x_{min}} \quad (1)$$

Where, x_i is the original value of the variable, x_{min} is the minimum variable range, x_{max} is the maximum variable range, X_i^N is the normalized value of the variable, a is the lower bound, and b is the upper bound. The data transformation was conducted using the [0,1] range, where the a = 0 and b = 1. After substituting the a and b values into Eq. (1), a simplified Eq. (2) was formed:

$$X_i^N = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (2)$$

The [0, 1] range is commonly adopted in min–max normalization due to its ability to provide bounded, dimensionless data while preserving the original distribution [17]. Such standardization is crucial for optimizing ANN performance by ensuring consistent feature scales [24].

2.2.2. ANN baseline model construction.

An ANN model construction was conducted based on the methods used by [19]. Three input parameters were considered in the ANN construction: (i) initial pH, (ii) temperature (°C), (iii) CO_2 inlet concentration (%), followed by one output variable: CO_2 fixation rate (g of CO_2 L^{-1} day^{-1}). The dataset compiled from section 2.1.3 was divided randomly into three subsets: 70% for training, 20% for testing, and 10% for validation. A feed-forward network and error backpropagation were employed in the ANN construction. Figure 2 illustrates the architecture of the ANN model.

ANN architecture consists of an input layer (representing the number of input variables), a hidden layer (the orange dots represent the neurons), and an output layer (representing the number of output variables). Although the flexibility of the model could be enhanced by increasing the number of hidden layers, two layers are fundamentally adequate for complex functions based on the universal approximation theorem [25].

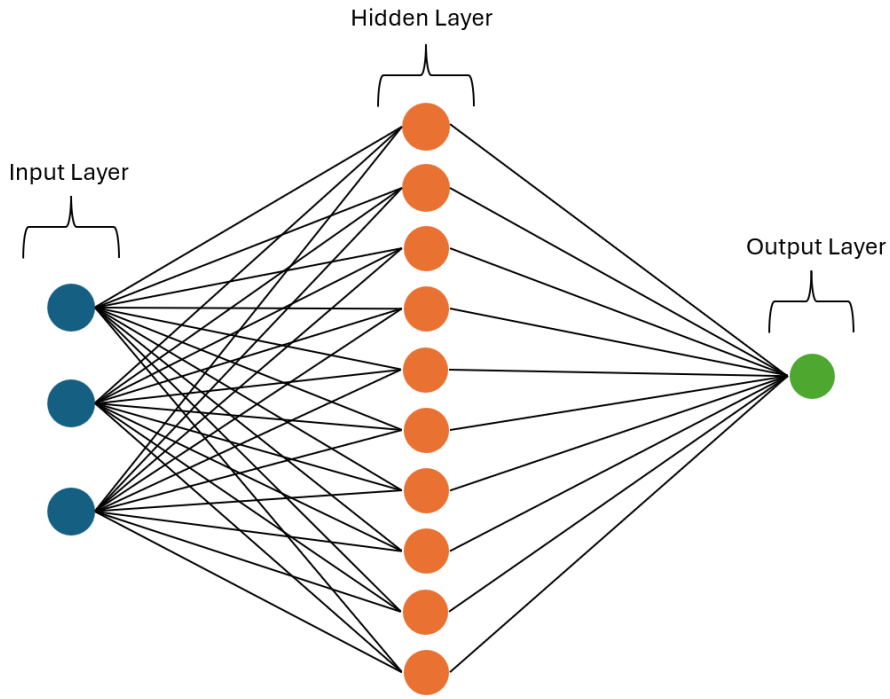

Figure 2. Example of the ANN model architecture used in predicting the CO_2 fixation rate by *Chlorella* species algae.

The upper limit of the number of hidden neurons required for the ANN model development was determined using the constraint described in Eq. (3):

$$N \leq \frac{N^{TR}}{N^{I}+1} \qquad (3)$$

Where N is the upper limit of the number of neurons in the hidden layer, N^{TR} is the number of training samples, and N^{I} is the number of inputs. Based on this criterion, the upper limit for hidden neurons was set at 14. The optimal number of hidden neurons was subsequently identified through trial and error to minimize model error. The neuronal connection between the layers is represented by biases and weights as shown in Eq. (4):

$$Y_j = \sum(X_i W_{ij}) + b_j \qquad (4)$$

Where Y_j represents the net input (or weighted sum) arriving at the j^{th} neuron in each layer, W_{ij} denotes the weight connection from neuron i to neuron j, X_i is the input at neuron i, b_j is the bias of neuron j, then the activation at the j^{th} neuron. Using the dataset collected in section 2.1.4, the networks were trained using three widely used backpropagation algorithms: Levenberg–Marquardt (trainlm), Bayesian regularisation (trainbr), and scaled conjugate gradient (trainscg). Each network configuration was trained ten times independently to minimize the influence of random weight initialization. This algorithm minimizes the error between the training data and the network output by adjusting its weights and biases.

During training, the hidden layer used the hyperbolic tangent sigmoid (tansig) activation function, which enables nonlinear mapping between the inputs and outputs. The output layer used a linear activation function (purelin), suitable for continuous regression. Training was terminated according to MATLAB's default stopping criteria, including a maximum of 1000 training epochs, a minimum performance gradient threshold of 1×10^{-7}, or

six consecutive validation failures, indicating that no further improvement in model performance could be achieved.

The ANN model performance during optimization was evaluated using K-fold cross-validation. The dataset was divided into K subsets, with each subset used once as the test set and the remaining subsets used for training. The predictive performance of the ANN was assessed using three statistical indicators: the coefficient of determination (R^2), Mean Absolute Error (MAE), and the root mean square error (RMSE). The calculations were conducted using MATLAB version 2024a (Mathworks Inc., Natick, USA) as shown in Eqs. (5), (6), and (7):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i^{actual} - Y_i^{predicted})^2} \quad (5)$$

$$MAE = \frac{1}{N}\left(\sum_{i=1}^{N}\left|Y_i^{actual} - Y_i^{predicted}\right|\right) \quad (6)$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}} \quad (7)$$

Where Y_i^{actual} is an actual value obtained from the literature (dataset), $\bar{Y}^{actual}$ is the mean actual value, $Y_i^{predicted}$ is the predicted value by the model, N is the number of datasets, SS_{res} is the sum of squared errors, and SS_{total} is the total sum of squares, in this case, the variance of the actual data.

2.2.3. ANN hyperparameter optimizations.

The baseline ANN model was further optimized by tuning its hyperparameters (specifically, the number of layers and neurons) to improve the overall architecture [26]. This step is crucial for identifying the optimal configuration to accurately predict the CO_2 fixation rate of *Chlorella* species. To achieve this, two distinct optimizers were employed: the Genetic Algorithm (GA) and the Bayesian optimization algorithm (BOA).

2.2.3.1. Genetic algorithm (GA) optimization.

GA is an evolutionary search algorithm that mimics natural selection and genetic science based on Charles Darwin's theory of natural evolution [27]. To overcome the limitations of random initialization and prevent the model from becoming trapped in local minima, GA was employed to optimize the Artificial Neural Network (ANN) globally. In this study, rather than altering the network architecture, each 'chromosome' in the population represented a specific set of initial weights and biases bounded between -1 and 1. The performance of these chromosomes was evaluated using a predefined fitness function within a 5-fold cross-validation framework, which computed RMSE, MAE, and R2 in the normalized space to estimate survival probability. Chromosomes demonstrating higher fitness (minimizing RMSE and MAE while maximizing R^2) were selected for reproduction. The configuration of the GA was conducted based on the parameters set by [28]. The evolutionary search was configured with a population of 80 individuals, evolving for up to 150 generations. The algorithm iteratively refined the network parameters by utilizing a crossover fraction of 0.85, a Gaussian mutation function to introduce genetic diversity, and an elite count of 4 to preserve the highest-performing configurations across generations. This cyclical process continued until the algorithm converged on the optimal weights and biases for predictive accuracy [29]. Once the ideal configuration was identified, the network underwent a final fine-tuning phase to achieve precise local convergence, using the Levenberg-Marquardt backpropagation algorithm

for up to 500 epochs. Table 2 summarises the GA parameter configurations and settings used in this study.

Table 2. GA parameter configurations and settings used in this study (the parameters are adapted from [28]).

Parameter	Value/Setting
Population size	80
Maximum generations	150
Crossover fraction	0.85
Mutation function	Gaussian Mutation (Scale: 0.3, Shrink: 0.5)
Elite count	4
Maximum objective evaluations	60
ANN tuning training algorithm	Levenberg–Marquardt
Maximum ANN epochs	500
Model evaluation strategy	5-fold Cross-Validation
Training repetitions	3
Performance metrics	R^2, MAE, and RMSE

2.2.3.2. Bayesian optimization algorithm (BOA).

The BOA determines the next search point based on knowledge gained from previous iterations. The BOA procedure conducted in this study is based on [30]. BOA was employed to optimize the number of neurons in a two-layer hidden architecture (h_1 and h_2). These variables were initially treated as continuous values during the optimization process, but were rounded to the nearest integer when constructing the ANN model. The search space for the first hidden layer (h_1) is constrained to [1, 14] neurons, while the second layer (h_2) ranges from [0, 14] neurons, where a value of zero denotes a single-layer configuration. The optimizer is restricted to 60 objective evaluations to ensure computational efficiency. BOA uses two main components: the surrogate model and the acquisition function. The surrogate model estimates the objective function using a Gaussian Process (GP). The GP approximates the objective function *f(x)*, which is expressed as the relationship between inputs and outputs through a mean function *m(x)* and the covariance kernel function (*k*), as described in Eq. (8):

$$f(x) \sim GP\ (m, k(x_i x_j) \qquad (8)$$

The acquisition function uses the surrogate model's predictions and uncertainty estimates. The function continuously updates and maximizes with iterations, effectively balancing exploration (searching unknown areas) and exploitation (refining promising regions). The mathematical expression for hyperparameter optimization using BOA is given as Eq. (9):

$$x^* = \arg\min f(x),\ _{x \epsilon X} \qquad (9)$$

Where $f(x)$ denotes an objective score to lower the error rate on the validation set, x^* denotes the set of hyperparameters that provides the low grade, and x is any value of the space X. This process guides the BOA optimization more strategically toward the global minimum.

The ANN model was retrained using a feedforward neural network architecture created with MATLAB's Neural Network Toolbox. The network was trained using the Levenberg–Marquardt backpropagation algorithm (trainlm) with a training limit of 500 epochs per iteration. To ensure robust model evaluation and prevent overfitting during the optimization process, K-fold cross-validation was applied. The normalized dataset consisting of input variables (X_n) and target variables (T_n) was divided into K equally sized subsets using MATLAB's "crossvalind('Kfold', N, K)" function, where N represents the total number of

samples. For each fold k, the k^{th} subset was used as the testing dataset while the remaining K−1 subsets were used for model training. This process was repeated until all subsets had served once as the test dataset. Each ANN model within a cross-validation fold was trained three times independently. For each initialization, the network weights were reset using the "init()" function before training. Within each fold, the network is trained using the "trainlm" function across three independent initializations (nInit = 3), with the mean performance across these runs serving as the objective value.

The BOA minimizes a composite weighted loss function designed to maximize the R^2 while minimizing the RMSE and MAE as shown in Eq. (10):

$$Objective = \alpha\left(1 - \overline{R^2}\right) + \beta\left(\overline{RMSE}\right) + \gamma\left(MAE\right) \quad (10)$$

Where weights are set to α = 0.5, β = 0.2, and γ = 0.3 to prioritize the goodness-of-fit.

Upon identifying the optimal hyperparameters (h_1 and h_2), the final ANN is trained for 500 epochs to ensure convergence. Similar to GA, the BOA objective was evaluated using a 5-fold Cross-Validation framework. The model's predictive performance is then reported as the mean and standard deviation of R^2, RMSE, and MAE across the cross-validation folds to provide a statistically significant measure of model reliability. Table 3 summarises the BOA parameter configurations and settings used in this study.

Table 3. BOA parameter configurations and settings used in this study (the parameters are adapted from [30]).

Parameter	Value/Setting
Surrogate model	Gaussian Process Regression
Acquisition function	Expected Improvement (EI)
Optimisation variables	Hidden layer neurons (h_1, h_2)
Objective function	$Objective = \alpha\left(1 - \overline{R^2}\right) + \beta\left(\overline{RMSE}\right) + \gamma\left(MAE\right)$
Maximum objective evaluations	60
ANN training algorithm	Levenberg–Marquardt
Maximum ANN epochs	500
Model evaluation strategy	5-fold Cross-Validation
Training repetitions	3
Performance metrics	R^2, MAE, and RMSE

2.3. Wilcoxon signed-rank test and sensitivity analysis.

To determine whether the performance differences (R^2, RMSE, and MAE) between the ANN, ANN-GA, and ANN-BOA models were statistically significant, a non-parametric Wilcoxon signed-rank test was conducted across the cross-validation folds. This test was selected because performance metrics derived from cross-validation folds often violate parametric assumptions of normality and strict independence; ranking differences provides a more robust, conservative comparison of the hybrid models' predictive capabilities [31]. The statistical significance threshold was set at $\alpha = 0.05$.

To quantify the individual contribution of each input variable to the model's predictions, a perturbation-based sensitivity analysis (PSA) was conducted. A baseline input vector was used, using the mean of the normalized dataset. Each independent variable was systematically perturbed by ±10% of its normalized range, while maintaining all other variables at their baseline values. Because the model utilizes a polynomial feature expansion, perturbations in a primary variable were simultaneously reflected in its corresponding squared and interaction terms before forward propagation through the optimized ANN. The sensitivity score (S_i) for each variable i was calculated as the absolute difference in the predicted output, as shown in Eq. (11)

$$S_i = |Y(x_i + \Delta) - Y(x_i - \Delta)| \qquad (11)$$

Where Y is the predicted output generated by the model, x_i is the value of the independent variable, and Δ is the systematic perturbation applied to the variable, in this case, 10%. Final results were expressed as 'Relative Importance' (%), representing the normalized impact of each variable relative to the total sensitivity of the system. This analysis provides a transparent view of which physical parameters most heavily influence the target outcome, ensuring the model's physical interpretability.

3. Results and Discussion

3.1. Dataset construction.

Figure 3 shows the process flow chart for paper screening, inclusion and exclusion, and data extraction, which were used for the ANN analysis.

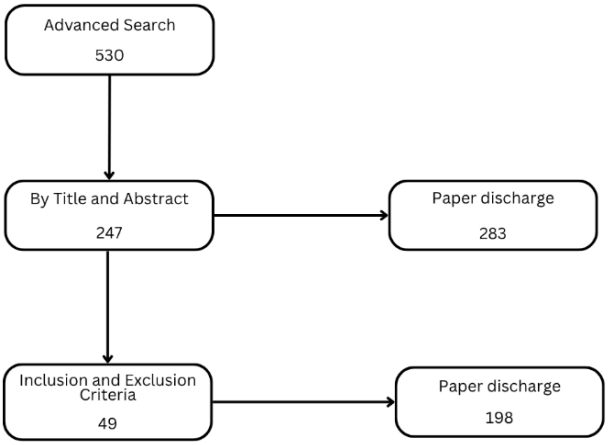

Figure 3. Flow chart of the paper screening, inclusion, exclusion, and data extraction process.

Referring to Figure 3, initially, 530 papers were obtained from Scopus and Web of Science. Subsequently, after filtering the keywords to "algae" or "CO_2 capture" in the abstract section, only 247 papers met the criteria. The 247 papers were re-evaluated to determine if they met the inclusion and exclusion criteria, and it was found that only 49 papers met the criteria. Values that are out of the ideal growth conditions of the *Chlorella* algae (pH = 6 – 9; temperature 25 °C – 45 °C; CO_2 inlet concentration = 0.5 % – 20 %) would be excluded from the review process, leaving only 46 papers [7]. The advantage of ANNs is that they can effectively work with small datasets (below 50), as demonstrated by [18] and [20]. The final data used for this review are listed in Table 4. Statistical analysis of the collected datasets, including mean, range, standard deviation, and variance, is presented in Table 5.

Table 4. The CO_2 fixation rate by *Chlorella* species algae at different operating parameters (temperature, initial pH, and CO_2 inlet concentration).

Reference	Species	CO_2 fixation rate (g of CO_2 L^{-1} day^{-1})	Temperature (°C)	Initial pH	CO_2 inlet (%)
[32]	*Chlorella pyrenoidosa Chick (IPPAS C2)*	0.25	21.4	7	9.5
[33]	*Chlorella sorokiniana*	0.421	35	7	15.65
[34]	*Chlorella vulgaris (BEIJ 1890)*	0.476	25	7	10
[35]	*Chlorella minutissima*	0.26	25	7.8	4
[36]	*Chlorella vulgaris ISC-23*	0.44	27	7.4	6
[37]	*Chlorella sp*	0.26	25	7	10
[38]	*Chlorella sorokiniana GS03*	0.66	28	7.2	5
[39]	*Chlorella sp.*	0.43	30	7	10

Reference	Species	CO_2 fixation rate (g of CO_2 L^{-1} day^{-1})	Temperature (°C)	Initial pH	CO_2 inlet (%)
[40]	*Chlorella vulgaris ESP-31*	0.2	28	7	25
[41]	*Chlorella sp.*	0.1	30	7	15
[42]	*Chlorella sp. BTA 9037*	0.183	25	7	5
		0.176	25	7	15
		0.123	25	7	5
[43]	*Chlorella sorokiniana TH01*	0.47	28	7.4	5
[21]	*Chlorella pyrenoidosa*	0.26	25	7	10
[44]	*Chlorella Pyrenoidosa*	0.12	25	7	5
		0.125	25	7	10
		0.052	25	7	0.04
[45]	*Chlorella vulgaris.*	0.096	28	6.8	1
		0.106	28	6.8	2
		0.163	28	6.8	3
		0.16	27	4	5
[46]	*Chlorella pyrenoidosa (SJTU-2)*	0.244	25	6.8	5
		0.26	25	8	10
[9]	*Chlorella sorokiniana (AG20740)*	0.14	25	8.4	15
	Chlorella sp. (AG10133)	1.785	30	8.4	15
		1.486			10
		0.957			5
[47]	*Chlorella sorokiniana (CS-01)*	0.23	26	5.8	5
		0.24			10
[48]	*Chlorella sorokiniana*	0.55	25	6.5	5
[49]	*Chlorella vulgaris (CCAP 211/11D)*	0.25	25	7	4
		0.28	25	7	8
[50]	*Chlorella sp. (PTCC 6010, Persian Type Culture Collection)*	0.12	26	7	9.45
		0.17	26	7	1.75
[51]	*Chlorella sp. (ABC-001)*	0.683	30	7	10
[52]	*Chlorella vulgaris*	0.225	25	7	15
	Chlorella sorokiniana	0.192	30	7	20
[53]	*Chlorella sp.*	0.256	18	8	0.03
		1.378	18	8	0.03
[54]	*Chlorella sp. NCTU-2*	1.147	26	6	10
[55]	*Chlorella sp. (AG10002)*	0.27	20	7.18	1
		0.31			2
		0.35			5
[56]	*Chlorella vulgaris*	0.318	28	6	5
[57]	*Chlorella emersonii*	0.13	25	8	15
[58]	*Chlorella vulgaris*	0.39	30	7	5
[59]	*C. sorokiniana*	0.44	25	7	10
[60]		0.67	25	7	5

Table 5. Statistical analysis of the datasets from Table 4 (mean, range, standard deviation, and variance) (Source: Data obtained from this study).

Statistical information	CO_2 fixation rate (g of CO_2 L^{-1} day^{-1})	Temperature (°C)	Initial pH	CO_2 inlet (%)
Mean	0.368	25.900	6.989	9.118
Range	0.052 – 1.378	18 - 35	4 - 8	0.03 - 25
Standard deviation	0.299	2.944	0.713	6.122
Variance	0.089	8.667	0.508	37.474

*T*he dataset in Table 4 includes multiple *Chlorella* species and strains compiled from different studies, introducing a degree of biological heterogeneity. Variations in metabolic activity, CO_2 tolerance, and environmental adaptability across species and strains may influence the observed CO_2 fixation rates. In the present study, stratified analysis was not conducted to avoid further reducing the already limited dataset used for ANN training. While this approach allows for the development of a generalized predictive model, it may also reduce the model's ability to capture species-specific behaviors, thereby affecting its generalisability.

3.2. Model fitness evaluation for ANN, ANN-GA, and ANN-BOA.

Table 6 shows the accuracy results (R^2, RMSE, MAE) from the developed models. Figure 4 shows the fitted line plots describing the relationships between the actual data from the literature and the outputs predicted by the ANN baseline, ANN-GA, and ANN-BOA.

Table 6. Accuracy results (R^2, RMSE, and MAE) from ANN, ANN-GA, and ANN-BOA)

Model	R^2	RMSE	MAE
ANN baseline	0.6423	0.2527	0.1764
ANN-GA	0.8495	0.1384	0.1038
ANN-BOA	0.7238	0.1780	0.1226

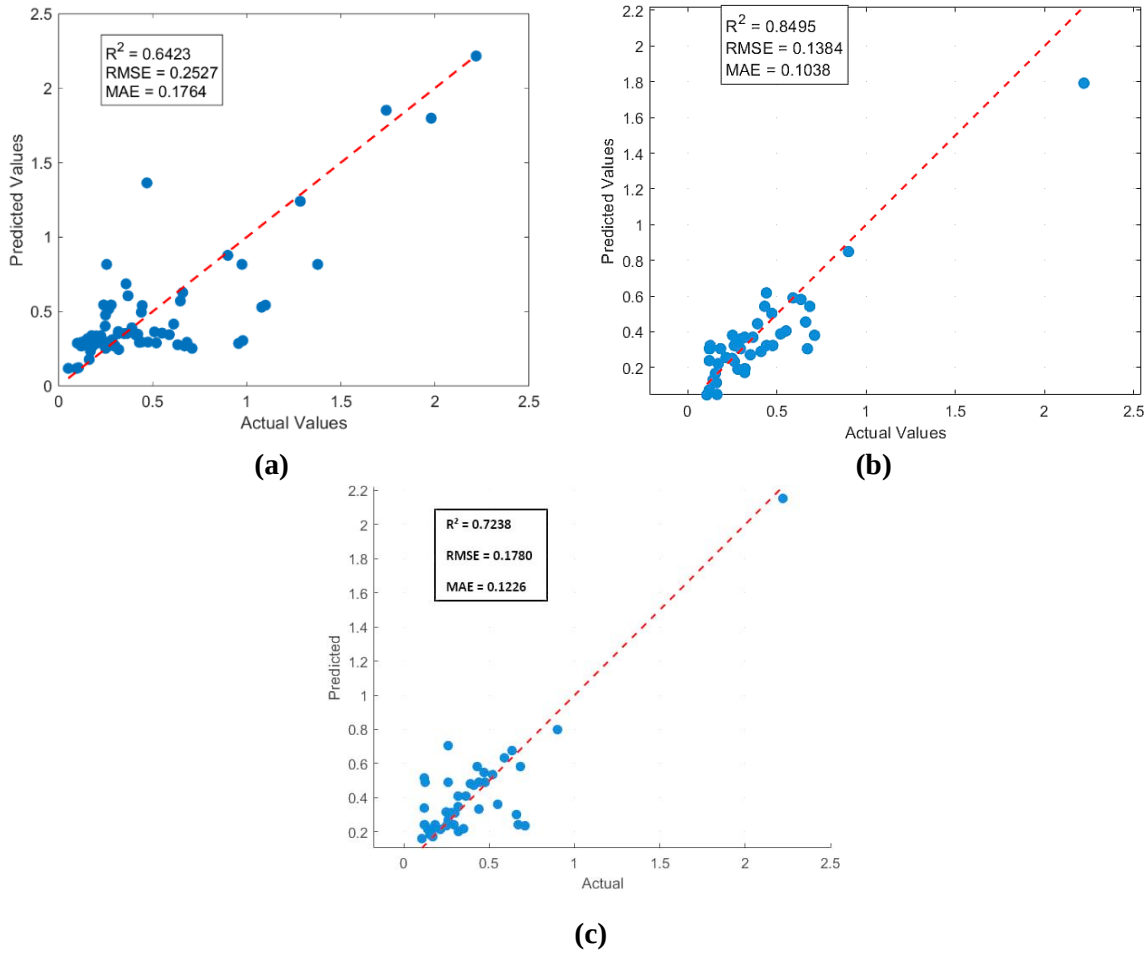

Figure 4 (a-c). Fitted line plot of the actual data from the literature vs output predicted by the (a) ANN, (b) ANN-GA, and (c) ANN-BOA. All fitted line plots display the entire dataset, comprising the training (70%), validation (20%), and testing subsets (10%).

Referring to Table 6 and Figure 4(a), the accuracy results (R^2, RMSE, and MAE) achieved from the ANN baseline model are poor compared to the optimized ANN model (ANN-GA and ANN-BOA). The poor ANN model performance is likely due to the nonlinear and interacting effects among the input parameters (CO_2 inlet concentration, pH, temperature) on the CO_2 fixation rate, which are notoriously difficult to learn from small or noisy experimental datasets. ANN baseline models tend to overfit or fail to capture higher-order interactions when training samples are limited [61]. Optimization tools such as GA and BOA can significantly improve the model by expanding the hyperparameter/weight search space and finding configurations that both improve fit and reduce generalization error on a limited dataset [62], as demonstrated by this study. Results from the Wilcoxon signed-rank test ($\alpha < 0.05$)

confirm that the improvements offered by both ANN-GA and ANN-BOA are statistically significant.

A comparative analysis of optimization frameworks reveals that ANN-GA outperforms ANN-BOA, achieving a higher R^2 value. This indicates a superior ability to map the non-linear relationships between operating parameters, specifically CO_2 inlet concentration, pH, and temperature, and the algal CO_2 fixation rate. The Wilcoxon signed-rank test ($\alpha < 0.05$) validates this result, confirming that the enhanced predictive accuracy of the GA-integrated model is statistically significant rather than stochastic variance. The success of GA optimizer is attributed to its population-based, global search dynamics. GA works by testing and improving many possible solutions simultaneously, keeping a diverse "population" of solutions and using simple biological-inspired rules such as selection, crossover, and mutation to generate new combinations [63]. This wide search enables GA to explore multiple directions in the dataset simultaneously and avoid getting stuck in local solutions. In contrast, BOA employs a sequential strategy, where a surrogate model (typically a Gaussian Process) predicts the performance of unexplored configurations. While BOA is computationally efficient for smooth objective functions, its surrogate model can become overconfident. This often results in excessive exploitation of local regions at the expense of global exploration, potentially bypassing superior distant optima [64]. Therefore, GA is more likely to find better overall (global) settings for the neural network than BOA, especially when the parameter relationship is complex and irregular [28].

In algae bioprocess modeling, achieving an R^2 of 0.85 using an optimized ANN is considered an acceptable performance since biological systems are naturally complex, highly nonlinear, multi-dimensional, and inherently noisy [15, 18, 20]. In this study, ANN-GA successfully meets this standard, demonstrating a satisfactory R^2 of 0.8495 alongside low error metrics (RMSE = 0.1384, MAE = 0.1038). In contrast, ANN-BOA exhibits limited predictive strength ($R^2 < 0.85$) despite maintaining comparably low absolute errors (RMSE = 0.1780, MAE = 0.1226). This apparent discrepancy highlights a critical limitation of the R^2 metric when applied to heavily constrained datasets. Referring to Table 5, the target variable (CO_2 fixation rate) has a narrow operating range (0.052 - 1.378 g of CO_2 L^{-1} day^{-1}) and a quantitatively low variance of 0.089. Referring to Eq(7), because R^2 measures explained variance relative to the total natural variance (SS_{total}), a low intrinsic variance results in a minuscule denominator [65]. Consequently, even minor residual errors (SS_{res}) cause a mathematically disproportionate reduction in the R^2 score. In such contexts, absolute metrics like RMSE and MAE provide a more reliable indication of true model performance than R^2 [65]. Furthermore, the configuration of the BOA optimizer contributes to the low R^2 score. BOA uses a surrogate model, such as a Gaussian process, and acquisition-function settings can bias the search toward exploitation, particularly in complex or high-dimensional landscapes [64]. This can cause BOA to converge on a locally good hyperparameter region that minimizes RMSE or MAE (or validation loss) rather than maximizing the model's ability to capture the global trend (R^2) [66]. Ultimately, this underscores how both dataset characteristics (e.g., variance and noise) and optimizer mechanics significantly dictate overall model performance.

While previous literature has extensively explored the ANN modeling of algae systems and CO_2 fixation rates, directly benchmarking the predictive metrics (R^2, RMSE, and MAE) of the current models against those in the literature is inherently flawed. Variations in algae species, cultivation conditions, bioreactor designs, dataset scales, and parameters used render quantitative cross-study comparisons invalid. Therefore, the primary contribution of this study

lies in methodological advancement rather than absolute metric superiority. Prior research has frequently relied on trial-and-error approaches and on single optimization tools for ANN architecture design in algae cultivation modeling. In contrast, this study systematically tests two advanced optimization techniques: a genetic algorithm (ANN-GA) and Bayesian optimization (ANN-BOA). By testing both optimization algorithms on the same dataset with the same ANN framework, this research provides a detailed, controlled evaluation of which method is better suited for predicting the complex behavior of biological CO_2 fixation.

The accuracy results from Table 6 and the fitted line plot (Figure 4) quantitatively demonstrate the efficacy of the ANN-GA model. While the literature notes that highly complex ANN architectures, such as ANFIS–GA [18], can achieve higher accuracy (R2 = 0.9786), these models were trained on highly restricted datasets (27 datasets), posing a severe risk of overfitting, in which the model memorizes data rather than generalizing. Therefore, the proposed ANN–GA represents a preferable and more trustworthy alternative. By evaluating the ANN-GA on a larger, more heterogeneous dataset, it successfully captures the highly nonlinear bioprocess relationships, achieving an R^2 value of 0.8495 and low RMSE (0.1384) and MAE (0.1038). The robustness of this GA-based optimization strategy is quantitatively supported by the Wilcoxon signed-rank test ($\alpha < 0.05$), proving that the model maintains statistically significant predictive stability and consistency across validation phases without the structural instability often seen in overly complex models.

3.3. Sensitivity analysis for ANN, ANN-GA, and ANN-BOA.

The sensitivity analysis evaluates the relative importance of three input parameters (temperature, Initial pH, and CO_2 inlet concentration) across the baseline ANN model and its two optimized models (ANN-GA and ANN-BOA). Table 7 summarises the sensitivity analysis results from all three models (ANN baseline, ANN-GA, and ANN-BOA).

Table 7. Sensitivity analysis results from ANN, ANN-GA, and ANN-BOA.

Model	Relative importance (%)		
	Temperature (°C)	Initial pH	CO_2 inlet concentration (%)
ANN baseline	30.80	49.2	20
ANN-GA	4.57	46.08	49.35
ANN-BOA	15.50	77.40	7.20

The results demonstrate that the choice of optimization algorithm drastically alters the weight the models assign to each parameter. The striking divergence in the relative importance assigned to input parameters across the three models is likely due to their internal parameter weighting, consequently affecting their functional logic depending on the optimization strategy employed. Referring to Table 7, the ANN baseline model has the most balanced distribution of parameter importance, placing the greatest weight on pH (49.2%) while still accounting for other parameters in making predictions. Under standard backpropagation or gradient descent training used by the baseline ANN, the model relies on a holistic combination of all three physical parameters to map the complex, non-linear biological responses in the dataset.

In terms of ANN-GA, the GA optimization completely minimizes the temperature importance (4.57%), shifting the model's focus to balance between CO_2 inlet concentration (49.53%) and initial pH (46.08%). Because GA explores the solution space using evolutionary principles (crossover and mutation), it is likely to have converged on a local optimum in which the dataset's variance could be sufficiently explained by the interplay of pH and carbon

availability. The GA effectively excludes temperature from the primary decision-making pathway to minimize the cost function.

Whereas in ANN-BOA, there is a significant shift in parameter importance, with initial pH (77.40 %) taking precedence over minimizing temperature (15.50 %) and CO2 inlet concentration (7.20 %) in CO_2 fixation rate predictions. Unlike GA, BOA builds a probabilistic surrogate model (e.g., a Gaussian Process) of the objective function to balance exploration of the weight space with exploitation of known low-error regions. In this case, the BOA optimizer clearly identified initial pH as the parameter whose variance most strictly dictated the network's error. By mathematically exploiting initial pH as the primary proxy, ANN-BOA efficiently minimized the loss function, isolating a highly specialized mathematical pathway that heavily sidelines the other two variables.

3.4. Relationship between CO_2 fixation rate by Chlorella species algae and the operating parameters.

To better understand the dynamic between operating conditions and their influence on CO2 fixation rate, a surface plot was constructed. Figure 5 illustrates the combined effect of the operating parameters (CO_2 inlet concentration, initial pH, and temperature) on the CO_2 fixation rate by the *Chlorella* species algae.

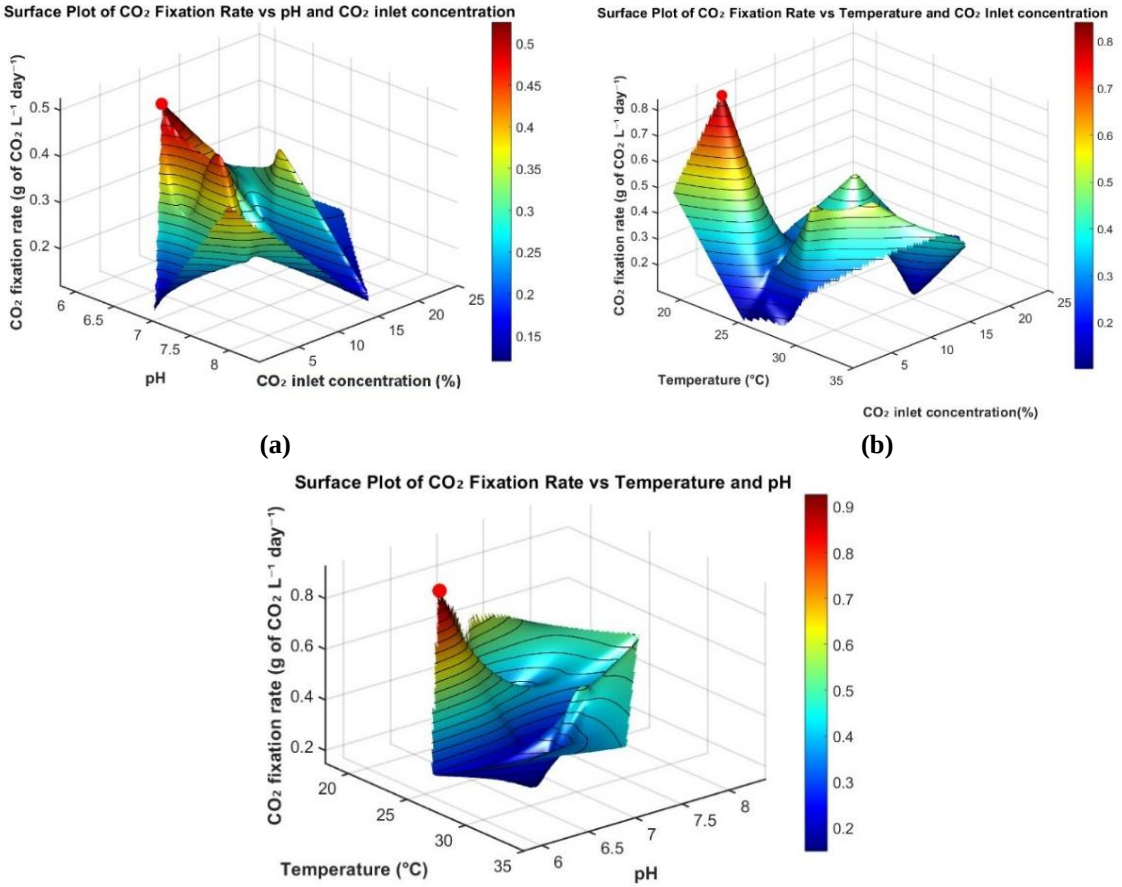

Figure 5 (a-c). Surface plot illustrating the simulated CO_2 fixation rate against temperature, pH, and CO_2 inlet concentration. The underlying ANN model generating this predictive surface was evaluated using a data partition of 70% training, 20% validation, and 10% testing subsets. Note: This surface represents the empirical interpolation of predictions generated by the trained ANN model within the experimental data range, rather than a mechanistic physical model.

The 3D surface plots, as shown in Figures 5 (a – c), visualize the predictive space of the trained ANN model. Since ANN interpolates directly from the empirical data rather than fitting a smoothed global equation, the topography of these surfaces exhibits distinct peaks and valleys. These complex shapes represent localized, non-linear interactions between variables (such as pH, temperature, and CO_2 inlet concentration), where the experimental data clusters. Regions with sharp gradients or peaks indicate high data density and high observed biological activity during the model's training. Conversely, smoother or transitional regions represent the model's interpolation across gaps in the training data space. Therefore, the visual complexity of these plots should be interpreted strictly as an empirical mapping of the experimental outcomes, rather than a definitive mechanistic law governing CO_2 fixation.

Figure 5(a) shows the interactive effects of the pH and CO_2 inlet concentration on the CO_2 fixation rate by *Chlorella* species algae. The model predicts the peak CO_2 fixation rate at approximately 0.45 g of CO_2 L^{-1} day^{-1} under neutral or acidic conditions (pH 6.5 – 7.5) combined with moderate CO_2 inlet concentration (5 – 10 %). These parameters align closely with established biological literature, which confirms that this pH range maximizes the activity of carbon-fixing enzymes such as RuBisCO [7]. A moderate level of CO_2 inlet concentration has been reported in previous studies to increase the gas–liquid concentration gradient and volumetric mass-transfer (k_La), improving the absorption of CO_2 by the algae [67]. Excessively high concentrations (beyond 15%) have been reported in the literature to induce culture acidification, which exerts an inhibitory effect on growth and overall photosynthetic efficiency, as noted in previous bioreactor optimization studies [68].

Figure 5(b) shows the synergistic effect of the CO_2 inlet concentration (%) and temperature on the CO_2 fixation rate by *Chlorella* species algae. CO_2 fixation rates scale positively with both variables up to a distinct optimal threshold: 30°C and 5% CO_2, which yields a maximum rate of 0.8 g of CO_2 L^{-1} day^{-1}. Beyond these setpoints, performance steadily declines. The model's predicted decline under high CO_2 concentrations (10% and above) is consistent with previous studies indicating that such environments may inhibit carbonic anhydrase, a key enzyme in the carbon-concentrating pathway, thereby slowing algae growth [8]. In terms of temperature, while *Chlorella* exhibits peak photosynthetic activity near 30 – 35°C, exceeding this range induces severe thermal stress. The predicted reduction in CO_2 fixation at temperatures above 35°C mirrors empirical studies showing that thermal stress reduces photosynthetic pigment content, limits chlorophyll synthesis, and decreases overall cell density [69].

Figure 5(c) illustrates the combined effect of temperature and pH on the CO_2 fixation rate by *Chlorella* species algae. The optimal operating window narrows to 30–32°C and a pH of 6.8–7.0, generating a peak fixation rate of roughly 0.8 g of CO_2 L^{-1} day^{-1}. Operating at temperatures below 25°C is likely to suppress the metabolic and enzymatic activities. On the other hand, exposure to temperatures above 35°C corresponds to known thresholds for thermal stress and photoinhibition [70]. Similarly, maintaining an optimal temperature cannot offset the penalty of an unbalanced pH. Operating at a pH outside of the optimal range (6.8-7.0) results in a sharp decline in CO_2 fixation, a phenomenon corroborated by external studies showing maximum *Chlorella vulgaris* uptake is highly restricted to 30–35°C and pH 6.5–7.5 [70]. Cultivating the algae beyond the optimum range is theorized to inhibit RuBisCO enzyme activity, thereby limiting overall photosynthesis, as highlighted in Figure 5(a).

Overall, the surface plot diagrams demonstrate that *Chlorella* species exhibit a distinct synergistic relationship among pH, temperature, and CO_2 inlet concentration and their impact

on the CO_2 fixation efficiency. These findings reinforce the importance of maintaining a balance between the three operating parameters to sustain an optimal CO_2 fixation rate. Crucially, it must be noted that the primary scope of this study is the ANN modeling of cultivation parameters and their impact on the CO_2 fixation rate. While the optimal conditions predicted by the optimized ANN models closely align with well-documented intracellular mechanisms, the specific biological pathways were not directly quantified in experimental assays in the current study. Nevertheless, the discussion of these enzymatic responses is presented as a literature-supported framework to contextualize the model outputs, rather than an experimentally validated conclusion.

4. Conclusions

The study compares the performance of GA and BOA optimization tools in enhancing the predictive accuracy of an ANN model in predicting the CO_2 fixation rate based on the operating conditions(temperature, initial pH, and CO_2 inlet concentration). It is found that ANN optimized with GA (ANN-GA) exhibits the highest fitness performance (R^2 = 0.8495, RMSE = 0.1384, MAE = 0.1038). The statistical superiority of GA over BOA in this predictive modeling framework was successfully confirmed through a Wilcoxon Signed-Rank test, demonstrating GA's robust capability in consistently minimizing prediction errors. The ANN-GA model identified the optimum parameter (pH = 7.0 - 8.5, temperature = 30 -35 °C, CO_2 inlet concentration = 5 – 10 %) required for a high CO_2 fixation rate.

The model framework proposed in this study provides a preliminary data-driven proof-of-concept for optimizing algae bioprocess systems under controlled settings. However, extending these findings as a direct guideline for the industrial upscaling of CO_2 capture technology is premature without further validation beyond the compiled dataset.

The dataset used to develop the ANN model is limited strictly to the *Chlorella* genus, meaning the framework might not simulate effectively for other algae genera. Furthermore, compiling data from various peer-reviewed sources introduces inherent dataset heterogeneity, including differences in bioreactor designs, light intensities, and culture media, which can affect the model's generalizability. Furthermore, the predictive ability of the ANN-GA model framework has yet to be validated through direct experimental trials. Future research should extend the model framework to other algal genera, incorporate larger and more diverse datasets while accounting for systemic heterogeneity, and rigorously validate the model via controlled experimental trials. Nevertheless, this study serves as an advancement in integrating AI-based modeling frameworks into algae CO_2 capture research.

Author Contributions

Conceptualization – N.H.B.Z.; Methodology – N.H.B.Z.; Software – K.Z.Y.; Validation – K.Z.Y.; Formal analysis – N.H.B.Z.; Investigation – K.Z.Y.; Resources – K.Z.Y., Y.H.H.; Data curation – K.Z.Y., Y.H.H.; Writing – original draft preparation – K.Z.Y.; Writing – review & editing – Y.M.B.M.J., Z.B.H.; Visualization – K.Z.Y.; Supervision – N.F.B.A.S.; Project administration – N.F.B.A.S.; Funding acquisition – N.F.B.A.S., N.H.B.Z., Y.M.B.M.J., Z.B.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data supporting the findings of this study are available upon reasonable request from the corresponding author.

Funding

This research was funded by Universiti Teknologi Malaysia Potential Academic Staff Grant - DTD Fund (PAS DTD), grant number R.J130000.8046.4J727.

Acknowledgments

The authors would like to acknowledge the research grant from Universiti Teknologi Malaysia, Potential Academic Staff Grant - DTD Fund (PAS DTD), with grant number R.J130000.8046.4J727.

Conflicts of Interest

The authors declare no conflict of interest.

Role of Funders

The funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Ritchie, H. Cars, planes, trains: where do CO_2 emissions from transport come from? *Our world in data* **2020**.
2. Worldometers. Available online: https://www.worldometers.info/co2-emissions/malaysia-co2-emissions/ (accessed on 30 December **2025**).
3. Yusuf, M.; Ibrahim, H. A comprehensive review on recent trends in carbon capture, utilization, and storage techniques. *J. Environ. Chem. Eng.* **2023**, *11*, 111393, https://doi.org/10.1016/j.jece.2023.111393.
4. Wilkes, M.D.; Ejeh, J.; Roberts, D.; Brown, S. Cost of small-scale dispatchable CO_2 capture: Techno-economic comparison and case study evaluation. *Int. J. Greenh. Gas Control* **2023**, *127*, 103931, https://doi.org/10.1016/j.ijggc.2023.103931.
5. Almomani, F.; Abdelbar, A.; Ghanimeh, S. A Review of the Recent Advancement of Bioconversion of Carbon Dioxide to Added Value Products: A State of the Art. *Sustainability* **2023**, *15*, 10438, https://doi.org/10.3390/su151310438.
6. Xu, P.; Li, J.; Qian, J.; Wang, B.; Liu, J.; Xu, R.; Chen, P.; Zhou, W. Recent advances in CO_2 fixation by microalgae and its potential contribution to carbon neutrality. *Chemosphere* **2023**, *319*, 137987, https://doi.org/10.1016/j.chemosphere.2023.137987.
7. Ashour, M.; Mansour, A.T.; Alkhamis, Y.A.; Elshobary, M. Usage of *Chlorella* and diverse microalgae for CO_2 capture - towards a bioenergy revolution. *Front. Bioeng. Biotechnol.* **2024**, *12*, 1387519, https://doi.org/10.3389/fbioe.2024.1387519.
8. Zhu, B.; Xiao, T.; Shen, H.; Li, Y.; Ma, X.; Zhao, Y.; Pan, K. Effects of CO_2 concentration on carbon fixation capability and production of valuable substances by *Spirulina* in a columnar photobioreactor. *Algal Res.* **2021**, *56*, 102310, https://doi.org/10.1016/j.algal.2021.102310.
9. Park, J.; Kumar, G.; Bakonyi, P.; Peter, J.; Nemestóthy, N.; Koter, S.; Kujawski, W.; Bélafi-Bakó, K.; Pientka, Z.; Muñoz, R.; Kim, S.-H. Comparative Evaluation of CO_2 Fixation of Microalgae Strains at

Various CO_2 Aeration Conditions. *Waste Biomass Valorization* **2021**, *12*, 2999-3007, https://doi.org/10.1007/s12649-020-01226-8.

10. Beigbeder, J.B.; Lavoie, J.-M. Effect of photoperiods and CO_2 concentrations on the cultivation of carbohydrate-rich *P. kessleri* microalgae for the sustainable production of bioethanol. *J. CO_2 Util.* **2022**, *58*, 101934, https://doi.org/10.1016/j.jcou.2022.101934.
11. Kim, W.; Park, J.M.; Gim, G.H.; Jeong, S.-H.; Kang, C.M.; Kim, D.-J.; Kim, S.W. Optimization of culture conditions and comparison of biomass productivity of three green algae. *Bioprocess Biosyst. Eng.* **2012**, *35*, 19-27, https://doi.org/10.1007/s00449-011-0612-1.
12. Lababpour, A. A dynamic model for the prediction of flue gas carbon dioxide removal by the microalga *Chlorella vulgaris* in column photobioreactor. *Alexandria Eng. J.* **2018**, *57*, 3311–3320, https://doi.org/10.1016/j.aej.2018.01.013.
13. Eze, V.C.; Velasquez-Orta, S.B.; Hernández-García, A.; Monje-Ramírez, I.; Orta-Ledesma, M.T. Kinetic modelling of microalgae cultivation for wastewater treatment and carbon dioxide sequestration. *Algal Res.* **2018**, *32*, 131-141, https://doi.org/10.1016/j.algal.2018.03.015.
14. Ojaniemi, U.; Tamminen, A.; Syrjänen, J.; Barth, D. CFD modeling of CO_2 fixation by microalgae cultivated in a lab scale photobioreactor. *Bioresour. Technol.* **2025**, *415*, 131715, https://doi.org/10.1016/j.biortech.2024.131715.
15. Liyanaarachchi, V.C.; Nishshanka, G.K.S.H.; Nimarshana, P.H.V.; Ariyadasa, T.U.; Attalage, R.A. Development of an artificial neural network model to simulate the growth of microalga *Chlorella vulgaris* incorporating the effect of micronutrients. *is J. Biotechnol.* **2020**, *312*, 44-55, https://doi.org/10.1016/j.jbiotec.2020.02.010.
16. de Queiróz Lamas, W. Algae's potential as a bio-mass source for bio-fuel production: MLR vs. ANN models analyses. *Fuel* **2025**, *395*, 134853, https://doi.org/10.1016/j.fuel.2025.134853.
17. Liyanaarachchi, V.C.; Nishshanka, G.K.S.H.; Sakarika, M.; Nimarshana, P.H.V.; Ariyadasa, T.U.; Kornaros, M. Artificial neural network (ANN) approach to optimize cultivation conditions of microalga *Chlorella vulgaris* in view of biodiesel production. *Biochem. Eng. J.* **2021**, *173*, 108072, https://doi.org/10.1016/j.bej.2021.108072.
18. Kushwaha, O.S.; Uthayakumar, H.; Kumaresan, K. Modeling of carbon dioxide fixation by microalgae using hybrid artificial intelligence (AI) and fuzzy logic (FL) methods and optimization by genetic algorithm (GA). *Environ. Sci. Pollut. Res.* **2023**, *30*, 24927-24948, https://doi.org/10.1007/s11356-022-19683-0.
19. Nayak, M.; Dhanarajan, G.; Dineshkumar, R.; Sen, R. 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.* **2018**, *201*, 1092-1100, https://doi.org/10.1016/j.jclepro.2018.08.048.
20. Hossain, S.M.Z.; Sultana, N.; Razzak, S.A.; Hossain, M.M. Modeling and multi-objective optimization of microalgae biomass production and CO_2 biofixation using hybrid intelligence approaches. *Renew. Sustain. Energy Rev.* **2022**, *157*, 112016, https://doi.org/10.1016/j.rser.2021.112016.
21. Liu, X.; Chen, G.; Tao, Y.; Wang, J. Application of effluent from WWTP in cultivation of four microalgae for nutrients removal and lipid production under the supply of CO_2. *Renew. Energy* **2020**, *149*, 708–715, https://doi.org/10.1016/j.renene.2019.12.092.
22. Alibrahim, H.; Ludwig, S.A. Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization. In Proceedings of the 2021 IEEE Congress on Evolutionary Computation (CEC), Kraków, Poland, 28 June-1 July 2021, IEEE: **2021**; pp. 1551-1559, https://doi.org/10.1109/CEC45853.2021.9504761.
23. Ahamad Sanadi, N.F.B.; Ibrahim, N.; Ong, P.Y.; Klemeš, J.J.; Li, C.; Lee, C.T. Dilution rate of compost leachate from different biowaste for the fertigation of vegetables. *J. Environ. Manag.* **2021**, *295*, 113010, https://doi.org/10.1016/j.jenvman.2021.113010.
24. Han, J.; Pei, J; Tong, H. 2 – Data, Measurements, And Data Preprocessing. In Data Mining: Concepts and Techniques, 4th Ed.; Morgan Kaufmann: Boston, **2022**; pp. 64 - 65, https://doi.org/10.1016/C2013-0-18660-6
25. Bishop, C. M.; Bishop, H. 6 – Deep Neural Networks. In Deep Learning: Foundations and Concepts, 1st ed.; Springer: Cham, Switzerland, **2024**; pp. 181 – 182, https://doi.org/10.1007/978-3-031-45468-4.
26. Daulton, S.; Cakmak, S.; Balandat, M.; Osborne, M.A.; Zhou, E.; Bakshy, E. Robust multi-objective bayesian optimization under input noise. In Proceedings of the International Conference on Machine

Learning, Baltimore, MD, USA, 17–23 July **2022**; pp. 4831–4866, https://doi.org/10.48550/arXiv.2202.07549.
27. Kouziokas, G. 2. Evolutionary Computation and Genetic Algorithm. In Swarm Intelligence and Evolutionary Computation: Theory, Advances and Applications in Machine Learning and Deep Learning, 1st ed.; CRC Press: Boca Raton, USA, **2023**; pp. 19 - 33, https://doi.org/10.1201/9781003247746.
28. Fernández Izquierdo, P.; Patiño Coral, M.; Ortiz Benavides, F. Application of an artificial neural network coupled to a genetic algorithm for the production of polyphenols in Parachlorella kessleri grown under mixotrophic conditions. *Algal Res.* **2024**, *77*, 103331, https://doi.org/10.1016/j.algal.2023.103331.
29. Mirjalili, S. Genetic Algorithm. In Evolutionary Algorithms and Neural Networks: Theory and Applications; Mirjalili, S., Ed.; Springer International Publishing: Cham, **2019**; Volume 780, pp. 43-55, https://doi.org/10.1007/978-3-319-93025-1.
30. Hossain, S.M.Z.; Sultana, N.; Jassim, M.S.; Coskuner, G.; Hazin, L.M.; Razzak, S.A.; Hossain, M.M. Soft-computing modeling and multiresponse optimization for nutrient removal process from municipal wastewater using microalgae. *J. Water Process Eng.* **2022**, *45*, 102490, https://doi.org/10.1016/j.jwpe.2021.102490.
31. Fagerland, M.W. t-tests, non-parametric tests, and large studies—a paradox of statistical practice? *BMC Med. Res. Methodol.* **2012**, *12*, 78, https://doi.org/10.1186/1471-2288-12-78.
32. Naď, M.; Brummer, V.; Lošák, P.; Máša, V.; Sukačová, K.; Tatarová, D.; Pernica, M.; Procházková, M. Waste-to-energy plants flue gas CO_2 mitigation using a novel tubular photobioreactor while producing *Chlorella* algae. *J. Clean. Prod.* **2023**, *385*, 135721, https://doi.org/10.1016/j.jclepro.2022.135721.
33. Kumar, K.; Dasgupta, C.N.; Das, D. Cell growth kinetics of *Chlorella sorokiniana* and nutritional values of its biomass. *Bioresour. Technol.* **2014**, *167*, 358-366, https://doi.org/10.1016/j.biortech.2014.05.118.
34. Kaštánek, F.; Šabata, S.; Šolcová, O.; Maléterová, Y.; Kaštánek, P.; Bránykiková, I.; Kuthan, K.; Zachleder, V. In-field experimental verification of cultivation of microalgae Chlorella sp. using the flue gas from a cogeneration unit as a source of carbon dioxide. *Waste Manag. Res.* **2010**, *28*, 961-966, https://doi.org/10.1177/0734242X10375866.
35. Tang, H.; Chen, M.; Simon Ng, K.Y; Salley, S.O. Continuous microalgae cultivation in a photobioreactor. *Biotechnol. Bioeng.* **2012**, *109*, 2468–2474, https://doi.org/10.1002/bit.24516.
36. Azari, A.; Tavakoli, H.; Barkdoll, B.D.; Haddad, O.B. Predictive model of algal biofuel production based on experimental data. *Algal Res.* **2020**, *47*, 101843, https://doi.org/10.1016/j.algal.2020.101843.
37. Yadav, G.; Panda, S.P.; Sen, R. Strategies for the effective solid, liquid and gaseous waste valorization by microalgae: A circular bioeconomy perspective. *J. Environ. Chem. Eng.* **2020**, *8*, 104518, https://doi.org/10.1016/j.jece.2020.104518.
38. Jin, X.; Gong, S.; Chen, Z.; Xia, J.; Xiang, W. Potential microalgal strains for converting flue gas CO_2 into biomass. *J. Appl. Phycol.* **2021**, *33*, 47–55, https://doi.org/10.1007/s10811-020-02147-8.
39. Yadav, G.; Meena, D.K.; Sahoo, A.K.; Das, B.K.; Sen, R. Effective valorization of microalgal biomass for the production of nutritional fish-feed supplements. *J. Clean. Prod.* **2020**, *243*, 118697, https://doi.org/10.1016/j.jclepro.2019.118697.
40. Chou, H.-H.; Su, H.-Y.; Song, X.-D.; Chow, T.-J.; Chen, C.-Y.; Chang, J.-S.; Lee, T.-M. Isolation and characterization of *Chlorella* sp. mutants with enhanced thermo- and CO_2 tolerances for CO2 sequestration and utilization of flue gases. *Biotechnol. Biofuels* **2019**, *12*, 251, https://doi.org/10.1186/s13068-019-1590-9.
41. Kassim, M.A.; Meng, T.K. Carbon dioxide (CO_2) biofixation by microalgae and its potential for biorefinery and biofuel production. *Sci. Total Environ.* **2017**, *584–585*, 1121–1129, https://doi.org/10.1016/j.scitotenv.2017.01.172.
42. Mondal, M.; Ghosh, A.; Gayen, K.; Halder, G.; Tiwari, O.N. Carbon dioxide bio-fixation by *Chlorella* sp. BTA 9031 towards biomass and lipid production: Optimization using Central Composite Design approach. *J. CO2 Util.* **2017**, *22*, 317-329, https://doi.org/10.1016/j.jcou.2017.10.008.
43. Van T. Do, C.; Dinh, C.T.; Dang, M.T.; Dang Tran, T.; Giang Le, T. A novel flat-panel photobioreactor for simultaneous production of lutein and carbon sequestration by *Chlorella sorokiniana* TH01. *Bioresour. Technol.* **2022**, *345*, 126552, https://doi.org/10.1016/j.biortech.2021.126552.
44. Yang, Q.P.; Wang, X.L.; Shi, X.Y.; Li, K.Q.; Yue, L.H. Study on Biological Fixation of High-Concentration CO_2 Using Chlorella Pyrenoidosa. *Adv. Mater. Res.* **2012**, *343-344*, 361-367, https://doi.org/10.4028/www.scientific.net/AMR.343-344.361.

45. Lam, M.K.; Lee, K.T. Effect of carbon source towards the growth of *Chlorella vulgaris* for CO_2 bio-mitigation and biodiesel production. *Int. J. Greenh. Gas Control* **2013**, *14*, 169–176, https://doi.org/10.1016/j.ijggc.2013.01.016.
46. Tang, D.; Han, W.; Li, P.; Miao, X.; Zhong, J. CO_2 biofixation and fatty acid composition of *Scenedesmus obliquus* and *Chlorella pyrenoidosa* in response to different CO_2 levels. *Bioresour. Technol.* **2011**, *102*, 3071-3076, https://doi.org/10.1016/j.biortech.2010.10.047.
47. Xia, J.-l.; Gong, S.-q.; Jin, X.-j.; Wan, M.-x.; Nie, Z.-y. Effects of simulated flue gases on growth and lipid production of *Chlorella sorokiniana* CS-01. *J. Cent. South Univ.* **2013**, *20*, 730-736, https://doi.org/10.1007/s11771-013-1541-8.
48. Zhang, X.; Liu, Y.; Cui, H.; Zhang, H.; Xue, J.; Ji, C.; Lee, Y.-s. Research on the enhancement of *Chlorella* growth, metabolism and carbon fixation effects by the chemical absorbent monoethanolamine (MEA). *Acta Laser Biol. Sin.* **2022**, *31*, 311–320, https://doi.org/10.3969/j.issn.1007-7146.2022.04.004.
49. Adamczyk, M.; Lasek, J.; Skawińska, A. CO_2 Biofixation and Growth Kinetics of *Chlorella vulgaris* and *Nannochloropsis gaditana*. *Appl. Biochem. Biotechnol.* **2016**, *179*, 1248-1261, https://doi.org/10.1007/s12010-016-2062-3.
50. Pourjamshidian, R.; Abolghasemi, H.; Esmaili, M.; Amrei, H.D.; Parsa, M.; Rezaei, S. Carbon dioxide biofixation by *Chlorella* sp. In a bubble column reactor at different flow rates and CO_2 concentrations. *Braz. J. Chem. Eng.* **2019**, *36*, 639-645, https://doi.org/10.1590/0104-6632.20190362s20180151.
51. Cho, J.M.; Oh, Y.-K.; Park, W.-K.; Chang, Y.K. Effects of nitrogen supplementation status on CO_2 biofixation and biofuel production of the promising microalga *Chlorella* sp. ABC-001. *J. Microbiol. Biotechnol.* **2020**, *30*, 1235, https://doi.org/10.4014/jmb.2005.05039.
52. Politaeva, N.; Ilin, I.; Velmozhina, K.; Shinkevich, P. Carbon Dioxide Utilization Using *Chlorella* Microalgae. *Environments* **2023**, *10*, 109, https://doi.org/10.3390/environments10070109.
53. Zhao, B.; Zhang, Y.; Xiong, K.; Zhang, Z.; Hao, X.; Liu, T. Effect of cultivation mode on microalgal growth and CO_2 fixation. *Chem. Eng. Res. Des.* **2011**, *89*, 1758-1762, https://doi.org/10.1016/j.cherd.2011.02.018.
54. Chiu, S.-Y.; Tsai, M.-T.; Kao, C.-Y.; Ong, S.-C.; Lin, C.-S. The air-lift photobioreactors with flow patterning for high-density cultures of microalgae and carbon dioxide removal. *Eng. Life Sci.* **2009**, *9*, 254-260, https://doi.org/10.1002/elsc.200800113.
55. Ryu, H.J.; Oh, K.K.; Kim, Y.S. Optimization of the influential factors for the improvement of CO_2 utilization efficiency and CO_2 mass transfer rate. *J. Ind. Eng. Chem.* **2009**, *15*, 471-475, https://doi.org/10.1016/j.jiec.2008.12.012.
56. Ayatollahi, S.Z.; Esmaeilzadeh, F.; Mowla, D. Integrated CO_2 capture, nutrients removal and biodiesel production using *Chlorella vulgaris*. *J. Environ. Chem. Eng.* **2021**, *9*, 104763, https://doi.org/10.1016/j.jece.2020.104763.
57. Borkenstein, C.G.; Knoblechner, J.; Frühwirth, H.; Schagerl, M. Cultivation of *Chlorella emersonii* with flue gas derived from a cement plant. *J. Appl. Phycol.* **2011**, *23*, 131-135, https://doi.org/10.1007/s10811-010-9551-5.
58. Hanifzadeh, M.M.; Sarrafzadeh, M.H.; Tavakoli, O. Carbon dioxide biofixation and biomass production from flue gas of power plant using microalgae. In Proceedings of the 2012 Second Iranian Conference on Renewable Energy and Distributed Generation, Tehran, Iran, 6-8 March 2012, IEEE: **2012**; pp. 61-64, https://doi.org/10.1109/ICREDG.2012.6190469.
59. Kim, S.; Park, J.-e.; Cho, Y.-B.; Hwang, S.-J. Growth rate, organic carbon and nutrient removal rates of Chlorella sorokiniana in autotrophic, heterotrophic and mixotrophic conditions. *Bioresour. Technol.* **2013**, *144*, 8–13, https://doi.org/10.1016/j.biortech.2013.06.068.
60. Martínez, L.; Otero, M.; Morán, A.; García, A.I. Selection of native freshwater microalgae and cyanobacteria for CO_2 biofixation. *Environ. Technol.* **2013**, *34*, 3137–3143, https://doi.org/10.1080/09593330.2013.808238.
61. Kokol, P.; Kokol, M.; Zagoranski, S. Machine learning on small size samples: A synthetic knowledge synthesis. *Sci. Prog.* **2022**, *105*, 00368504211029777, https://doi.org/10.1177/00368504211029777.
62. Jakšić, Z.; Devi, S.; Jakšić, O.; Guha, K. A Comprehensive Review of Bio-Inspired Optimization Algorithms Including Applications in Microelectronics and Nanophotonics. *Biomimetics* **2023**, *8*, 278, https://doi.org/10.3390/biomimetics8030278.
63. Liang, Q.; Gongora, A.E.; Ren, Z.; Tiihonen, A.; Liu, Z.; Sun, S.; Deneault, J.R.; Bash, D.; Mekki-Berrada, F.; Khan, S.A.; Hippalgaonkar, K.; Maruyama, B.; Brown, K.A.; Fisher III, J.; Buonassisi, T.

Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains. *npj Comput. Mater.* **2021**, *7*, 188, https://doi.org/10.1038/s41524-021-00656-9.
64. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623, https://doi.org/10.7717/peerj-cs.623.
65. Miller, C.; Portlock, T.; Nyaga, D.M.; O'Sullivan, J.M. A review of model evaluation metrics for machine learning in genetics and genomics. *Front. Bioinform.* **2024**, *4*, 1457619, https://doi.org/10.3389/fbinf.2024.1457619.
66. Papenmeier, L.; Nardi, L.; Poloczek, M. Bounce: reliable high-dimensional Bayesian optimization for combinatorial and mixed spaces. In Proceedings of the Proceedings of the 37th International Conference on Neural Information Processing Systems, New Orleans, LA, USA,10 – 16 December 2023, Curran Associates Inc: NY, USA, **2023**; 86, 1764 – 1793.
67. Gonçalves, A.L.; Almeida, F.; Rocha, F.A.; Ferreira, A. Improving CO_2 mass transfer in microalgal cultures using an oscillatory flow reactor with smooth periodic constrictions. *J. Environ. Chem. Eng.* **2021**, *9*, 106505, https://doi.org/10.1016/j.jece.2021.106505.
68. 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. Microalgae-based wastewater treatment for nutrients recovery: A review. *Bioresour. Technol.* **2019**, *291*, 121934, https://doi.org/10.1016/j.biortech.2019.121934.
69. Brück, W.M.; Alfonso, E.; Rienth, M.; Andlauer, W. Heat Stress Resistance in *Chlorella vulgaris* Enhanced by Hydrolyzed Whey Proteins. *Agronomy* **2024**, *14*, 2854, https://doi.org/10.3390/agronomy14122854.
70. Mendes, A.R.; Spínola, M.P.; Lordelo, M.; Prates, J.A.M. Advances in Bioprocess Engineering for Optimising *Chlorella vulgaris* Fermentation: Biotechnological Innovations and Applications. *Foods* **2024**, *13*, 4154, https://doi.org/10.3390/foods13244154.

Publisher's Note & Disclaimer