Renewable and Sustainable Energy Reviews 157 (2022) 112016

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Modeling and multi-objective optimization of microalgae biomass production and CO_2 biofixation using hybrid intelligence approaches

S. M. Zakir Hossain [a], Nahid Sultana [b], Shaikh A. Razzak [c,d], Mohammad M. Hossain [c,e,*]

[a] Department of Chemical Engineering, University of Bahrain, Zallaq, Bahrain
[b] Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[c] Department of Chemical Engineering, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia
[d] Center for Membranes & Water Security, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia
[e] Center for Refining & Advanced Chemicals, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

ARTICLE INFO

Keywords:
Microalgae
CO_2 biofixation
Modeling
Optimization
Artificial intelligence

ABSTRACT

This study investigates the impacts of temperature, light-dark cycles (LD), and nitrogen-phosphorus ratios (NP) on *Chlorella vulgaris* microalgae biomass productivity (BP) and CO_2 biofixation (R_{CO2}). Three artificial intelligence (AI) modeling approaches - boosted regression tree (BRT), artificial neural networks (ANN), and support vector regression (SVR) were applied. Bayesian optimization algorithm (BOA) was combined with each AI approach to predict BP and R_{CO2}. Real-life experimental data, according to Box-Behnken design (BBD) were employed to assess the models using relative error (RE), coefficient of determination (R^2), mean absolute error (MAE), mean absolute relative error (MARE), root mean square error (RMSE), and fractional bias (FB). The performance of the ANN and SVR models are comparable. However, the SVR model performs much better than the BRT and ANN models. Regarding R_{CO2}, the SVR model yields low errors (MAE of 0.0128, MARE of 0.4131, RMSE of 0.0189) with a high R^2 of 0.911. The value of FB is close to zero (0.0088), suggesting the model is reliable. The SVR model shows a better prediction capability of R_{CO2} compared to BBD with a performance improvement of 17.16%. MARE of BBD for R_{CO2} is 0.7409, which is higher than that of SVR model. Finally, the crow search algorithm was combined with the SVR for multi-objective optimization to determine the global optimal conditions for maximizing BP and R_{CO2}, respectively. The optimum conditions were calculated to be 40 °C, 1:1 of N/P, and 12/12 h/h of LD with BP and R_{CO2} of 0.0979 g $L^{-1}d^{-1}$ and 0.1408 g $L^{-1}d^{-1}$, respectively.

1. Introduction

Rapid urbanization and population growth have resulted in the energy demand also increasing in parallel. Meeting the energy demand is increasingly posing a serious challenge owing to the finite and dwindling fossil energy sources. In addition, these energy sources lead to global warming due to the release of toxic gases (e.g., CO_2, NO_x, SO_x, etc.) to the environment upon combustion [1,2]. International Energy Agency (IEA) data reveal that global emission of CO_2 at the end of 2019 was 33 GT [3]. Hence, the global scientific community is keen on developing CO_2 capture and utilization (CCU) techniques [4,5]. Among the typical techniques available for CCU are injection into geological formations or deep oceans, solvent-based (viz., amines, amine blends, sodium carbonate, ionic liquids, amino acids, etc.) chemisorption, carbonate looping, so-called oxyfuel combustion as well as the use of various absorbent/adsorbent materials [4,6,7]. Even though most of these methods are promising and have acceptable efficiencies, the high cost (capital and running) [8] and environmental pollution (due to hydrate generation) are some of their shortcomings [9].

Biological techniques are among the popular environmentally sustainable approaches for decreasing CO_2 in the atmosphere. Microalgae, photosynthetic microorganisms, are considered a promising alternative for in-situ CO_2 biofixation. The technique also produces a large amount of biomass that can be used as the feedstock or adsorbent to produce valuable products (e.g., biofuels, organic acids, bio-fertilizers) and wastewater treatment [10–13]. Chen et al. (2012) have reported that CO_2 can be used sustainably via a microalgae photobioreactor system [14]. Microalgae use atmospheric CO_2 during their growth via photosynthesis and sequester about 80% of CO_2 from the culture media [15,16]. Microalgae can be cultivated in both open and closed systems,

* Corresponding author. Department of Chemical Engineering, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia.
E-mail address: mhossain@kfupm.edu.sa (M.M. Hossain).

https://doi.org/10.1016/j.rser.2021.112016
Received 21 May 2021; Received in revised form 20 October 2021; Accepted 19 December 2021
Available online 7 January 2022

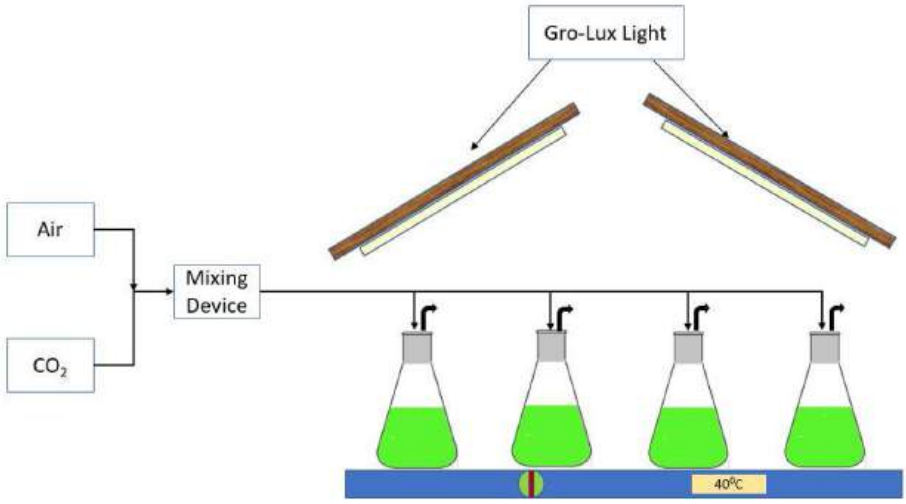

Fig. 1. Schematic illustration of the experimental setup applied for the cultivation of microalgae (*Chlorella vulgaris* sp.).

including seas, raceway ponds, circular ponds, tubular photobioreactors, flat panel photobioreactors, plastic bags photobioreactors, and membrane photobioreactors [10]. Thus, the amount of microalgae biomass must be maximized to optimize CO_2 fixation, thereby achieving optimum process variables. Process variables of microalgae cultures that should be optimized include temperature (T), light intensity, light-dark cycle (LD), pH of the media, nutrient or substrate concentration, and nitrogen to phosphorus ratio (NP) [17–20]. As CO_2 biofixation by microalgae depends heavily on its culture conditions, optimizing all of these parameters is crucial [18,21]. Some of the above-mentioned parameters were examined and optimized in earlier investigations via batch photobioreactors by this group [22,23]. Studies on the effect of each independent variable on maximizing microalgae growth as well as CO_2 biofixation involve a substantial number of laboratory trials, and thus, require a significant amount of time, trained hands, and other resources.

Predictive estimation is a sophisticated and effective approach, which allows overcoming these hurdles in investigating complex systems with multi-response variables [24,25]. Artificial intelligence (AI) techniques, viz., boosted regression tree (BRT), artificial neural network (ANN), support vector regression (SVR), etc., are popular and effective modeling tools for predicting process performance due to their high competence, accuracy, and applications in several areas of science and engineering [26–28]. BRT uses ensemble designs that minimize the generalization error, responding with good precision [29–31]. The ensemble method is considered one of the state-of-the-art solutions to overcome many artificial intelligence challenges [32]. This approach increases the anticipative performance of a single model by training several models and combining their predictions. ANN is a type of computational model produced using hundreds of artificial neurons with associated coefficients/weights generating neural structures that simulate biological processes [33,34]. As such, ANN functions simply as a human brain. SVR is also a powerful artificial intelligence (AI) tool that predicts useful features of a system, including a suitable hyperplane for implementation, performance level, empirical risk, stability, and non-convergence to a local minimum [35,36]. As the performance of each AI technique depends considerably on its hyperparameters, optimizing all hyperparameters is crucial for developing a tuned model. In this respect, a Bayesian optimization algorithm (BOA) can be applied to tune the hyperparameters automatically associated with each AI technique, resulting in hybrid optimal intelligence approaches. Several reports on investigations of the performance of artificial intelligence approaches in predicting CO_2 solubility in ionic liquids are available [37,38]. ANN and decision tree have also been adapted for predicting efficient CO_2 utilization [39,40]. In a few studies, the SVR prediction of microalgae growth has been evaluated [41,42]. To the best of our knowledge, comparative modeling using hybrid intelligence approaches to predict microalgae growth and CO_2 biofixation have not been published in the literature.

In this study, the impacts of three key variables, viz., temperature, LD, and NP on CO_2 biofixation (R_{CO2}) and biomass productivities (BP) of *Chlorella vulgaris* were evaluated. The success of enhancing biomass productivities and CO_2 fixation lies in the understanding of the characteristics of these parameters and their relationship to their impact on cultivation. Detailed experimental studies were undertaken and the data were used to understand the behavior and influence of the process parameters on the culture, biomass growth, and CO_2 fixation. Modeling and process optimization are required to understand the impact and behavior of these three parameters on growth. In this regard, three hybrid intelligence approaches, viz., BOA-BRT, BOA-ANN, and BOA-SVR were used to predict the microalgae BP and R_{CO2}. Based on the aforementioned literature survey and discussion, the following key tasks were performed:

(i) *K*-fold cross-validation and a BOA were applied in all cases (i.e., BRT, ANN, and SVR) to automatically optimize the hyperparameters.
(ii) The developed hybrid models were then analyzed and compared to each other using several performance grading indicators (e.g., RE, R^2, MAE, MARE, RMSE, FB) to determine the best predictive model. The performance of the best hybrid intelligence model was also compared with the statistical Box-Behnken design (BBD) model.
(iii) Optimization of the process variables is vital for maximizing BP and R_{CO2}. Hence, the developed best hybrid prediction model was coupled with the crow search algorithm (CSA) for optimization. CSA is a fairly new nature-inspired metaheuristic algorithm for global optimization of independent parameters in different research fields [43,44].

A multi-objective optimization method was applied to maximize BP and R_{CO2}. This study, for the first time, describes modeling and optimization of a microalgae-based CO_2 biofixation process particularly designed for simultaneously maximizing biomass growth and CO_2 utilization. The optimal input conditions were confirmed by comparing them with laboratory data.

This paper is organized as follows: Section 2 describes the experimental procedure and data acquisition aspects. The details of the

Table 1
The range and level of the input variables.

Independent Factors	Symbol	Range and levels		
		−1	0	1
Temperature	x_1	20	30	40
Nitrogen-phosphorus ratio (N/P)	x_2	1:1	3:1	6:1
Light-dark cycles (LD)	x_3	6/18	12/12	24/0

procedure for obtaining real-life experimental data and the model development process are also outlined in this section. Section 3 provides details of the results and a discussion. Finally, the concluding remarks are provided in Section 4.

2. Experimental procedure and data acquisition

2.1. Microalgae species and growth conditions

Chlorella vulgaris (UTEX 2714) was obtained from the University of Texas, USA. The Bold's Basal Medium (BBM) was used as the microalgae growth medium. The nitrogen and phosphorus ratio (NP) of the BBM medium was modified using $NaNO_3$ and $KHPO_4$. The composition of BBM with respect to other chemicals, including trace metals and vitamins, was not changed. Microalgae *Chlorella vulgaris* sp. were cultivated in PYREX 1 L Erlenmeyer flasks (batch photobioreactors) as shown in Fig. 1 with a working volume of 500 mL. The initial inoculum dose was nearly 2.2×10^7 cells mL^{-1} for all laboratory trials. Photobioreactors were positioned in a water bath on a bench, with four Gro-lux fluorescent light tubes (average intensity of 65 μmol $m^{-2}s^{-1}$) illuminating them. The *Chlorella vulgaris* was cultured with light-dark cycles (LD) of 12/12 and temperature was adjusted by changing the water temperature in the bath. A 4% mixture of CO_2 in the air was supplied to the Photobioreactors.

2.2. Estimation of microalgae growth parameters and CO_2 biofixation

The kinetic factors of microalgae growth, including specific growth rate (μ) (d^{-1}) and biomass productivity (BP) (g L^{-1} d^{-1}), were determined using Eqs. (1) and (2), respectively.

$$\text{Specific growth rate}, \mu = \frac{\ln\left(X_2/X_1\right)}{t_2 - t_1} \tag{1}$$

$$\text{Biomass productivity}, \ \mathbf{BP} = \frac{\mathbf{X_t} - \mathbf{X_o}}{\mathbf{t_x} - \mathbf{t_o}} \tag{2}$$

where X_1 and X_2 are the amount of biomass (g/L) at the beginning t_1 and end t_2 of the exponential growth phase, respectively, X_t is the biomass concentration (g/L) at the end of the cultivation t_x, and X_o is the initial biomass concentration (g/L) at t_0 (days).

CO_2 biofixation (R_{CO2}) (g L^{-1} d^{-1}) was determined via Eq. (3).

$$\text{CO}_2 \text{ biofixation rate}, \mathbf{R_{CO_2}} = \mathbf{C_c} \times \mathbf{BP} \times \left(\frac{\mathbf{M_{CO_2}}}{\mathbf{M_c}}\right) \tag{3}$$

where C_c is the carbon fraction of the biomass, BP is the biomass productivity, and $\frac{M_{CO2}}{M_C}$ denotes the ratio of molecular weight (MW) of CO_2 to carbon. The C_c was estimated using a TOC analyzer [45].

2.3. Techniques used in modeling and optimization of the parameters

Three AI approaches, viz., BRT, ANN, and SVR were used to develop predictive models. BBD, one of the response surface methodologies (RSMs), was applied for the design of the experiments. BOA was integrated with each AI approach to construct novel hybrid models to predict BP and R_{CO2}.

2.3.1. Design of experiments using BBD

Response surface methodology is a well-known hybrid statistical and mathematical tool that is used to design experiments (DoE), modeling, and optimization [22,46]. Two key designs of RSM, viz., BBD, and central composite design (CCD) are available for the design of experiments. In this study, BBD was selected instead of CCD as, generally, BBD requires fewer experimental trials than CCD, even though both methods provide similar results [47]. The general operational association between independent and dependent parameters in RSM is shown in Eq. (4).

$$y = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \sum_{i=1}^{N} \beta_{ii} x_i^2 + \sum_{i<j} \sum \beta_{ij} x_i x_j + \varepsilon \tag{4}$$

where y, x_i, β_o, β_i, β_{ii}, β_{ij}, and ϵ denote response, coded independent variable, intercept, linear effect, squared effect, interaction effect, and error, respectively.

BBD provides three coded levels for each parameter (T, LD, or NP), namely, high (+1), center (0), and low (−1). BP and R_{CO2} are the dependent variables or the responses. A total of 15 treatments with different operating parameters were used in a random order to obtain results free of bias. The range and coded level of each independent variable is given in Table 1. The association between coded and natural parameters is expressed by Eq. (5).

$$\text{Coded value} = \frac{(actual\ value - mean)}{half\ of\ range} \tag{5}$$

The experimental matrix is shown in Table 2. All experiments were conducted based on this matrix and the experimental data were used to develop the models using artificial intelligence approaches. All the analyses were conducted using the data obtained from microalgae cultivation on Day 8.

2.3.2. Boosted regression tree (BRT)

A regression tree (RT) is used in data mining with a numerical target variable. In this study, an RT ensemble was used to determine the best predictive model. Generally, ensemble models combine the decisions from numerous models to enhance the overall performance. The RT ensemble is composed of a weighted combination of multiple regression trees, which is categorized into two main methods, namely bagging and boosting. Bagging involves a simple average of results to attain an overall prediction, while boosting uses an iterative procedure and incorporates a weighted average of results, thereby achieving a better prediction. As such, boosted regression tree (BRT) was chosen as one of the effective AI tools in this study.

The inputs used in the tree are T, LD, and N/P. The hyperparameters in BRT modeling include a method, number of the ensemble learning cycle, leaf size, and learning rate. Low or high values of these factors may lead to a high generalization error, overfitting, or underfitting. Thus, these parameters must be tuned to guarantee a good model performance when predicting unseen data. Several common methods to optimize the hyperparameters are available in the open literature, including the random search algorithm (RSA), grid search algorithm (GSA), particle swarm optimization (PSO), and BOA. GSA requires numerous trials and thus it is a time-consuming process. PSO, which is a well-known approach, also takes a longer time. In contrast, BOA provides good results much faster than the other methods as it applies an acquisition function that determines the next point to evaluate. BOA was used in the current study, as it is also an orderly process for tuning that does not need derivatives [48,49]. K-fold cross-validation was also used along with BOA to prevent overfitting. A brief description of k-fold cross-validation and BOA are given below.

K-fold cross-validation: In this technique, the data are equally distributed into *k* subclasses. One subclass is chosen as a test, while the others are applied as training subgroups. This method is repeated *k* times and hence, each subclass is applied just one time for testing. Even though

Table 2
Experimental and anticipated outputs of specific growth rate (μ), biomass productivity (BP), and CO_2 biofixation (R_{CO2}).

Experiment number	Coded values			Natural values			Responses		
							Specific growth rate, μ	Biomass productivity, BP	CO_2 biofixation rate, R_{CO2}
	x_1	x_2	x_3	T (°C)	N/P	LD	(d^{-1})	($g\ L^{-1}d^{-1}$)	($g\ L^{-1}d^{-1}$)
1	0	1	−1	30	6:1	6/18	0.669	0.219	0.157
2	−1	0	−1	20	3:1	6/18	0.246	0.054	0.02
3	0	0	0	30	3:1	12/12	0.017	0.001	0.02
4	−1	−1	0	20	1:1	12/12	0.686	0.041	0.048
5	−1	1	0	20	6:1	12/12	0.662	0.175	0.134
6	0	0	0	30	3:1	12/12	0.022	0.001	0.02
7	1	0	1	40	3:1	24/0	0.237	0.041	0.02
8	0	−1	1	30	1:1	24/0	0.178	0.038	0.026
9	−1	0	1	20	3:1	24/0	0.284	0.058	0.039
10	1	1	0	40	6:1	12/12	0.49	0.021	0.018
11	0	−1	−1	30	1:1	6/18	0.784	0.097	0.182
12	1	−1	0	40	1:1	12/12	0.753	0.104	0.149
13	0	0	0	30	3:1	12/12	0.472	0.012	0.004
14	0	1	1	30	6:1	24/0	0.591	0.048	0.119
15	1	0	−1	40	3:1	6/18	0.017	0.001	0.02

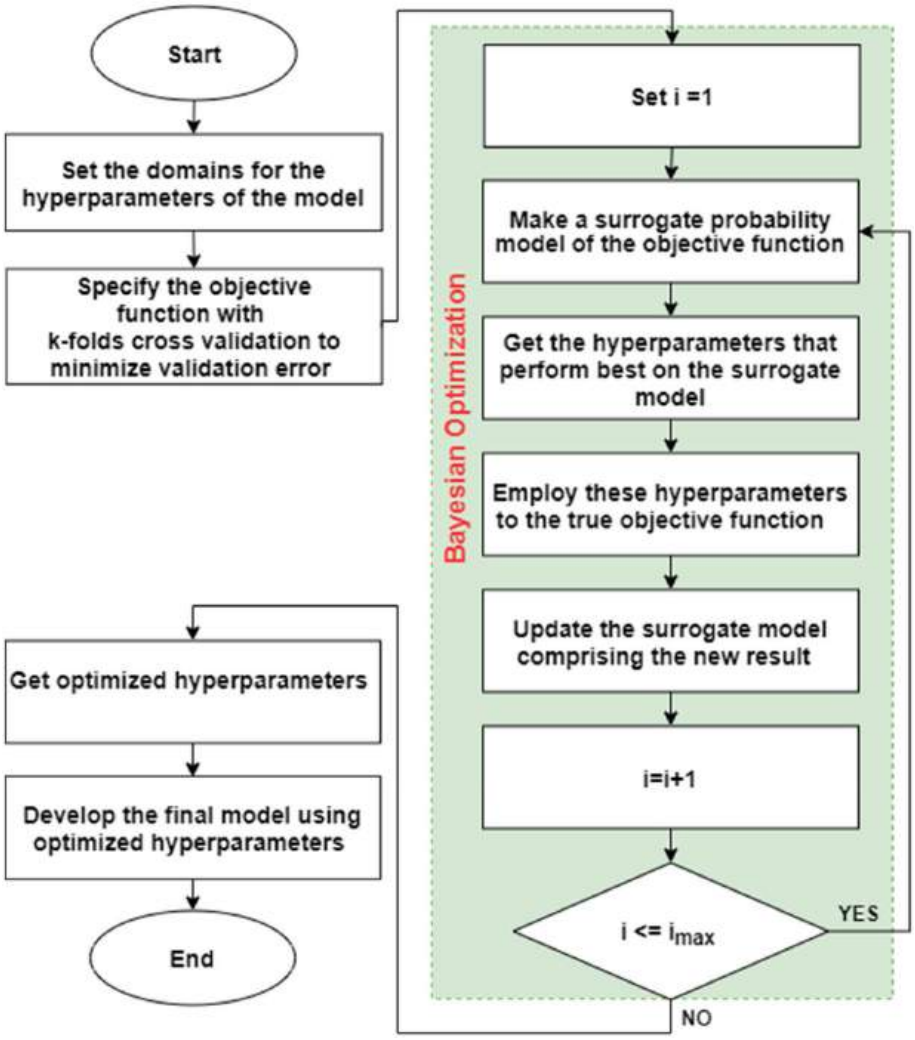

Fig. 2. A computational flowchart of the hybrid k-fold cross-validation and BOA approach.

the performance of the model is enhanced in terms of RMSE with the increase of the k value (2, 3, 4, 5), it also increases the computational time. Hence, the operator should apply a trade-off.

BOA: BOA involves a surrogate probability model via the Bayes' rule, as described by Eq. (6) [50].

$$p(w|D) = \frac{p(D|w)\ p(w)}{p(D)} \tag{6}$$

where $p(w)$ is the earlier distribution, $p(D)$ indicates evidence, $p\ (D|w)$

Input: T = Maximum number of iterations, N = crow's flock, pd = Problem dimension, AP = Awareness probability, fl = Flight length

Initialize:

- Randomly initialize the position of a flock of N crows within the search space
- Evaluate the position of the crows
- Initialize the memory of each crow

$t = 0$

while $t \leq T$

  for $i = 1$ to N do

    extract randomly one crow from the flock to follow it (for example j)

    Generate $r_i \in [0,1]$

    if $r_i \geq AP$ then

      Generate the new position for crow i as follows

      $x^{i,t+1} = x^{i,t} + r_i \times fl^{i,t} \times \left(m^{j,t} - x^{i,t}\right)$,

      where $x^{i,t}$ is the position of crow i at iteration t in the search space, and $m^{j,t}$ represents the memory (position of hiding place) of crow j at iteration t

    Else

      Generate the new position for crow i as follows

      $x^{i,t+1}$ = a random position of search space

    End if

  End for

  Examine the feasibility of new position

  Evaluate the new position of the crows

  Update the memory of each crow as follows

$$m^{i,t+1} = \begin{cases} x^{i,t+1} & \text{if } f\left(x^{i,t+1}\right) \text{ is better than } f\left(m^{i,t}\right) \\ m^{i,t} & otherwise \end{cases}$$

  where $f(\cdot)$ denotes the objective function value.

  $t = t + 1$

end while

Output: Best position

Fig. 3. Pseudocode for crow search algorithm (CSA).

represents the probability, and p (w| D) refers to the posterior distribution.

In this technique, the values for the next iteration are determined using the results of previous iterations. Thus, the optimum point can be achieved more effectively as opposed to when random selection is used. Two models, namely substitute and acquisition functions are used in BOA. The substitute model assesses the objective function applying the Gaussian process (GP). The GP process on the function $f(x)$ can be addressed using the mean function (m) and the covariance kernel function (k), as described by Eq. (7).

$$f(x) \sim \mathrm{GP}\left(\mathrm{m}(\mathrm{x}),\ \ \mathrm{k}\left(x_i x_j\right)\right) \tag{7}$$

Readers are referred to a published article for more details [51].

The acquisition function depends on the prior observations, and it is maximized with reiterations. The acquisition function model implies the next point to reiterate using the results of the substitute model. Hyperparameter tuning using BOA can be expressed mathematically by Eq. (8).

$$x^{\star} = \underset{x \varepsilon X}{\operatorname{argmin}} f(x) \tag{8}$$

where $f(x)$ refers the objective score to lower RMSE, $x^{\star}$ denotes a set of hyperparameters that yields the lowest objective score, and x is the arbitrary value of space X. In this investigation, BOA was used because it is more efficient than other typical techniques (e.g., grid search, random search, manual search, etc.). BOA is a systematic method for global optimization [52,53] and a computational flowchart for the BOA-based hybrid model is shown in Fig. 2.

2.3.3. Artificial neural networks (ANN)

ANN is one of the familiar artificial intelligence techniques for predictive modeling. The multilayer perceptron (MLP) neural network model was generated to predict BP and R_{CO2}. MLP uses a feed-forward network where the input signals pass from left to right via several layers (viz., input, hidden, and output). The relationship among input, hidden, and output layers can be defined by the following simple equation (9).

$$Z_j = f \sum_{i=1}^{N} |xw + b|_i \tag{9}$$

where w indicates the weight, b represents the bias, f denotes the operating function, and x and Z represent the ith input and jth output, respectively.

A hyperbolic tangent activation function was used in the hidden layer, while a purelin transfer function was adapted in the output layer. Laboratory data were divided randomly into training (70%), validation (10%), and testing (20%) sample groups. The inputs used in the network were T, LD, and NP. Several training algorithms (e.g., the gradient descent with momentum, gradient descent, scaled conjugate gradient (trainscg), quasi-newton, and levenberg-marquat (trainlm)) were evaluated to train the data. The ANN hyperparameters, including the optimum number of neurons in the hidden layer and learning rate, were optimized using k-fold cross-validation and BOA, as discussed above.

2.3.4. Support vector regression (SVR)

SVR is another form of the artificial intelligence approach that specifies the association between descriptors or inputs and targets. SVR uses kernel functions that form a suitable hyperplane and minimize the generalized error boundary [26]. The popular kernel functions used for SVR model development, are linear, polynomial, and Gaussian. The performance of the SVR model relies heavily on the suitable selection of

Table 3
Experimental and predicted values for biomass productivity (BP) using different artificial intelligence approaches with relative error.

Experiment number	Experimental values	Predicted values			Relative error		
		BRT	ANN	SVR	BRT	ANN	SVR
1	0.219	0.222	0.218	0.214	0.013	0.005	0.025
2	0.054	0.069	0.054	0.059	0.274	0.004	0.100
3	0.001	0.027	0.011	0.006	26.130	10.039	5.415
4	0.041	0.098	0.109	0.064	1.384	1.656	0.572
5	0.175	0.098	0.109	0.064	0.441	0.378	0.632
6	0.001	0.027	0.011	0.006	26.130	10.039	5.415
7	0.041	0.042	0.041	0.046	0.033	0.008	0.132
8	0.038	0.029	−0.024	0.043	0.245	1.626	0.141
9	0.058	0.064	0.058	0.063	0.104	0.003	0.093
10	0.021	0.037	0.021	0.026	0.756	0.001	0.256
11	0.097	0.033	0.095	0.092	0.655	0.021	0.055
12	0.104	0.041	0.102	0.099	0.608	0.023	0.052
13	0.012	0.027	0.011	0.006	1.261	0.080	0.465
14	0.048	0.050	−0.004	0.053	0.044	1.093	0.112
15	0.001	0.047	−0.002	0.006	46.103	2.655	5.415

Table 4
Experimental and predicted values for CO_2 biofixation rate (R_{CO2}) using different artificial intelligence approaches with relative error.

Experiment. number	Experimental values	Predicted values			Relative error		
		BRT	ANN	SVR	BRT	ANN	SVR
1	0.157	0.121	0.157	0.149	0.229	0.001	0.052
2	0.020	0.055	0.065	0.028	1.733	2.249	0.413
3	0.020	0.016	0.016	0.012	0.214	0.207	0.393
4	0.048	0.090	0.117	0.071	0.867	1.439	0.474
5	0.134	0.090	0.117	0.071	0.331	0.126	0.472
6	0.020	0.016	0.016	0.012	0.214	0.207	0.393
7	0.020	−0.002	−0.008	0.028	1.080	1.375	0.410
8	0.026	0.078	0.024	0.034	1.990	0.088	0.318
9	0.039	0.028	0.042	0.047	0.287	0.077	0.212
10	0.018	0.060	0.019	0.026	2.345	0.082	0.452
11	0.182	0.168	0.161	0.174	0.077	0.118	0.045
12	0.149	0.114	0.134	0.141	0.234	0.104	0.055
13	0.004	0.016	0.016	0.012	2.931	2.965	2.034
14	0.119	0.096	0.122	0.111	0.197	0.022	0.068
15	0.020	0.032	0.017	0.028	0.603	0.150	0.405

the hyperparameters, viz., kernel function type, kernel scale (γ), epsilon (ϵ), and box constraints (C). Hence, these hyperparameters must be optimized to confirm the model performance when predicting unseen data. All the hyperparameters were optimized using k-fold cross-validation and BOA, as outlined above. The mathematical explanation of SVR is well established and is available in the literature [54–56].

2.3.5. Process parameters optimization

CSA was coupled with the hybrid BOA-SVR method for multi-objective optimization. The best model was chosen as the fitness function for CSA. All responses are maximized concurrently via trade-offs with a set of optimal parameters. The generation of the CSA computer code (step by step) using MATLAB (version R-2019a) has been reported in previous studies [57,58]. A brief pseudo-code for CSA is shown in Fig. 3.

In the multi-objective optimization technique, the objective functions are normally transformed into a linear or nonlinear function or processed into Pareto fronts [59]. However, an extra computational endeavor is needed in the latter technique. In this study, a linear composite objective function was generated by the combination of two objective functions, as described by Eq. (10).

$$\text{Maximize } Y = p_1{}^*Y1 + p_2{}^*Y2 \tag{10}$$

Subject to: $p_1 + p_2 = 1$ and $0 \le p_i \le 1, \forall i$.

Where Y1, Y2, and Y represent the objective functions for BP, R_{CO2}, and composite BP and R_{CO2}, respectively.

2.4. Estimation of the performance of the models

The performance of the models was assessed employing the following indicators:

$$\text{Relative error, RE} = \frac{Y_{Exp} - Y_P}{Y_{Exp}} \times 100 \tag{11}$$

$$\text{Coefficient of determination, } R^2 = 1 - \frac{\sum_{i=1}^{N}\left(Y_{Exp} - Y_P\right)^2}{\sum_{i=1}^{N}\left(Y_{Exp} - \overline{Y}_{Exp}\right)^2} \tag{12}$$

$$\text{Predicted coefficient of determination, } R^2_{pred} = 1 - \frac{\sum_{i=1}^{n}\frac{\left(Y_{Exp} - Y_P\right)^2}{(1-H_{ii})^2}}{\sum_{i=1}^{n}\left(Y_{Exp} - \overline{Y}_{Exp}\right)^2} \tag{13}$$

$$\text{Mean absolute error, MAE} = \frac{1}{N}\left(\sum_{i=1}^{N}\left|Y_{Exp} - Y_P\right|\right) \tag{14}$$

$$\text{Mean absolute relative error, MARE} = \frac{1}{N}\left(\sum_{i=1}^{N}\left|\frac{Y_{Exp} - Y_P}{Y_{Exp}}\right|\right) \tag{15}$$

$$\text{Root mean square error, RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_{Exp} - Y_P\right)^2} \tag{16}$$

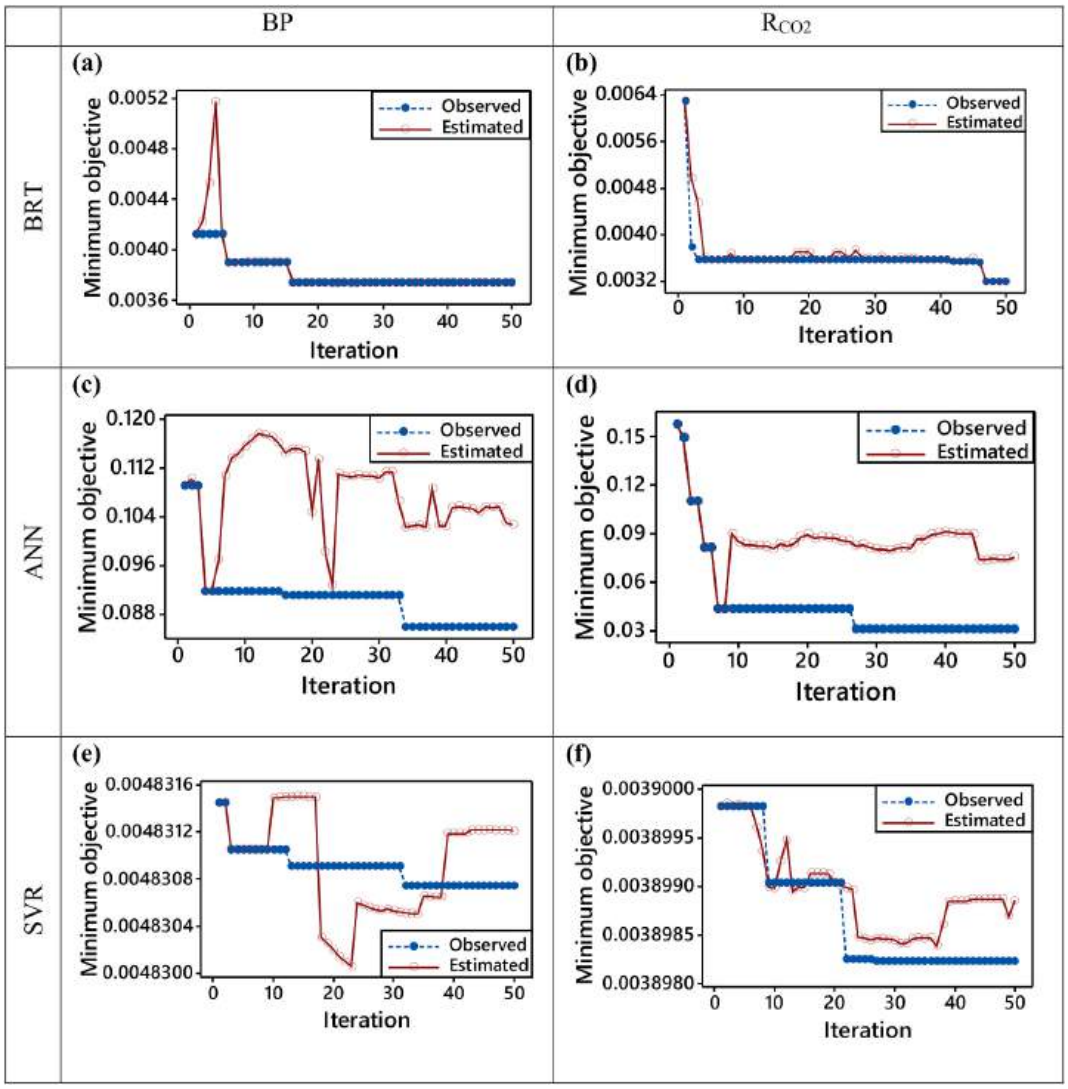

Fig. 4. Minimum objective score versus iterations plots for (a) biomass productivity (BP) and (b) CO_2 biofixation (R_{CO2}) during Bayesian optimization.

Table 5
Tuned hyperparameters of BRT models for biomass productivity (BP) and CO_2 biofixation (R_{CO2}).

Parameters	BP	R_{CO2}
Technique	LSBoost	LSBoost
Number of the ensemble learning cycle	13	12
Learning rate	0.95566	0.52588
Min leaf size	4	1

$$\text{Fractional bias, FB} = \frac{2\sum_{i=1}^{N}\left(Y_{Exp} - Y_P\right)}{\sum_{i=1}^{N}\left(Y_{Exp} + Y_P\right)} \quad (17)$$

where Y_{Exp}, $\overline{Y}_{Exp}$, Y_{P}, and N are the experimental value, the mean laboratory trials value, the predicted value, and the amount of data, respectively. H_{ii} are the i^{th} diagonal entry of the hat matrix $H = X(X^TX)^{-1}X^T$ with $X \in \mathbb{R}^{n\times p}$ be the matrix of data.

2.5. Software usage

Experimental design matrix (or BBD matrix) was generated using Minitab (version 18), while MATLAB (version R-2019a) was applied to develop hybrid BOA-BRT, BOA-ANN, and BOA-SVR models. Crow

Table 6
Tuned hyperparameters of ANN models for biomass productivity (BP) and CO_2 biofixation (R_{CO2}).

Parameters	Optimum value chosen	
	BP	R_{CO2}
Network type	Feed forward back propagation	Feed forward back propagation
Training function	Levenberg-Marquat (trainlm)	Levenberg-Marquat (trainlm)
Hidden layer	1 hidden layer with 10 neurons	1 hidden layer with 7 neurons
Learning rate	0.9730	0.0010
Transfer function	Hyperbolic tangent transfer function for the hidden layer and purelin transfer function for the output layer	Hyperbolic tangent transfer function for the hidden layer and purelin transfer function for the output layer

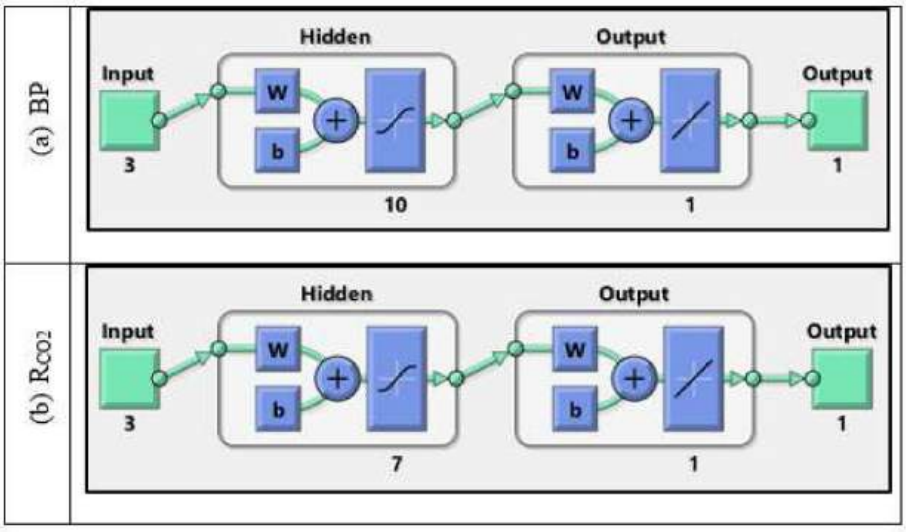

Fig. 5. The optimized ANN structure for (a) biomass productivity (BP) and (b) CO_2 biofixation (R_{CO2}).

Table 7
Tuned hyperparameters of SVR models for biomass productivity (BP) and CO_2 biofixation (R_{CO2}).

Hyperparameters	BP	R_{CO2}
Kernel function	Gaussian	Gaussian
Epsilon (ϵ)	0.0054	0.0082
Box Constraint	7.8142	1.0998
Kernel Scale	0.0003	0.0333

search algorithm was coupled with hybrid BOA-SVR for multi-objective optimization using MATLAB (version R-2019a) environment.

3. Results and discussion

3.1. Development of hybrid BOA-BRT models

Hybrid BOA-BRT models for biomass productivity (BP) and CO_2 biofixation (R_{CO2}) were developed using real-life laboratory data given in Table 3 (for BP), and Table 4 (for R_{CO2}), respectively. The inputs used in the boosted regression tree are T, LD, and N/P. The performance of the model heavily depends on the hyperparameters (viz., method, number of the ensemble learning cycle, leaf size, and learning rate). In this regard, an integrated 5-fold cross-validation and BOA approach was applied to obtain the optimal values of all hyperparameters, which provides the optimum BRT model automatically. The progress of the tuning of hyperparameters is depicted in Fig. 4. The best observed viable position is a point where the value of the observed objective score is a minimum. The values of the minimum observed objective for BP (see Fig. 4a) and R_{CO2} (see Fig. 4b) are 0.003796 and 0.003183, at iterations of 16 and 47, respectively. The optimized hyperparameters used for the optimized BRT models are shown in Table 5. Of two available ensemble methods, namely bootstrap aggregation (Bag) and least-squares boosting (LSBoost), the second one was chosen due to its high performance. LSBoost is one of the popular adaptive boosting (AdaBoost) learning algorithms. A noteworthy aspect of AdaBoost is that it uses multiple iterations to generate a single compound strong learner (model with small RMSE). AdaBoost produces a strong learner by iteratively adding weak learners (model with large RMSE). During each cycle of training, a new weak learner is added to the ensemble and a weighting vector is adapted to reach the minimum RMSE.

The number of ensemble learning cycles is one of the vital hyperparameters of BRT models, which provides the minimum number of cycles required to transform weak learners (large RMSE) into strong learners (low RMSE). The optimum number of ensemble learning cycles found using Bayesian optimization for BP, and R_{CO2} are 13 and 12, respectively. The learning rate (lr) is used to shrink the impact of each tree as it is added to the model. A slow lr increases the number of trees, while a fast lr provides relatively few trees and does not attain the

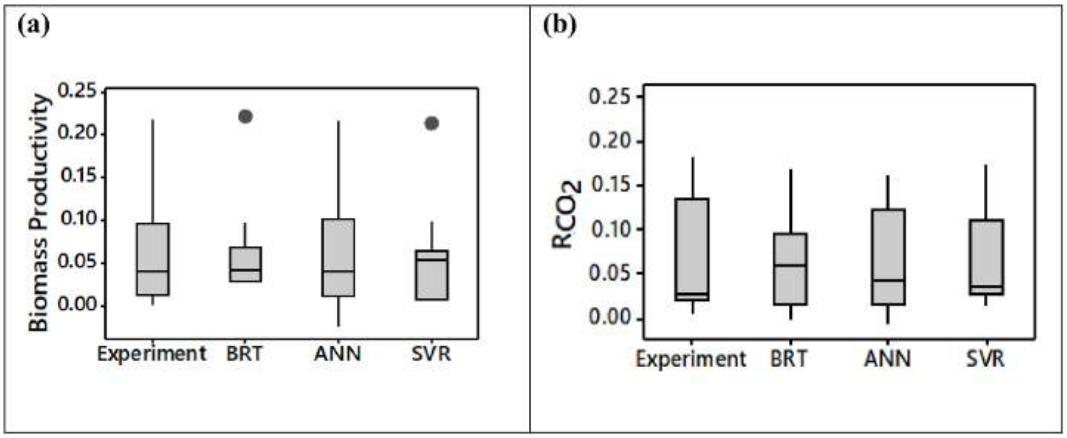

Fig. 6. Box plots for (a) biomass productivity (BP) and (b) CO_2 biofixation (R_{CO2}) via hybrid intelligence approaches.

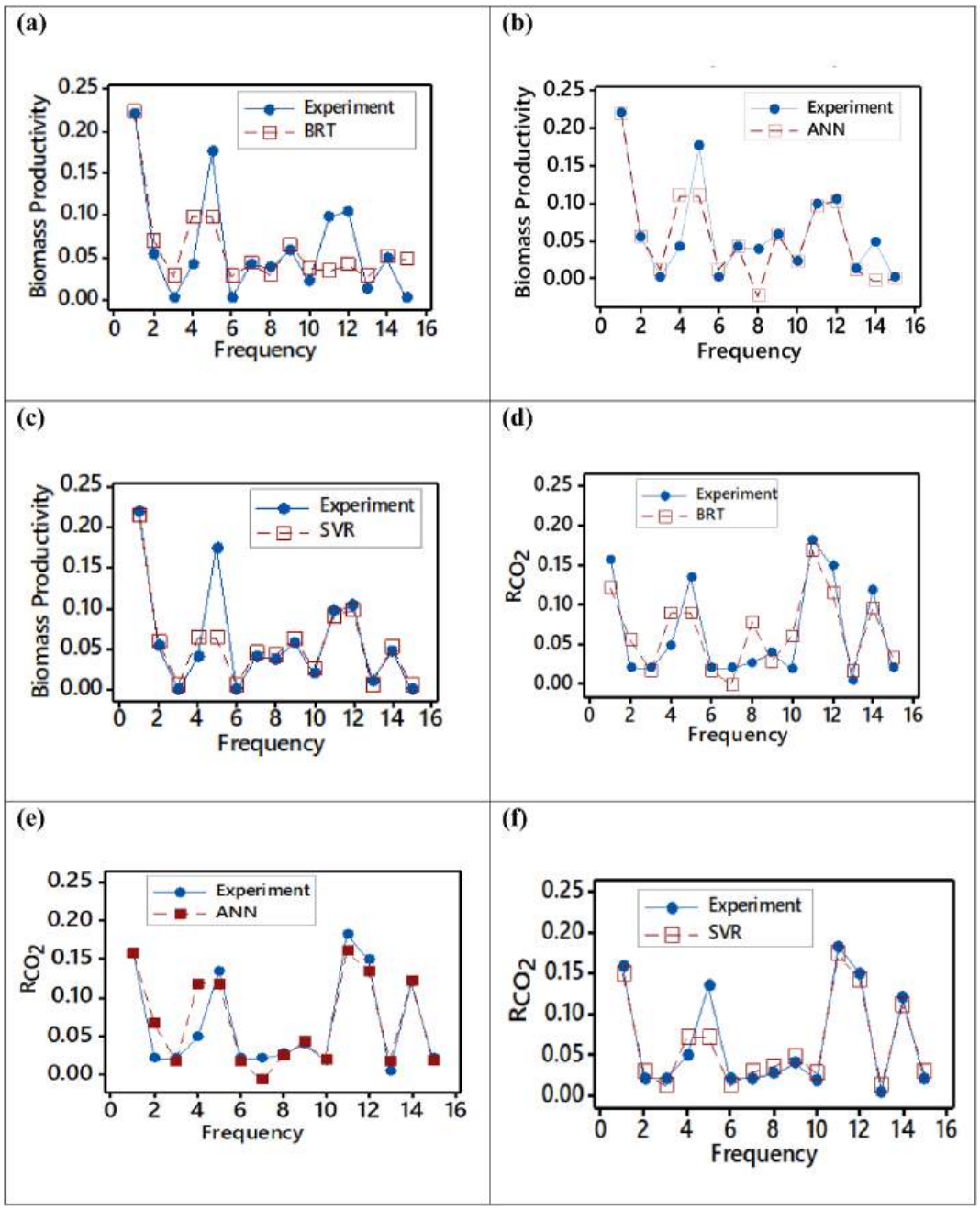

Fig. 7. Comparison of experimental and predicted values of three artificial intelligence approaches for all responses (e.g., BP and R_{CO2}) to frequency.

minimum error. Thus, finding an optimum learning rate is crucial. The optimum learning rates for BP and R_{CO2} are 0.95566, and 0.52588, respectively (see Table 5). The leaf size is the number of cases or observations in any leaf. A small leaf size provides more splits, resulting in a deep tree, which may cause overfitting. In contrast, a large leaf size stops tree growing after one split or a few splits, resulting in poor predictive performance. The optimum leaf size for BP and R_{CO2} are 6 and 1, respectively (see Table 5).

3.2. Development of hybrid BOA-ANN models

Hybrid BOA-ANN models were implemented for the estimation of microalgae BP and R_{CO2}. For the ANN model development, a feed-forward back propagation neural network was selected to anticipate BP and R_{CO2}. The network comprises three inputs and one output. The inputs are T, LD, and N/P affecting either BP or R_{CO2} as the response variables. Real-life laboratory data, listed in Table 3 (BP), and Table 4 (R_{CO2}), were allocated randomly to training, testing, and validation datasets. Training data were applied to estimate model parameters (neuron weights and biases), while the validation data were employed to determine network generalization.

From several training algorithms (gradient descent with momentum, gradient descent, quasi-Newton, scaled conjugate gradient, and Levenberg-Marquat), Levenberg-Marquat (trainlm) was chosen based on the high prediction performance of the ANN model. An important aspect of the performance of the ANN model is that it depends mainly on the proper choice of its hyperparameters (e.g., learning rate, number of neurons in the hidden layer, etc.). An integrated 5-fold cross-validation and BOA approach was applied to obtain the optimum values of all hyperparameters, which automatically provide the optimum configuration for the neural network. Fig. 4 depicts the progress of the Bayesian hyperparameter optimization and the best feasible location is at a point

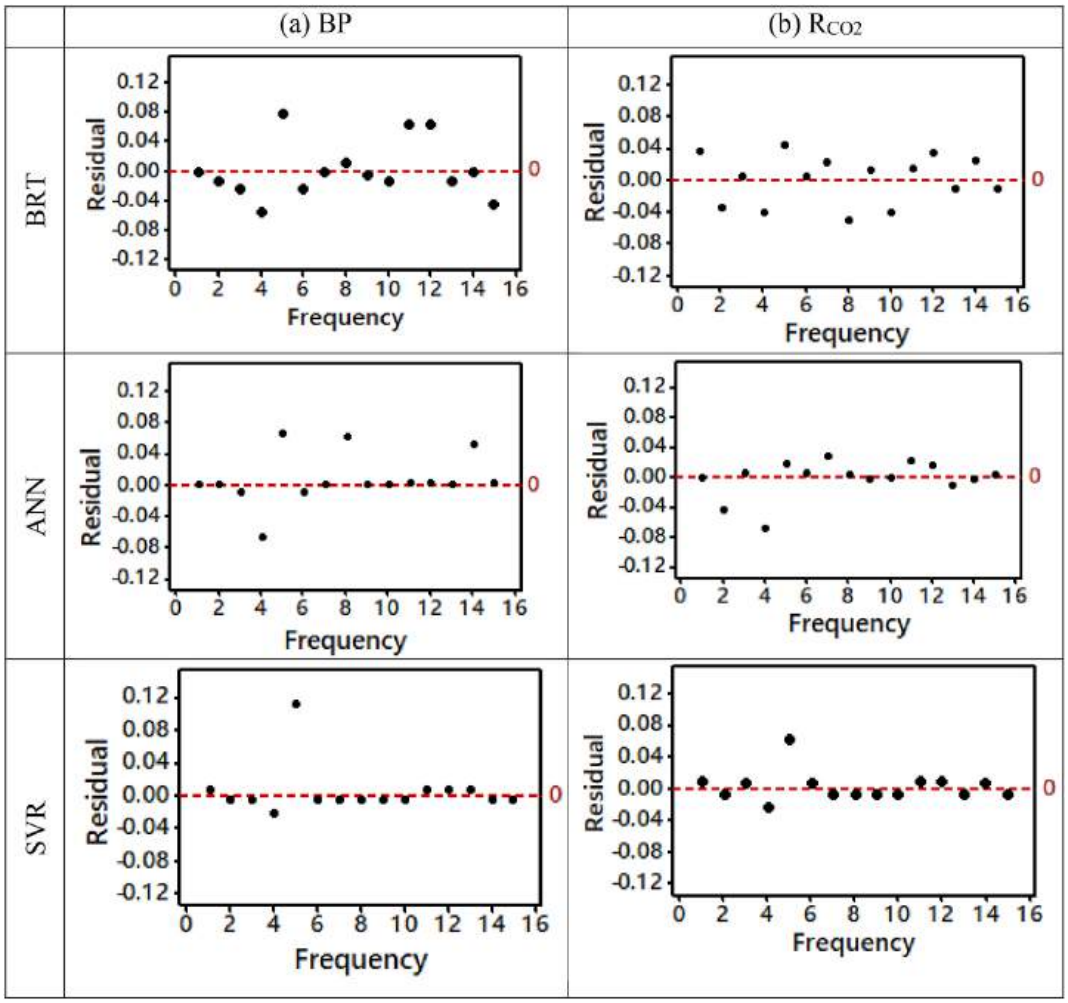

Fig. 8. Residual versus frequency figures for (a) biomass productivity (BP) and (b) CO_2 biofixation (R_{CO2}) via different artificial intelligence approaches.

where the value of the objective score is a minimum. The minimum values of the observed objective for BP (see Fig. 4c) and R_{CO2} (see Fig. 4d) are 0.085811 and 0.030664 at iterations 34 and 27, respectively. The optimized hyperparameters of the established ANN model for both BP and R_{CO2} are shown in Table 6.

The learning rate (or step length for the weights update) is one of the most important hyperparameters for ANN models. Generally, the value of the learning rate varies from 0.0 to 1.0. A very high learning rate may lead the model to converge too rapidly to a suboptimal solution, whereas a very low value can cause the process to hang up. Thus, a trade-off between high and low values is needed. The optimum learning rates determined using Bayesian optimization for BP, and R_{CO2} are 0.973 and 0.001, respectively. Fig. 5 shows the optimized ANN structures for all responses (BP, R_{CO2}) comprising three input neurons and one output neuron. Ten and seven neurons in the hidden layer are optimum for BP (see Fig. 5a), and R_{CO2} (see Fig. 5b), respectively.

3.3. Development of hybrid BOA-SVR models

Hybrid BOA-SVR models were developed to predict microalgae BP and R_{CO2}. In this respect, real-life experimental data (15 sets of input-output pairs), listed in Table 3 (BP) and Table 4 (R_{CO2}), were used. A noteworthy aspect is that the performance of the SVR model heavily depends on the proper selection of the relevant hyperparameters, viz., the kernel function, epsilon (ε), the box constraint (C), and kernel scale (γ). A high or low score of these hyperparameters may lead to over- or under-fitting. Thus, the kernel function type and the values of ε, γ, and C were optimized using the BOA and the predicting abilities of the models were evaluated. Fig. 4 shows the improvement of the tuning of the SVR hyperparameters as well as the optimal set. The scores for the minimum objectives observed for BP (see Fig. 4e) and R_{CO2} (see Fig. 4f) are 0.0048307, and 0.0038982 at iteration numbers of 32 and 27, respectively. The maximum accuracy was attained automatically for BP and R_{CO2} with the Gaussian kernel function. The optimized SVR models were obtained using the optimized hyperparameters listed in Table 7.

3.4. Effectiveness of the models

The estimated results obtained from the three hybrid intelligence approaches for both BP and R_{CO2} are listed in Tables 3 and 4, respectively. These modeling techniques were analyzed and compared to one another to determine the best predictive approach offering the best results with the highest level of accuracy.

Prior to this investigation, it is important to identify the distribution of responses acquired in both laboratory trials and by model predictions. A boxplot is a well-known graphical presentation demonstrating the dispersion of data based on a 5-number summary (viz., minimum score, lower quartile, median, upper quartile, and maximum score). Fig. 6 indicates the boxplots of experimental and predicted responses for both BP (Fig. 6a) and R_{CO2} (Fig. 6b). Supplementary Table S1 provides the descriptive statistics involving mean and standard deviation. The results indicate that the distribution patterns of laboratory and predicted data for all predictive models are similar and right-skewed except the SVR model data in the case of BP and BRT model data in the case of R_{CO2}. The

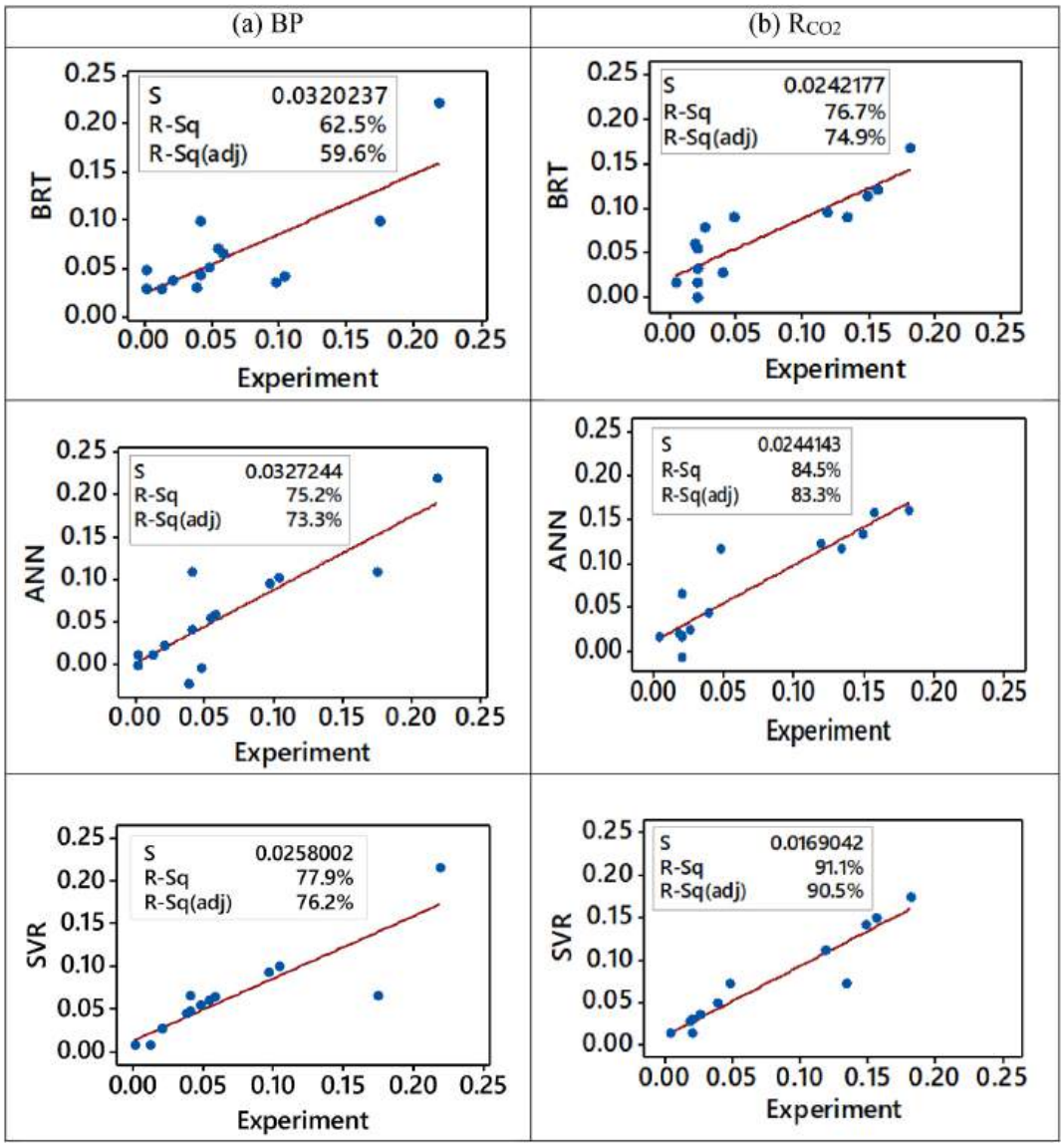

Fig. 9. Comparison of experimental and anticipated data for (a) biomass productivity (BP) and (b) CO_2 biofixation (R_{CO2}) using hybrid intelligence approaches. Here, R-Sq, R-Sq (adj), and S represent the coefficient of determination, the adjusted coefficient of determination, and the standard deviation, respectively.

Table 8
Model performances with MAE, MARE, RMSE, and FB values.

Criterion	Removal parameter	BRT model	ANN model	SVR model	Improvement in SVR wrt BRT (%)	Improvement in SVR wrt ANN (%)
MAE	BP	0.0284	0.0185	0.0136	52.11	26.55
	R_{CO2}	0.0259	0.0152	0.0128	50.56	15.87
MARE	BP	6.9454	1.8420	1.2587	81.87	31.66
	R_{CO2}	0.8888	0.6140	0.4131	53.52	32.72
RMSE	BP	0.0379	0.0324	0.0296	21.89	8.68
	R_{CO2}	0.0300	0.0241	0.0189	36.95	21.57
FB	BP	0.000	0.1183	0.0061	–	–
	R_{CO2}	0.0001	0.0385-	0.0088	–	–

median lines were observed within the boxes indicating that there is no difference between the datasets. The box lengths for both BP (~0.0401–0.0906) and R_{CO2} (~0.0.0798–0.114) strengths do not vary significantly, resulting in data that are not distributed broadly. Moreover, outliers are not present in the case of R_{CO2} (for all models), while only two outliers are present in the case of BP, obtained separately by BRT and SVR models.

The effectiveness of the developed models (e.g., BRT, ANN, and SVR) was assessed based on several performance indicators, viz., RE, R^2, MAE, MARE, RMSE, and FB. The values of RE for BP and R_{CO2} are listed in Tables 3 and 4, respectively. The BRT model provides the highest RE compared to both ANN and SVR, while ANN and SVR models provide comparable RE values for all responses (BP and R_{CO2}). Fig. 7 shows a comparison of experimental and predicted values of the three artificial intelligence approaches for all responses (e.g., BP and R_{CO2}. The results obtained using the ANN and SVR models are comparable, while many

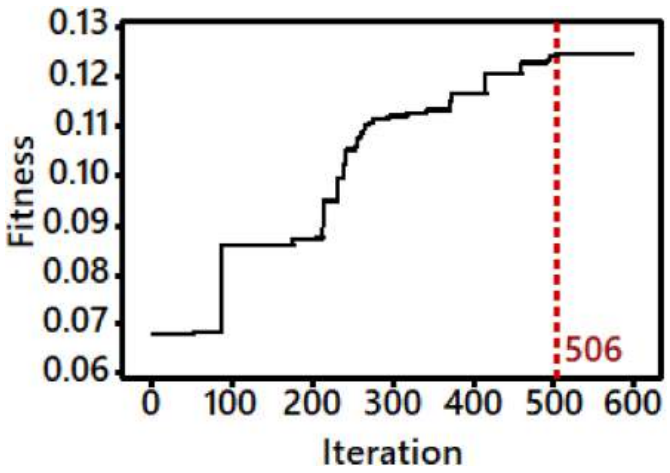

Fig. 10. Concurrent convergence plot for BP and R_{CO2} when multiobjective optimization is used. Stable values of BP and R_{CO2} were obtained at 506 iterations.

data points predicted by the BRT model do not agree with experimental data. Almost all experimental data points agree well with the corresponding output of the SVR model, indicating that the SVR model has a much better prediction capability. The plots of residuals, shown in Fig. 8, indicate this aspect clearly. The residuals of the data points obtained using the SVR and ANN models for BP (see Fig. 8a), and RCO_2 (see Fig. 8b) are randomly distributed around reference 0 line, indicating comparable capabilities of both models. However, SVR model estimates are closer to the experimental data.

Fig. 9 shows the fitted line plot, which describes the relationship between the experimental data and outputs predicted by BRT, ANN, and SVR. The R^2 values for BP (see Fig. 9a) obtained using the BRT, ANN, and SVR models are 0.625, 0.7524, and 0.7787, respectively, while these values for R_{CO2} (see Fig. 9b) are 0.7674, 0.8449, and 0.9114, respectively. High R^2 values are provided by the SVR model for all responses (e.g., BP and R_{CO2}), compared to those provided by the BRT and ANN models. A performance improvement for BP of 24.59% was provided by the SVR model compared to the BRT model, while it is 3.45% for the ANN model. Similarly, a performance improvement of 18.76% for R_{CO2} was provided by SVR compared to that provided by BRT, while it is 7.87% compared to that provided by ANN. The data indicate that the performance of both ANN and SVR is comparable for BP and R_{CO2}. However, the overall results indicate that the SVR predictions agree well with the experimental data.

The values of other performance grading measures, including MAE, MARE, RMSE, and FB are provided in Table 8. The SVR model provides better results for BP and R_{CO2} compared to the BRT and ANN models. Using MAE to assess the performance of the models, the SVR model outperformed the BRT and ANN models for the prediction of BP with an improvement of 52.11% and 26.55%, respectively. While for the prediction of R_{CO2}, the enhancement of the performance of the SVR model over BRT and ANN models is 50.56% and 15.87%, respectively. Based on MARE, the performance of the SVR model for the prediction of BP and R_{CO2} is better than that of BRT and ANN with an enhancement of 81.87% and 31.66%, and 53.52% and 32.72%, respectively. Also, based on RMSE, the performance of SVR is higher than that of the BRT and ANN models for the prediction of both BP and R_{CO2} with a performance enhancement of 21.89% and 8.68%, and 36.95% and 21.57%, respectively. The performance of a model is satisfactory if the value of absolute fractional bias, $|FB| \leq 0.5$ [60]. The FB values of all hybrid intelligence approaches for BP and R_{CO2} are around zero (see Table 8). The results indicate that the predictions of all the models used in this investigation are highly reliable. However, based on the overall performance, the results confirm that the predictions of the SVR model are in closer agreement with real-life laboratory data compared to the predictions of the other AI-based models (viz., ANN and BRT).

Furthermore, to compare the performance of the AI-based SVR model with a statistical model, both BP and R_{CO2} were predicted using the RSM methodology with BBD. In this regard, the experimental data obtained using the BBD matrix (see Table 2) were assessed using multiple regression. Coded independent factors were used for the development of models to ensure that the design of the experiment (DoE) is orthogonal [61,62]. A noteworthy aspect of this method is that orthogonality is one of the characteristics of bias-free assessment. Regression analyses provide models described by Eqs. (18) and (19) for predicting BP and R_{CO2}, respectively.

$$BP = 0.0047–0.0201\ x_1 + 0.0229\ x_2 - 0.0233\ x_3 + 0.0093\ x_1^2 + 0.0713\ x_2^2 + 0.0245\ x_3^2 - 0.0543\ x_1x_2 + 0.0090\ x_1x_3 - 0.0280\ x_2x_3 \quad (18)$$

$$R_{CO2} = 0.0147–0.0043\ x_1 + 0.0029\ x_2 - 0.0219\ x_3 - 0.0118\ x_1^2 + 0.0844\ x_2^2 + 0.0219\ x_3^2 - 0.0542\ x_1x_2 - 0.0047\ x_1x_3 + 0.0295\ x_2x_3 \quad (19)$$

where, BP and R_{CO2} denote the biomass productivity and CO_2 biofixation rate, respectively; x_1, x_2, and x_3 indicate the coded input parameters of the natural T, N/P ratio, and LD, respectively.

The results obtained using the BBD methodology are compared with those obtained using the optimized SVR model based on the predicted coefficient of determination (R^2_{pred}). A point that should be noted is that R^2_{pred} determines how well a model estimates new observations. This parameter prevents overfitting as it is determined using experimental data instead of model predictions. The R^2_{pred} values for BP and R_{CO2} obtained using BBD are 0.599 and 0.705, respectively. While these values obtained using the SVR model are 0.597 and 0.826, respectively. The ability of SVR and BBD approaches in predicting BP is comparable. However, the SVR model predicts R_{CO2} more accurately compared to the BBD approach with an enhancement of the performance of 17.16%. Besides, MARE for the prediction of BP and R_{CO2} by BBD is 2.9901 and 0.7409, respectively, which are comparatively higher than those obtained with the SVR model (see Table 8). The results indicate that the established SVR model performs very well in predicting unseen data. Overall, the results indicate that the prediction ability of the SVR model is better than that of the BBD model. Another noteworthy aspect is that the SVR model can perform well even with a small dataset [36,63]. The

Table 9
A comparison of the results of this study with those in reported studies.

Culture Media	Microalgae	Temp. (°C)	CO_2 (%)	Light Intensity (μmol $m^{-2}s^{-1}$)	BP (g $L^{-1}d^{-1}$)	R_{CO2} (g $L^{-1}d^{-1}$)	Ref.
Modified BG11	*Chlorella pyrenoidosa*	25	5	180	0.133	0.244	[72]
BG11 with N source	*Auxenochlorella pyrenoidosa*	40	4	148	0.134	–	[67]
Synthetic Municipal wastewater	*Neochloris oleoabundans*	25	6	60	0.100	0.145	[71]
OSPW	*Chlorella kessleri*	21	28	70	–	0.063	[75]
Synthetic Municipal wastewater	*Chlorella vulgaris*	25	4	60	0.078	0.200	[70]
Modified BBM	*Chlorella kessleri*	25	10	65	0.045	0.077	[76]
Modified BBM	*Chlorella vulgaris*	25	4	23	0.079	0.182	[21]
Modified BBM	*Chlorella vulgaris*	30	4	65	0.147	0.142	[22]
Modified BBM	*Chlorella vulgaris*	25	3	65	0.092	0.065	[23]
Modified BBM	*Chlorella vulgaris*	40	4	65	0.098	0.141	This study

findings of this study are consistent with this observation.

3.5. Multi-objective optimization using CSA

Multi-objective optimization approach is more suitable for obtaining the optimal operating conditions for processes with more than one response. A single optimal point of the input parameters is obtained for several responses via trade-offs [64–66]. In this study, the best hybrid BOA-SVR model was applied with CSA to obtain the optimal conditions for both responses (e.g., BP and R_{CO2}) simultaneously. Fig. 10 shows a convergence plot that is used to maximize BP and R_{CO2} concurrently. Normally, a convergence plot produces a stable output resulting from the collective impacts of all independent parameters. The red broken line in the Figure indicates the iteration number and optimal point where both responses BP and R_{CO2} are maximized. The response of iterations below the optimal point is unstable and it becomes stable from the 506th iteration. At the optimal point, the optimum coded set of T, N/P, and LD are 1, -1, and 0.00, which are equivalent to 40 °C, 1:1 of N/P, and 12/12 h/h of LD, respectively. At the optimal point, the maximum BP and R_{CO2} are 0.0979 g $L^{-1}d^{-1}$ and 0.1408 g $L^{-1}d^{-1}$, respectively. Some studies have evaluated the impact of temperature on microalgae growth as well as CO_2 biofixation [67,68]. A significant extent of BP resulted from *Auxenochlorella pyrenoidosa* microalgae at 40 °C [67]. However, with the increase of temperature (>40 °C) biomass productivity decreases. An LD cycle of 12/12 (h/h) is close to the daily cycle, while an N/P ratio of 1:1 is very close to that in the secondary discharge of municipal wastewater treatment plants (rich in nitrate-nitrogen and phosphate-phosphorus) [69]. Thus, the optimal conditions determined in this study can easily be adapted to outdoor cultivation of microalgae using wastewater available in the Middle East region without controlling any parameters. Elevated temperature (~35–40 °C) with an LD cycle close to 12/12 (h/h) prevails in the Middle East region almost throughout the year.

3.6. Validation of the optimal set

A set of triplicate laboratory trials was conducted with the optimum operating conditions (40 °*C*, 12/12 h/h of LD, and 1:1 of N/P) for validation. The percentage error concerning the predicted value is less than 5% in both cases (e.g., maximum BP and R_{CO2}), suggesting that the optimal point attained by the hybrid BOA-SVR-CSA is consistent and reliable. The optimal values of both responses BP and R_{CO2} were compared with those obtained in several reported studies, although AI-based advanced modeling and optimization of data for these responses are limited in the literature. A summary of such a comparison is presented in Table 9. Some values of BP and R_{CO2} of *Chlorella vulgaris* in the current investigation are either less, comparable, or higher than those reported [21–23,70]. Razzak [71] has reported similar values for BP and R_{CO2} using *Neochloris oleoabundans* microalgae. While Tang et al. [72] found a higher amount of BP and R_{CO2} using the *Chlorella pyrenoidosa* microalgae. A comparable BP (0.092 g $L^{-1}d^{-1}$) has been reported using an RSM-based investigation, while the R_{CO2} uptake (0.065 g $L^{-1}d^{-1}$) reported in the same study is lower than that of the current study [23]. Pires et al. (2013) have reported that *Chlorella vulgaris* produces a high R_{CO2} uptake of 0.305 g L^{-1} d^{-1} at room temperature and under continuous fluorescent light (LD: 24/0) with a light intensity of 72 μmol m^{-2} s^{-1} [73]. Hariz et al. (2018) have reported a CO_2 fixation rate of 0.1208 g L^{-1} d^{-1} (which is comparable to that of the current study) under optimum conditions of 10% v/v of CO_2, 1670 mL min^{-1} of aeration rate, and 24.8% v/v inoculum dose [74]. Kasiri et al. (2015) [75] have reported a maximum CO_2 uptake rate of 0.063 g L^{-1} d^{-1} (which is lower than that of the current study) by *Chlorella kessleri* sp. at the optimum conditions of 28% of CO_2, 29 mM of phosphate, and 70 μmol m^{-2} s^{-1} of irradiation. Omar et al. [76] have reported a maximum BP and CO_2 uptake rate of 0.045 g L^{-1} d^{-1} and 0.077 g L^{-1} d^{-1}, respectively (which are lower than that of the current study) by *Chlorella kessleri* under 25 °C, 10% CO_2, and irradiation of 65 μmol m^{-2} s^{-1}. The inconsistencies of the responses can be justified as the microalgae strain type, growth conditions, and bioreactor type of this study are completely different from those used in others. In addition, a systematic statistical design of the experimental technique is not used in most of the other studies.

Although L/D culture, N/P ratio, and temperature are the three main important parameters evaluated in this study, further studies are required to determine the impact of other parameters in open/closed cultivation in the outdoor environment. These are pH and light irradiation impacts on the culture conditions, reactor design, effects of scaling up to the pilot plant scale to industrial photobioreactor design, and applications.

4. Conclusions

The effects of three essential process variables, viz., temperature, LD cycle, and NP ratio of the culture on the biomass productivity (BP) and CO_2 biofixation (RCO_2) were evaluated using *Chlorella vulgaris*. Three hybrid intelligence approaches, viz., BOA-BRT, BOA-ANN, and BOA-SVR were applied to generate novel predictive models to predict microalgae biomass productivity (BP) and CO_2 biofixation (R_{CO2}). BOA was combined with each intelligence technique to tune hyperparameters and thereby optimize the models. The utility of these models was evaluated via several performances quantifying indicators, viz., RE, R^2, MAE, MARE, RMSE, and FB. The data indicate that the performance of the ANN and SVR models is comparable. However, based on the overall results, the hybrid SVR model performs better with a higher prediction precision than ANN. With the SVR model, a low MAE (BP: 0.0136, and R_{CO2}: 0.0128), MARE (BP: 1.2587, and R_{CO2}: 0.4131), and RMSE (BP: 0.0296, and R_{CO2}: 0.0189) values and high R^2 values (BP: 0.7787, and R_{CO2}: 0.9114) were obtained. Hence, the technique based on the SVR model is a more efficient analysis and diagnosis platform for simulating and assessing the non-linear nature of microalgae biomass productivity and CO_2 biofixation. Furthermore, the performance of the developed SVR model was compared with the statistical BBD model in terms of the R^2_{pred} and MARE. In SVR, a performance improvement of 17.16% (concerning the R^2_{pred}) with a lower MARE value was observed for R_{CO2} in comparison to BBD. Then, CSA was hybridized into the hybrid BOA-SVR model to obtain tuned variables. High biomass productivity and CO_2 biofixation of 0.0979 g $L^{-1}d^{-1}$ and 0.1408 g $L^{-1}d^{-1}$, respectively, under optimal conditions of 40 °*C*, 12/12 h/h of LD, and 1:1 of N/P were obtained. Finally, the predicted optimal values were confirmed by comparing them with experimental data for both responses with less than a 5% error. Overall, this novel platform can be readily applied as a key performance evaluating tool to other CO_2 sequestration and utilization (CCU) processes.

Credit to authors

S. M. Zakir Hossain - Conceptualization, methodology, writing original draft of the manuscript; N. Sultana - Fund acquisition, formal analysis, software, methodology; S.A. Razzak - Contributed to writing, reviewing and editing the manuscript; M. M. Hossain – Leading and managing the overall tasks, reviewing and editing the manuscript.

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.

Acknowledgments

The authors would like to thank the Deputyship for Research &

Innovation, Ministry of Education in Saudi Arabia for funding this research work through project number RDO-2019-001-CSIT.

Nomenclature

ANN	Artificial Neural Network
BOA	Bayesian Optimization Algorithm
BBD	Box-Behnken Design
BBM	Bold's Basal Medium
BRT	Boosted Regression Tree
BP	Biomass Productivity
FB	Fractional Bias
LD	Light Dark cycle
MAE	Mean Absolute Error
MARE	Mean Absolute Relative Error
MW	Molecular Weight
NP	Nitrogen–Phosphorus ratio
OSPW	Oil Sands Process Water
R^2	Coefficient of determination
RE	Relative Error
RMSE	Root Mean Square Error
RSM	Response Surface Methodology
R_{CO2}	CO_2 biofixation
RT	Regression Tree
SVR	Support Vector Regression
μ	Specific growth rate

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rser.2021.112016.

References

[1] Bilanovic D, Andargatchew A, Kroeger T, Shelef G. Freshwater and marine microalgae sequestering of CO_2 at different C and N concentrations - response surface methodology analysis. Energy Convers Manag 2009;50:262–7. https://doi.org/10.1016/j.enconman.2008.09.024.

[2] Zhao X, Zhou H, Sikarwar VS, Zhao M, Park AHA, Fennell PS, et al. Biomass-based chemical looping technologies: the good, the bad and the future. Energy Environ Sci 2017;10:1885–910. https://doi.org/10.1039/c6ee03718f.

[3] IEA. Global CO2 emissions in 2019. 2019. https://www.iea.org/articles/global-co2-emissions-in-2019. [Accessed 30 December 2021].

[4] Hanak DP, Anthony EJ, Manovic V. A review of developments in pilot-plant testing and modelling of calcium looping process for CO_2 capture from power generation systems. Energy Environ Sci 2015;8:2199–249. https://doi.org/10.1039/c5ee01228g.

[5] Boot-Handford ME, Abanades JC, Anthony EJ, Blunt MJ, Brandani S, Mac Dowell N, et al. Carbon capture and storage update. Energy Environ Sci 2014;7:130–89. https://doi.org/10.1039/c3ee42350f.

[6] Aghaie M, Rezaei N, Zendehboudi S. A systematic review on CO_2 capture with ionic liquids: current status and future prospects. Renew Sustain Energy Rev 2018;96:502–25. https://doi.org/10.1016/j.rser.2018.07.004.

[7] Song C, Liu Q, Ji N, Deng S, Zhao J, Li Y, et al. Alternative pathways for efficient CO_2 capture by hybrid processes—a review. Renew Sustain Energy Rev 2018;82:215–31. https://doi.org/10.1016/j.rser.2017.09.040.

[8] Molazadeh M, Ahmadzadeh H, Pourianfar HR, Lyon S, Rampelotto PH. The use of microalgae for coupling wastewater treatment with CO_2 biofixation. Front Bioeng Biotechnol 2019;7. https://doi.org/10.3389/fbioe.2019.00042.

[9] Rochelle GT. Conventional amine scrubbing for CO_2 capture. Absorb Based Post Combustion capture Carbon Dioxide 2016:35–67. https://doi.org/10.1016/B978-0-08-100514-9.00003-2.

[10] Razzak SA, Hossain MM, Lucky RA, Bassi AS, De Lasa H. Integrated CO_2 capture, wastewater treatment and biofuel production by microalgae culturing - a review. Renew Sustain Energy Rev 2013;27:622–53. https://doi.org/10.1016/j.rser.2013.05.063.

[11] Hossain SMZ, Razzak SA, Al-Shater AF, Moniruzzaman M, Hossain MM. Recent advances in enzymatic conversion of microalgal lipids into biodiesel. Energy Fuel 2020;34:6735–50. https://doi.org/10.1021/acs.energyfuels.0c01064.

[12] Hossain SMZ. Biochemical conversion of microalgae biomass into biofuel. Chem Eng Technol 2019;42:2594–607. https://doi.org/10.1002/ceat.201800605.

[13] Razzak SA, Ali SAM, Hossain MM, deLasa H. Biological CO_2 fixation with production of microalgae in wastewater – a review. Renew Sustain Energy Rev 2017;76:379–90. https://doi.org/10.1016/j.rser.2017.02.038.

[14] Chen PH, Liu HL, Chen YJ, Cheng YH, Lin WL, Yeh CH, et al. Enhancing CO_2 bio-mitigation by genetic engineering of cyanobacteria. Energy Environ Sci 2012;5:8318–27. https://doi.org/10.1039/c2ee21124f.

[15] Lee SY, Park SJ. A review on solid adsorbents for carbon dioxide capture. J Ind Eng Chem 2015;23:1–11. https://doi.org/10.1016/j.jiec.2014.09.001.

[16] Aziz MMA, Kassim KA, Shokravi Z, Jakarni FM, Lieu HY, Zaini N, et al. Two-stage cultivation strategy for simultaneous increases in growth rate and lipid content of microalgae: a review. Renew Sustain Energy Rev 2020;119. https://doi.org/10.1016/j.rser.2019.109621.

[17] Razzak SA. Biomass and lipid productivity of Neochloris oleoabundans for CO_2 biofixation and biodiesel application. Chem Eng Technol 2018;41:2177–85. https://doi.org/10.1002/ceat.201800330.

[18] Razzak SA, Ilyas M, Ali SAM, Hossain MM. Effects of CO_2 concentration and pH on mixotrophic growth of nannochloropsis oculata. Appl Biochem Biotechnol 2015;176:1290–302. https://doi.org/10.1007/s12010-015-1646-7.

[19] Wu J-Y, Lay C-H, Chen C-C, Wu S-Y. Lipid accumulating microalgae cultivation in textile wastewater: environmental parameters optimization. J Taiwan Inst Chem Eng 2017;79:1–6. https://doi.org/10.1016/J.JTICE.2017.02.017.

[20] Chen C-Y, Chang Y-H. Engineering strategies for enhancing C. vulgaris ESP-31 lipid production using effluents of coke-making wastewater. J Biosci Bioeng 2018;125:710–6. https://doi.org/10.1016/J.JBIOSC.2018.01.008.

[21] Kazeem MAA, Hossain SMZ, Hossain MM, Razzak SA. Application of central composite design to optimize culture conditions of chlorella vulgaris in a batch photobioreactor: an efficient modeling approach. Chem Prod Process Model 2018;13. https://doi.org/10.1515/cppm-2017-0082.

[22] Hossain SMZ, Alnoaimi A, Razzak SA, Ezuber H, Al-Bastaki N, Safdar M, et al. Multiobjective optimization of microalgae (*Chlorella sp* .) growth in a photobioreactor using Box-Behnken design approach. Can J Chem Eng 2018;96:1903–10. https://doi.org/10.1002/cjce.23168.

[23] Hossain SMZ, Hossain MM, Razzak SA. Optimization of CO_2 biofixation by *Chlorella vulgaris* using a tubular photobioreactor. Chem Eng Technol 2018;41:1313–23. https://doi.org/10.1002/ceat.201700210.

[24] Chopra P, Sharma R, Applied MK-IJ. U. Artificial neural networks for the prediction of compressive strength of concrete. Int J Appl Sci Eng 2015;13:187–204. 2015.

[25] Gillard JW, Kvasov DE. Lipschitz optimization methods for fitting a sum of damped sinusoids to a series of observations. Stat Interface 2017;10:59–70. https://doi.org/10.4310/SII.2017.v10.n1.a6.

[26] Alade IO, Oyehan TA, Popoola IK, Olatunji SO, Bagudu A. Modeling thermal conductivity enhancement of metal and metallic oxide nanofluids using support vector regression. Adv Powder Technol 2018;29:157–67. https://doi.org/10.1016/j.apt.2017.10.023.

[27] Hameed M, Sharqi SS, Yaseen ZM, Afan HA, Hussain A, Elshafie A. Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. Neural Comput Appl 2017;28:893–905. https://doi.org/10.1007/s00521-016-2404-7.

[28] Rao KDVSK, Premalatha M, Naveen C. Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: a case study. Renew Sustain Energy Rev 2018;91:248–58. https://doi.org/10.1016/j.rser.2018.03.096.

[29] Naghibi SA, Pourghasemi HR, Dixon B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ Monit Assess 2016;188:1–27. https://doi.org/10.1007/s10661-015-5049-6.

[30] Yu H, Cooper AR, Infante DM. Improving species distribution model predictive accuracy using species abundance: application with boosted regression trees. Ecol Model 2020;432:109202. https://doi.org/10.1016/j.ecolmodel.2020.109202.

[31] Hu Y, Dai Z, Guldmann JM. Modeling the impact of 2D/3D urban indicators on the urban heat island over different seasons: a boosted regression tree approach. J Environ Manag 2020;266:110424. https://doi.org/10.1016/j.jenvman.2020.110424.

[32] Torres-Barrán A, Alonso Á, Dorronsoro JR. Regression tree ensembles for wind energy and solar radiation prediction. Neurocomputing 2019;326–327:151–60. https://doi.org/10.1016/j.neucom.2017.05.104.

[33] Kazem HA, Yousif J, Chaichan MT, Al-Waeli AHA. Experimental and deep learning artificial neural network approach for evaluating grid-connected photovoltaic systems. Int J Energy Res 2019;43:8572–91. https://doi.org/10.1002/er.4855.

[34] Sultana N, Hossain SMZ, Alam MS, Islam MS, Al-Abtah MA. Soft computing approaches for comparative prediction of the mechanical properties of jute fiber reinforced concrete. Adv Eng Software 2020;149:102887. https://doi.org/10.1016/j.advengsoft.2020.102887.

[35] Silitonga AS, Masjuki HH, Ong HC, Sebayang AH, Dharma S, Kusumo F, et al. Evaluation of the engine performance and exhaust emissions of biodiesel-bioethanol-diesel blends using kernel-based extreme learning machine. Energy 2018;159:1075–87. https://doi.org/10.1016/j.energy.2018.06.202.

[36] Adewunmi AA, Ismail S, Owolabi TO, Sultan AS, Olatunji SO, Ahmad Z. Hybrid intelligent modelling of the viscoelastic moduli of coal fly ash based polymer gel system for water shutoff treatment in oil and gas wells. Can J Chem Eng 2019;97:2969–78. https://doi.org/10.1002/CJCE.23436.

[37] Sedghamiz MA, Rasoolzadeh A, Rahimpour MR. The ability of artificial neural network in prediction of the acid gases solubility in different ionic liquids. J CO2 Util 2015;9:39–47. https://doi.org/10.1016/j.jcou.2014.12.003.

[38] Venkatraman V, Alsberg BK. Predicting CO_2 capture of ionic liquids using machine learning. J CO2 Util 2017;21:162–8. https://doi.org/10.1016/j.jcou.2017.06.012.

[39] Günay ME, Türker L, Tapan NA. Decision tree analysis for efficient CO_2 utilization in electrochemical systems. J CO2 Util 2018;28:83–95. https://doi.org/10.1016/j.jcou.2018.09.011.
[40] Sun Y, Yang G, Wen C, Zhang L, Sun Z. Artificial neural networks with response surface methodology for optimization of selective CO_2 hydrogenation using K-promoted iron catalyst in a microchannel reactor. J CO2 Util 2018;24:10–21. https://doi.org/10.1016/j.jcou.2017.11.013.
[41] Yan Q, Wang G. Prediction model of alga's growth based on support vector regression. Proc - 2009 Int Conf Environ Sci Inf Appl Technol ESIAT 2009 2009;2:673–5. https://doi.org/10.1109/ESIAT.2009.170.
[42] Wang Y, Xie Z, Lou I, Ung WK, Mok KM. Algal bloom prediction by support vector machine and relevance vector machine with genetic algorithm optimization in freshwater reservoirs. Eng Comput 2017;34:664–79. https://doi.org/10.1108/EC-11-2015-0356.
[43] Arumugam S, Chengareddy P, Tamilarasan A, Santhanam V. RSM and crow search algorithm-based optimization of ultrasonicated transesterification process parameters on synthesis of polyol ester-based biolubricant. Arabian J Sci Eng 2019;1–14. https://doi.org/10.1007/s13369-019-03847-1.
[44] Sayed GI, Hassanien AE, Azar AT. Feature selection via a novel chaotic crow search algorithm. Neural Comput Appl 2019;31:171–88. https://doi.org/10.1007/s00521-017-2988-6.
[45] Yusuf HA, Hossain SMZ, Khamis AA, Radhi HT, Jaafar AS. Optimization of CO_2 biofixation rate by microalgae in a hybrid microfluidic differential carbonator using response surface methodology and desirability function. J CO2 Util 2020;42:101291. https://doi.org/10.1016/j.jcou.2020.101291.
[46] Irfan M F MF, Hossain SMZ, Tariq I, Khan NA, Tawfeeqi A, Goeva A, et al. Modeling and optimization of aqueous mineral carbonation for cement kiln dust using response surface methodology integrated with box-behnken and central composite design approaches. Min Metall Explor 2020;37:1367–83. https://doi.org/10.1007/s42461-020-00222-9.
[47] Aslan N, Cebeci Y. Application of box-behnken design and response surface methodology for modeling of some Turkish coals. Fuel 2007;86:90–7. https://doi.org/10.1016/j.fuel.2006.06.010.
[48] Dimitratos SD, Hommel AS, Konrad KD, Simpson LM, Wu-Woods JJ, Woods DF. Biosensors to monitor water quality utilizing insect odorant-binding proteins as detector elements. Biosensors 2019;9:62. https://doi.org/10.3390/bios9020062.
[49] Feurer M, Hutter F. Hyperparameter optimization. Autom. Mach. Learn. Cham: Springer; 2019. p. 3–33. https://doi.org/10.1007/978-3-030-05318-5_1.
[50] Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: a review of Bayesian optimization. Proc IEEE 2016;104:148–75. https://doi.org/10.1109/JPROC.2015.2494218.
[51] Rasmussen CKIW CE. Gaussian processes for machine learning: book webpage. MIT press; 2006.
[52] Mockus J. Global optimization and the bayesian approach. D. Reidel Publishing Company; 1989. https://doi.org/10.1007/978-94-009-0909-0_1.
[53] Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst 2012;4:2951–9.
[54] Alade IO, Bagudu A, Oyehan TA, Rahman MAA, Saleh TA, Olatunji SO. Estimating the refractive index of oxygenated and deoxygenated hemoglobin using genetic algorithm – support vector regression model. Comput Methods Progr Biomed 2018;163:135–42. https://doi.org/10.1016/j.cmpb.2018.05.029.
[55] Owolabi TO, Akande KO, Olatunji SO. Development and validation of surface energies estimator (SEE) using computational intelligence technique. Comput Mater Sci 2015. https://doi.org/10.1016/J.COMMATSCI.2015.01.020. C:143–51.
[56] Sultana N, Hossain SMZ, Taher S, Khan A, Razzak SA, Haq B. Modeling and optimization of non-edible papaya seed waste oil synthesis using data mining approaches. S Afr J Chem Eng 2020;33:151–9. https://doi.org/10.1016/J.SAJCE.2020.07.009.
[57] Sultana N, Hossain SMZ, Alam MS, Hashish MMA, Islam MS. An experimental investigation and modeling approach of response surface methodology coupled with crow search algorithm for optimizing the properties of jute fiber reinforced concrete. Construct Build Mater 2020;243:118216. https://doi.org/10.1016/j.conbuildmat.2020.118216.
[58] Hossain SMZ, Taher S, Khan A, Sultana N, Irfan MF, Haq B, et al. Experimental study and modeling approach of response surface methodology coupled with crow search algorithm for optimizing the extraction conditions of papaya seed waste oil. Arabian J Sci Eng 2020;1–13. https://doi.org/10.1007/s13369-020-04551-1.
[59] Deb K. Multi-objective optimization using evolutionary algorithms. John Wiley & Sons; 2001.
[60] Kumar A, Bellam NK, Sud A. Performance of an industrial source complex model: predicting long-term concentrations in an urban area. Environ Prog 1999;18:93–100. https://doi.org/10.1002/ep.670180213.
[61] Montgomery DC. Design and analysis of experiments. 8 edition. Wiley; 2012. April 10, 2012.
[62] Oprime PC, Pureza VMM, De Oliveira SC. Systematic sequencing of factorial experiments as an alternative to the random order. Gest e Prod 2017;24:108–22. https://doi.org/10.1590/0104-530X1266-16.
[63] Almansour NA, Syed HF, Khayat NR, Altheeb RK, Juri RE, Alhiyafi J, et al. Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study. Comput Biol Med 2019;109:101–11. https://doi.org/10.1016/j.compbiomed.2019.04.017.
[64] Niu B, Wang H, Wang J, Tan L. Multi-objective bacterial foraging optimization. Neurocomputing 2013;116:336–45. https://doi.org/10.1016/j.neucom.2012.01.044.
[65] Rao, Ms K, Manisha M K, Roy DS. Hybrid disease diagnosis using multiobjective optimization with evolutionary parameter optimization. J Healthc Eng 2017;1:1–27.
[66] Deb K. Multi-objective optimisation using evolutionary algorithms: an introduction. In: Multi-objective evol. Optim. Prod. Des. Manuf. Springer London; 2011. p. 3–34. https://doi.org/10.1007/978-0-85729-652-8_1.
[67] Mello B, Chemburkar M. Effect of temperature and pH variation on biomass and lipid production of Auxenochlorella pyrenoidosa. Life Sci Informatics Publ 2018;4:378–87.
[68] Hossain SMZ, Al-Bastaki N, Alnoaimi AMA, Ezuber H, Razzak SA, Hossain MM. Mathematical modeling of temperature effect on algal growth for biodiesel application. Cham: Springer; 2020. p. 517–28. https://doi.org/10.1007/978-3-030-18488-9_41.
[69] Hossain SMZ, Sultana N, Mohammed ME, Razzak SA, Hossain MM. Hybrid support vector regression and crow search algorithm for modeling and multiobjective optimization of microalgae-based wastewater treatment. J Environ Manage. 2022;301:113783. https://doi.org/10.1016/j.jenvman.2021.113783.
[70] Razzak SA, Ali SAM, Hossain MM, Mouanda AN. Biological CO_2 fixation using Chlorella vulgaris and its thermal characteristics through thermogravimetric analysis. Bioproc Biosyst Eng 2016;39:1651–8. https://doi.org/10.1007/s00449-016-1640-7.
[71] Razzak SA. In situ biological CO_2 fixation and wastewater nutrient removal with Neochloris oleoabundans in batch photobioreactor. Bioproc Biosyst Eng 2019;42:93–105. https://doi.org/10.1007/s00449-018-2017-x.
[72] 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;102:3071–6. https://doi.org/10.1016/j.biortech.2010.10.047.
[73] Pires JCM, Gonçalves AL, Martins FG, Alvim-Ferraz MCM, Simões M. Effect of light supply on CO_2 capture from atmosphere by Chlorella vulgaris and Pseudokirchneriella subcapitata. Mitig Adapt Strategies Glob Change 2013;19:1109–17. https://doi.org/10.1007/S11027-013-9463-1. 197 2013.
[74] Hariz HB, Takriff MS, Ba-Abbad MM, Mohd Yasin NH, Mohd Hakim NIN. CO_2 fixation capability of Chlorella sp. and its use in treating agricultural wastewater. J Appl Phycol 2018;30:3017–27. https://doi.org/10.1007/S10811-018-1488-0. 306 2018.
[75] Kasiri S, Abdulsalam S, Ulrich A, Prasad V. Optimization of CO2 fixation by Chlorella kessleri using response surface methodology. Chem Eng Sci 2015;127:31–9. https://doi.org/10.1016/J.CES.2015.01.008.
[76] Faruque MO, Ilyas M, Hossain MM, Razzak SA. Influence of nitrogen to phosphorus ratio and CO_2 concentration on lipids accumulation of scenedesmus dimorphus for bioenergy production and CO_2 biofixation. Chem Asian J 2020;15:4307–20. https://doi.org/10.1002/asia.202001063.