Article

Application of Artificial Intelligence and Image Processing for the Cultivation of *Chlorella* sp. Using Tubular Photobioreactors

Thananop Tummawai, Thongchai Rohitatisha Srinophakun, Surapol Padungthon, and Somboon Sukpancharoen*
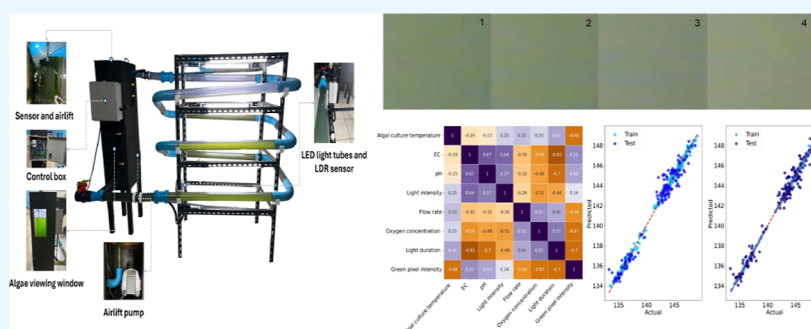
Cite This: ACS Omega 2024, 9, 46017−46029

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations

ABSTRACT: By integrating innovative technologies to enhance the efficiency and sustainability of production, this study specifies the establishment of a cutting-edge growing system for *Chlorella* sp. microalgae. Improvement of a system for the real-time, noninvasive observation and management of algae growth employing a closed tubular photobioreactor (PBR) engineered with computational fluid dynamics (CFD), combined with the Internet of things (IoT), artificial intelligence (AI), and image processing technologies was the major goal of this research. The fitting of seven types of sensors to identify key characteristics such as temperature, pH, light intensity, electrical conductivity (EC), flow rate, oxygen content, and light exposure duration was included in the research method. To manage the gaining of sensor data and system operations, an ESP8266 microcontroller was used as the main control unit, while 33×33 pixel images were taken with an ESP32 camera at 30 min intervals to assess growth by evaluating color intensity, enabling real-time evaluation of algal density without sampling or disturbing growth. Forecasting and enhancing farming situations was the goal of producing these machine learning (ML) models. Uniformly dispersed between 12 and 24 h light cycles, the data set comprised 602 samples. Considerable improvements were observed in the results for biomass productivity, with constant 24 h lighting yielding a 7.19% increase, counter to a 2.09% increase seen in the 12 h cycle. Temperature and light intensity are the most significant parameters for growth, as revealed by analysis of Feature Importance. The eXtreme Gradient Boosting (XGBoost) model showed remarkable effectiveness in terms of projecting growth, attaining an R^2 value of 0.9997 for the training data set. With important benefits for the development of renewable energy, food supply, and environmental modification in the future, this research highlights the competence of intelligent technology to strengthen microalgae production.

1. INTRODUCTION

Microalgae cultivation has emerged as a promising field with vast potential for addressing global challenges in sustainable energy, food security, and environmental remediation. These microscopic photosynthetic organisms have garnered significant attention due to their diverse applications in biofuels, food supplements, pharmaceuticals, and wastewater treatment.[1,2] Among the myriad of microalgal species, *Chlorella* sp. stands out as a particularly promising candidate. Its high protein content (40−60%), ability to thrive in both freshwater and saltwater environments, and adaptability to various cultivation systems make it an ideal subject for intensive research and commercial applications.[3,4]

The cultivation of microalgae, however, is not without its challenges. Successful production, particularly of species like *Chlorella* sp., relies heavily on optimizing growth conditions and meticulously monitoring key parameters to ensure optimal biomass production.[5] These challenges have led to the development of advanced cultivation systems, with photo-

Figure 1. Schematic diagram and components of the closed tubular PBR system for *Chlorella* sp. cultivation.

bioreactors (PBRs) emerging as a frontrunner in controlled microalgae cultivation technology.

PBRs, especially tubular designs, offer significant advantages over traditional open pond systems, including reduced risk of contamination, enhanced control over growth conditions, and substantially higher biomass yields.[6,7] The high surface-to-volume ratio of tubular PBRs allows for efficient light utilization and improves mass transfer, making them particularly suitable for high-density microalgae cultivation.[8,9] However, the complex nature of microalgal growth and the multitude of factors influencing their productivity necessitate the development of advanced monitoring and control systems to truly optimize the cultivation process.[4]

Recent technological advancements have opened new avenues for addressing these challenges and improving microalgae cultivation. The integration of Internet of things (IoT) technology and artificial intelligence (AI) in microalgal cultivation has the potential to revolutionize the industry. These technologies enable real-time monitoring, data-driven decision-making, and automated control of key parameters.[5,10] IoT sensors can continuously collect data on critical growth parameters such as temperature, pH, light intensity, and nutrient levels, providing invaluable insights into the cultivation process.[6,7] AI algorithms, in turn, can analyze these data to identify patterns, predict growth trends, and optimize control strategies, leading to improved biomass productivity and resource utilization.[8,9]

Complementing these advancements, image processing techniques have shown great promise in monitoring microalgal growth and assessing biomass concentration.[11,12] By analyzing images of the microalgal culture, changes in cell density, morphology, and color can be quantified, providing a noninvasive and cost-effective method for growth assessment.[13,14] The integration of image processing with IoT and AI technologies can further enhance the accuracy and reliability of growth monitoring, enabling real-time adjustments to cultivation conditions.[15]

Within the realm of AI, machine learning (ML) stands out as a key technique for advancing microalgae cultivation. ML

algorithms can process complex data from IoT sensors and image analysis to uncover hidden patterns and relationships in cultivation processes. These models can predict optimal growth conditions, forecast biomass yields, and even anticipate potential issues before they occur. ML techniques, such as Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and support vector machine,[16−18] offer the potential for more sophisticated modeling and control of microalgae systems. By continuously learning from new data, ML models can adapt to changing conditions and improve their predictions over time, leading to increasingly efficient and productive cultivation processes.

Previous studies on *Chlorella* sp. and other microalgae have focused on integrating AI and image-processing techniques for improved cultivation and monitoring. ML models have been developed for the accurate classification of microalgae species.[19,20] Image analysis has been utilized to optimize light distribution in PBRs[21,22] and monitor biomass growth.[23,24] Advanced techniques like multivariate image regression have been employed to estimate biomass concentration and chlorophyll content.[24] AI-driven approaches have also been applied to predict growth signatures and optimize cultivation conditions.[25] The integration of AI with microalgae research extends beyond cultivation, encompassing genetic optimization, system design, and product development.[26]

While these individual technological advancements have shown promise, there remains a significant gap in the integration of IoT, AI, and image processing into a comprehensive system for microalgae cultivation. Previous studies have explored these technologies separately, but few have attempted to combine them in a holistic approach. This study aims to address this gap by developing an advanced tubular PBR system for *Chlorella* sp. cultivation that integrates real-time IoT monitoring, AI-driven predictive modeling, and image-based growth assessment.[27]

Building upon these advancements, this research aims to develop and implement a novel, integrated system for optimizing *Chlorella* sp. cultivation. The primary objectives of this study are (1) to design and construct a tubular PBR
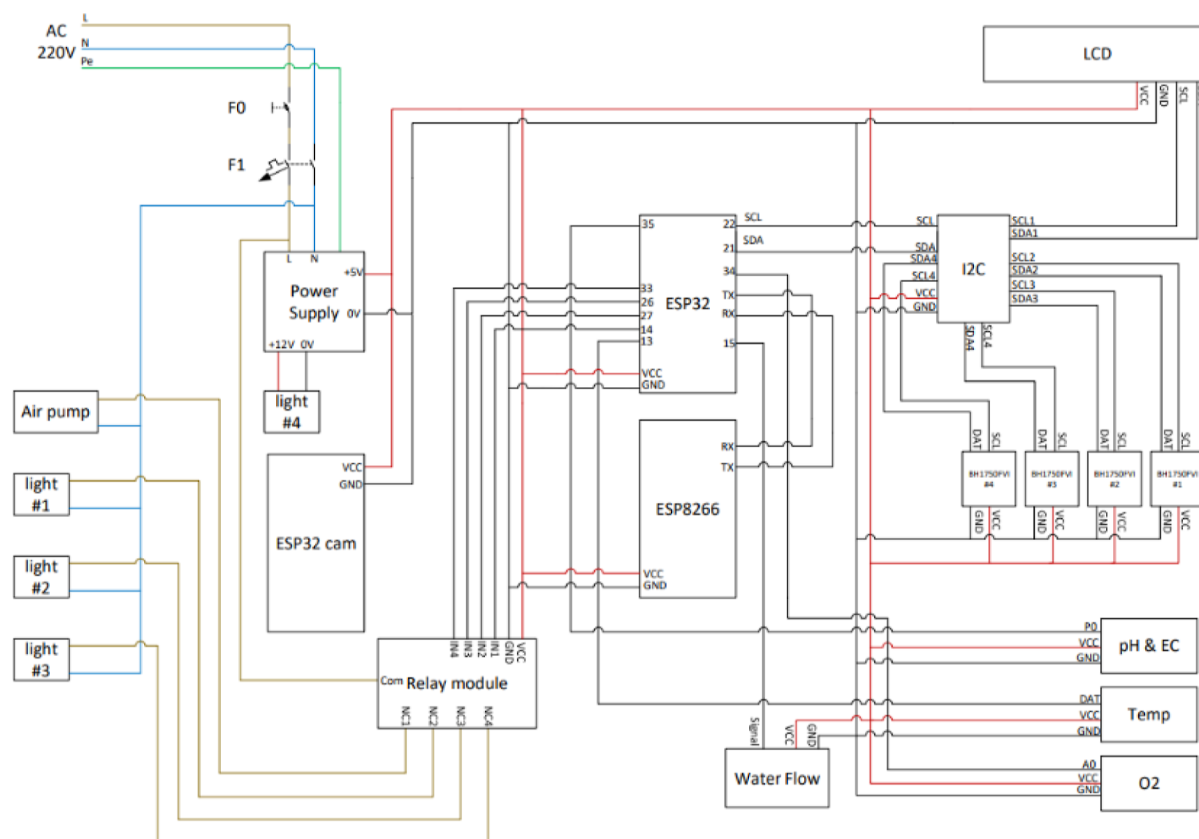
Figure 2. Schematic diagram of the instrumentation and control system for the PBR.

with a comprehensive sensor array for real-time monitoring of critical growth parameters, (2) to develop and validate ML models, specifically RF and XGBoost, for accurate prediction of *Chlorella* sp. growth based on multiparameter inputs, (3) to implement noninvasive image processing techniques for continuous biomass assessment, and (4) to investigate the effects of different light exposure durations on algal growth and productivity. By combining these technologies, it is hypothesized that more precise control over growth conditions can be achieved, leading to improved biomass productivity and quality. This integrated approach has the potential to revolutionize microalgae cultivation by enabling real-time optimization and automation, potentially reducing production costs and improving biomass yield.

2. MATERIALS AND METHODS

2.1. Tubular Photobioreactors Design and Construction. A closed tubular PBR was designed and constructed for the cultivation of *Chlorella* sp. The system was installed and tested indoors to control environmental variables such as temperature and light intensity, as shown in Figure 1. The system comprises the following components:

2.1.1. Structure. The frame of the PBR was constructed from robust square steel tubes, which were cut to the required lengths and welded together to form a rectangular structure measuring 60 cm (width) × 100 cm (length) × 140 cm (height). The frame was coated with black anticorrosion paint to prevent rust and deterioration from environmental conditions.

2.1.2. Tubes. Transparent UPVC tubes had a diameter of 6 cm and a length of 100 cm. Six tubes of this specification were used, along with five additional tubes 6 cm in diameter. The
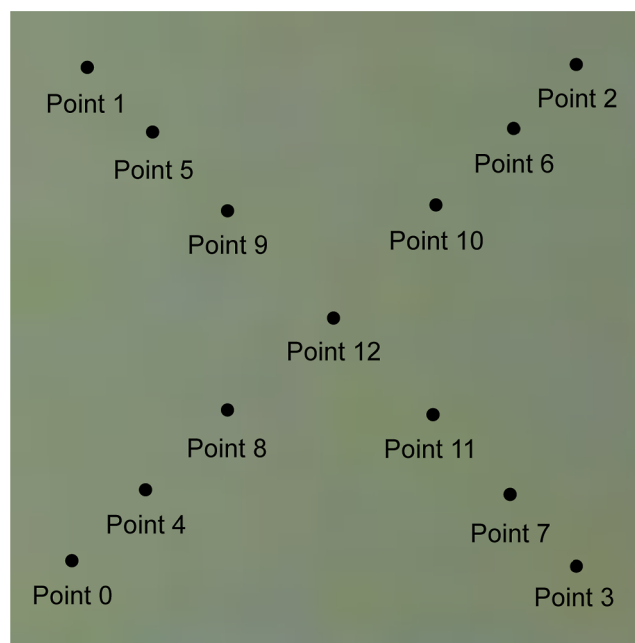

Figure 3. Distribution of 13 measurement points for color intensity analysis on an image of algae culture.

tubes were arranged vertically in a serpentine configuration. They were connected using PVC elbows and 90-degree joints to create a continuous circuit. The tube ends were connected to mixing tanks at the top and bottom of the PBR.

2.1.3. Tanks. An acrylic tank enclosed within an opaque black cabinet was installed at the top of the PBR frame and
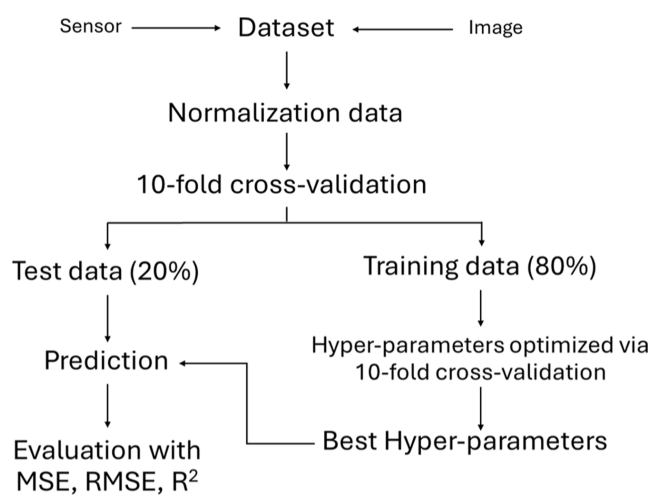
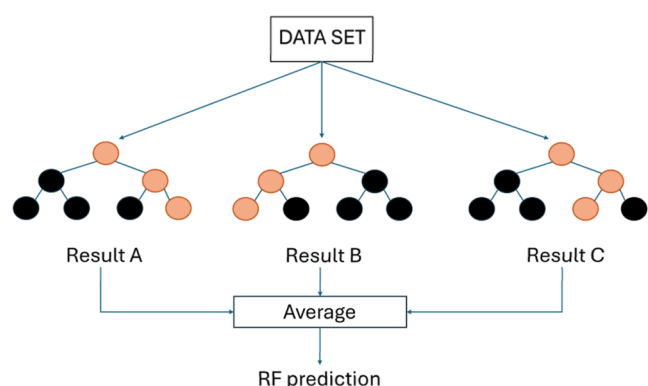**Figure 4.** ML model development and evaluation process.

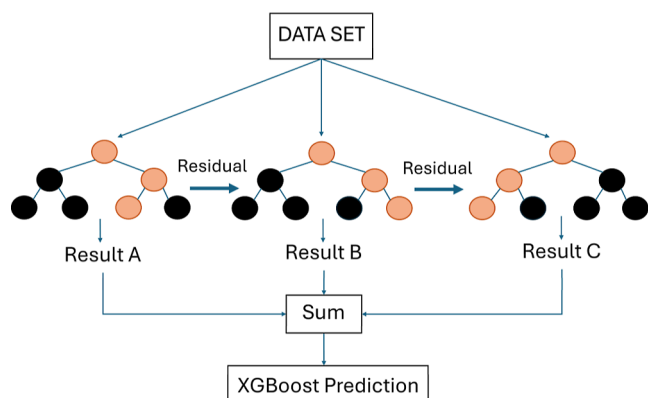**Figure 5.** Schematic representation of the RF algorithm.

Figure 6. Schematic representation of the XGBoost algorithm.

served as a reservoir for mixing and sampling the algal culture. The upper tank had a removable transparent lid to allow access for cleaning and maintenance. The tank was connected to the PVC tubes using 6 cm diameter PVC ball valves.

2.1.4. Aeration System. A stainless-steel sparger (2 in. in diameter, 1.5 m in length) was used for air injection into the algal culture. The sparger was positioned at the bottom of the PBR, beneath the lower tank, and was connected to an air pump using a flexible silicone tube.

2.1.5. Lighting System. The PBR was illuminated by 6 waterproof LED lamps, each measuring 12.3 cm (length) × 3.4 cm (width) × 4.5 cm (height). The LED lamps were mounted

on the PBR frame, parallel to the PVC tubes, with an average distance of 10 cm between the lamps and the tubes. Each LED lamp had a power output of 18 W. For experimental purposes, light exposure was compared between continuous 24 h lighting and a 12 h light cycle (from 6 AM to 6 PM).

The PBR was designed using 3D design software, which facilitated the creation of a three-dimensional model of the reactor and allowed for the simulation of its dimensions, flow patterns, and structural integrity. The flow behavior within the PBR was analyzed using computational fluid dynamics (CFD) simulations. The simulations demonstrated that the PBR design provides uniform flow distribution and effective mixing of the algal culture. The PBR components were fabricated from locally sourced materials and assembled on-site. The steel frame was welded and painted in-house, while the PVC tubes and acrylic tank were cut and drilled according to design specifications. The LED lamps, air pump, and other accessories were purchased from commercial suppliers and installed on the PBR. Following assembly, the PBR underwent leak testing by filling it with water and operating the air pump for 24 h. Any detected leaks were sealed using silicone sealant.

2.2. Instrumentation and Control System. The tubular PBR system incorporates various sensors to monitor and control several parameters crucial for optimal algae growth. These sensors provide real-time data on the conditions within the PBR, enabling the system to make informed decisions and adjustments to maintain an ideal environment for algae cultivation.

2.2.1. Temperature Sensor (DS18B20). The DS18B20 detects temperatures ranging from −55 to +125 °C with an accuracy of ±0.5 °C, thereby facilitating optimal thermal conditions for algal metabolism and photosynthetic efficiency.

2.2.2. pH Sensor (E-201-C PH Sensor). The E-201-C PH sensor measures pH values from 0 to 14 with an accuracy of ±0.1 pH, essential for regulating acidity or alkalinity for nutrition absorption and inhibiting toxin buildup.

2.2.3. Light Intensity Sensor (MCU-GUVA-S12SD Sunlight UV Intensity Sensor). The MCU-GUVA-S12SD quantifies UV light intensity between 200 and 370 nm, facilitating meticulous regulation of light exposure to enhance photosynthesis and avert UV-induced harm to algal cells.

2.2.4. Electrical Conductivity Sensor (DFRobot Gravity: Analog Electrical Conductivity Sensor). This sensor quantifies EC, offering information about nutrient concentration and overall water quality, which is crucial for optimizing conditions for algal development.

2.2.5. Dissolved Oxygen Sensor (SEN0237-A). The SEN0237-A measures oxygen levels in the water, essential for regulating photosynthesis and respiration in algal cultures.

2.2.6. Water Flow Rate Sensor (YF-DN50). The YF-DN50 facilitates enough circulation within the PBR, ensuring regular dispersion of nutrients and gases, essential for consistent algae growth.

The integration of sensors and the ESP8266 microcontroller plays a crucial role in the automated tubular PBR system, enabling real-time monitoring, control, and optimization of the algae cultivation environment. The sensors are strategically connected to the microcontroller via general-purpose input/output pins, facilitating seamless data acquisition and processing, as shown in Figure 2.

Upon receiving the sensor data, the ESP8266 microcontroller processes it using control algorithms and programmed logic to automatically adjust the system's compo-

nents, such as water pumps, air pumps, and LED lights, via relay modules. This ensures that the PBR maintains optimal conditions for algae growth, maximizing biomass productivity. Moreover, the microcontroller communicates and transmits the acquired data to a cloud platform, such as Google Sheets, via wireless Internet connectivity. This enables remote real-time monitoring and analysis of the system's performance, facilitating convenient system management and optimization from afar.

2.2.7. Sensor Calibration and Validation. The accuracy and reliability of sensor measurements are critical for the effective monitoring and control of the cultivation conditions in the PBR. The sensors used in the system were carefully calibrated and validated against standard instruments to ensure data quality. The acceptable value can be calculated from eq 1

$$\text{percentage error} = (V_{\text{cal}} - V_{\text{sensor}}) \times 100/V_{\text{cal}} \tag{1}$$

where V_{cal} is the value measured by the standard instrument used for calibration, and V_{sensor} is the value measured by the sensor being calibrated.

2.3. Image Processing. This study utilized image processing techniques to monitor and analyze the growth of *Chlorella* sp. algae in the tubular PBR system. The integration of image processing allows for the noninvasive, real-time assessment of algae growth, providing valuable insights into the system's performance and enabling timely adjustments to optimize growth conditions. An ESP32 camera module was installed inside the PBR to capture images of the algae culture every 30 min. The camera was placed in a light-proof box to ensure a clear view of the culture medium, minimizing any obstructions or reflections that could interfere with image quality.

The captured images, each 33×33 pixels in size, were transmitted to a connected computer for further processing and analysis. The image processing pipeline was implemented using the OpenCV library in Python. The captured images underwent a series of preprocessing steps to enhance their quality and extract relevant features. These steps included grayscale conversion, image segmentation, and feature extraction.

Each image was measured for color intensity at 13 points distributed across the entire image, as shown in Figure 3. The obtained color intensities were averaged to estimate algae density more accurately than single-point or whole-image analysis, as it accounts for potential spatial variations in algae distribution within the PBR. The RGB images were converted to grayscale so as to simplify processing and reduce computational complexity. The grayscale values at these 13 points were then analyzed to determine the green color intensity.

The RGB scale ranges from 0 to 255, where lower values indicate higher opacity, corresponding to greater biomass density. This color intensity analysis serves as a proxy for algae biomass density. Lower grayscale values (darker regions) indicate higher algae concentration, while higher values (lighter regions) suggest lower concentration. The relationship between the grayscale intensity and algae biomass concentration is inverse: Lower grayscale values indicate higher algae concentration or significant growth, while higher grayscale values suggest lower algae concentration or minimal growth.

2.4. Machine Learning Algorithms. ML algorithms play a crucial role in predicting and optimizing algae growth based on the collected sensor data. The selection of appropriate algorithms is essential to ensure accurate predictions and efficient system control. In this study, two prominent ML algorithms, namely RF and XGBoost, were employed to analyze the relationship between various environmental factors and algae growth.

The ML framework was implemented using Python's Scikit-learn library to determine the optimal algorithm for our specific feature space. This evaluation was necessary because each data set possesses a unique structure, and no single algorithm is universally optimal for all scenarios.[28,29]

Figure 4 illustrates the ML data process following data set standardization and scaling. To ensure an unbiased performance evaluation of the prediction models, *k*-fold cross-validation was employed, with $k = 10$ in this study. This method involved randomly shuffling the entire data set and dividing it into *k* nonoverlapping partitions (folds). The candidate ML model was then trained using $k - 1$ folds as training data and tested on the remaining fold (test data). This process was repeated *k* times, each time using a different fold as the test set, ensuring that all data were utilized for both training and testing purposes.

During the training phase, the hyper-parameters for each model were optimized using a comprehensive grid search approach. This involved exploring all possible combinations of hyper-parameters in conjunction with a nested *k*-fold cross-validation method. This thorough optimization process allowed us to fine-tune each model's performance for our specific data set and prediction task.[29]

To evaluate the performance of these ML models, several metrics were employed. Mean squared error (MSE) and root mean squared error (RMSE) are commonly used to measure the average squared difference between the predicted and actual values. Lower MSE and RMSE values indicate better model performance, as they quantify the magnitude of prediction errors.[18]

Loss functions measure the difference between the predicted values and the actual values. In this study, two common loss functions were used: MSE and RMSE. MSE is defined by eq 2

$$\text{MSE} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 \tag{2}$$

where *n* is the number of samples, y_i is the actual value, and $\widehat{y_i}$ is the predicted value [18].

RMSE is the square root of MSE shown by eq 3

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y_i})^2} \tag{3}$$

RMSE is a popular metric because it is in the same unit as the target variable, making it easy to interpret.[18]

The coefficient of determination (*R*-squared, R^2) is another important metric that measures the proportion of variance in the dependent variable (algae growth) that can be explained by the independent variables (environmental factors). An R^2 value closer to 1 indicates a better fit of the model to the data, implying that the model captures a significant portion of the variability in algae growth.[18] It is defined by eq 4

$$R^2 = 1 - \left(\frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}}\right) \tag{4}$$
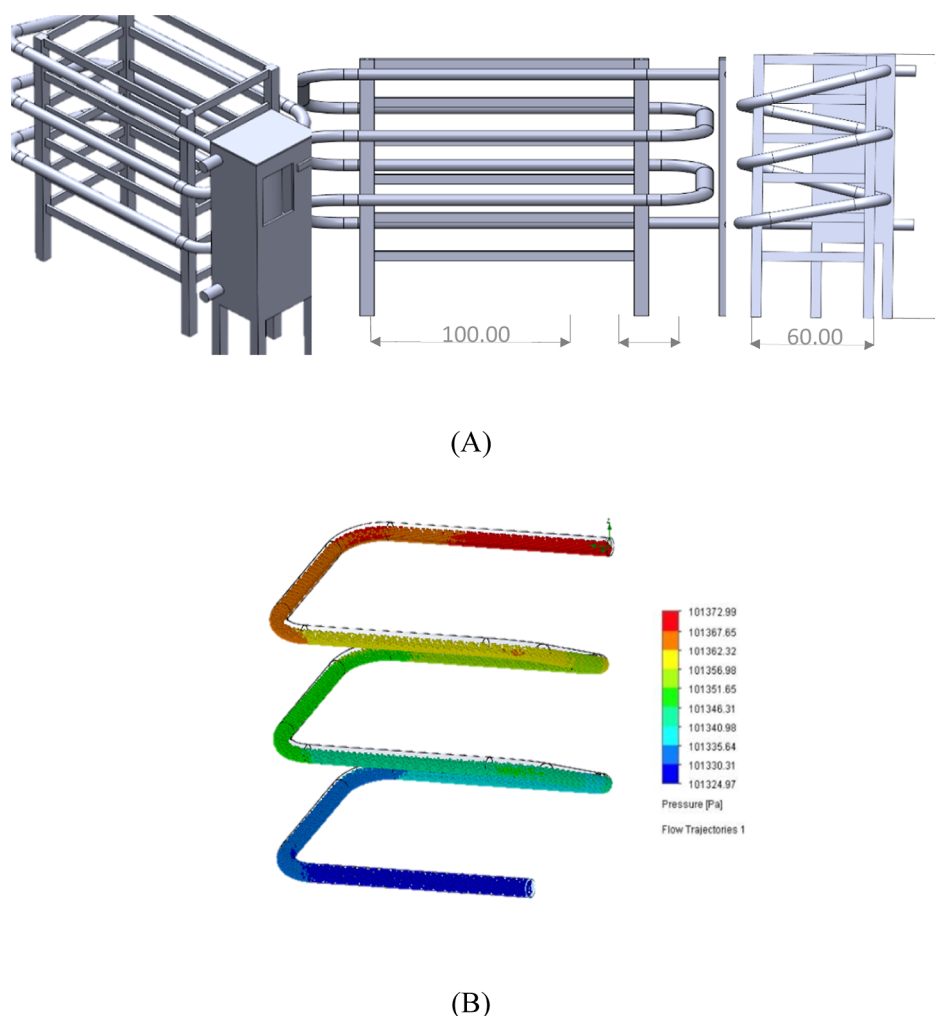
(A)

(B)

Figure 7. Tubular PBR design and CFD analysis showing (A) 3D model with dimensions (cm), and (B) pressure distribution.
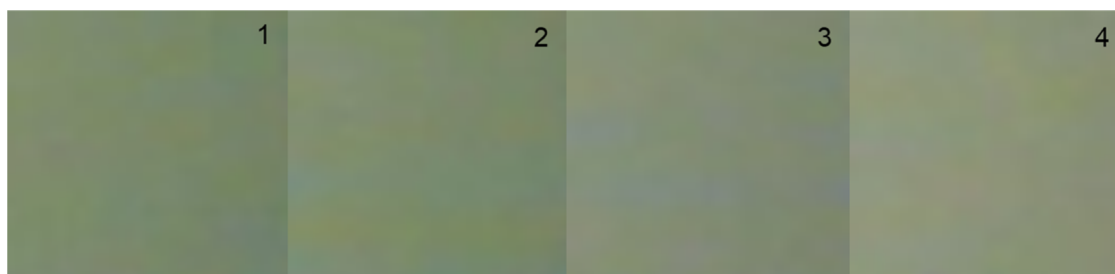

Figure 8. Example of image color intensity analysis for microalgae growth monitoring.

where SS_{res} is the sum of squares of residuals and SS_{tot} is the total sum of squares by eq 5.

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \right) \tag{5}$$

where $\overline{y}$ is the mean of the actual values.

2.4.1. Random Forest. RF is an ensemble learning method that operates by constructing multiple decision trees during the training phase, as shown in Figure 5. It uses bagging (bootstrap aggregating) to create diverse subsets of the training data for each tree and employs random feature selection at each split to ensure tree diversity. For regression tasks, RF outputs the mean prediction of the individual trees. This approach helps to reduce overfitting and enables RF to handle high-dimensional data effectively. The algorithm's strength lies in its ability to capture complex interactions between features without requiring extensive tuning of the hyperparameters, making it robust across various applications, including biological system modeling.[30]

2.4.2. eXtreme Gradient Boosting. XGBoost is an optimized implementation of gradient boosting that has gained popularity due to its performance and efficiency, as shown in Figure 6. Unlike RF, XGBoost constructs trees sequentially, with each new tree aiming to correct the errors of the previously combined ensemble. It uses a more regularized model formalization to control overfitting, often resulting in superior performance. XGBoost employs advanced techniques

such as gradient-based one-side sampling and feature bundling to enhance both speed and accuracy. The ability of the algorithm to handle missing data, its built-in cross-validation capability, and its effectiveness in capturing nonlinear relationships make it particularly suitable for complex prediction tasks in areas like algae growth modeling in PBR.[31]

3. RESULTS AND DISCUSSION

3.1. Computational Fluid Dynamics Analysis and Design Optimization of a Tubular Photobioreactors for

Table 1. Calibration Results of the Sensors

sensors	parameter	mean error (%)
algal culture temperature	°C	1.466
pH	pH	0.549
EC	$\mu S/cm$	1.615
light intensity	$\mu mol/m^2/s$	1.020

Algal Cultivation. Beneficial insights into the flow characteristics and mixing efficacy of the system were gained using CFD simulations performed with SolidWorks software during the design phase of the PBR. The 3D model and results of the CFD simulation for the PBR are given in Figure 7.

The tubular design of the PBRs, with significant dimensions that were improved for efficient algal cultivation, are illustrated in the 3D model shown in Figure 7A, while Figure 7B provides the results of the CFD simulation, demonstrating the pressure distribution in the PBRs. Uniform flow distribution was observed within the PBR, as revealed by the CFD simulation results. A steady pressure drop is indicated by the pressure variation along the tubes' length, which encourages turbulent flow and effective mixing. The pressure range represented by the color gradient in the model is 101325−101373 Pa, with a pressure differential of roughly 48 Pa between the highest and lowest places. This pressure differential can guarantee adequate mixing without exposing the algal cells to unnecessary mechanical stress. As shown in Figure 7A, the required flow characteristics are attained by the PBR design, as established by the simulation results, guaranteeing that algae get uniform exposure to nutrients and light system-wide. The optimization of the flow dynamics in the design is demonstrated by the tubular arrangement and pressure distribution shown in Figures 7A and 8B, respectively. To enhance algal growth and productivity, uniform distribution is vital.

How the structural design of the PBR affects the fluid dynamics within the system can be seen by assessing Figure 7A,B. The link between the physical design and fluid behavior is highlighted by the progressive color changes, as shown in Figure 7B, which match the twisting course of the tubes in Figure 7A. The development of a PBR that compromises between effective mixing and low stress on algae cells was

Table 3. Descriptive Statistics of Input Features and Output Variable for *Chlorella* sp. Cultivation in the PBR

features	min	max	mean	standard deviation
algal culture temperature (°C)	23.81	30.12	27.01	1.73
EC ($\mu S/cm$)	410.32	479.97	444.73	28.12
pH	6.66	8.64	7.98	0.58
light intensity ($\mu mol/m^2/s$)	251.67	430.00	299.33	48.33
flow rate (l/h)	0.00	16.00	3.50	3.99
oxygen concentration (%)	69.00	100.00	80.96	6.03
light duration (h)	12.00	24.00	18.00	6.00
green pixel intensity	133.31	149.15	141.23	4.32

Table 4. Optimal Hyperparameters for RF and XGBoost Models

parameters	XGBoost		RF	
	boundary	value	boundary	value
eta	0.05−0.5	0.4	-	-
max_depth	4−10	6	20−80	80
subsample	0.5−1	1	-	-
colsample_bytree	0.5−1	1	-	-
n_estimators	10−1000	120	10−1000	60
min_samples_split	-	-	2−10	2

made possible by the combination of design and CFD analysis, as shown in Figure 7.

3.2. Sensor Calibration and Validation. In order to confirm accurate data collection, a thorough calibration process was utilized for all sensors in the PBR. Encompassing the anticipated measurement range, with an acceptable error limit of ±3% for all sensors, the process employed standard solutions and conditions.

All sensors attained mean errors within acceptable limits, as shown by the results of calibration. As given in Table 1, the mean error of the temperature sensor was 1.466%, while the pH sensor was 0.549%, EC sensor was 1.615%, and light intensity sensor was 1.020%. The sensors were accurately calibrated and could provide accurate measurements for the functions of monitoring and control, as established by the results.

3.3. Growth Monitoring Using Image Analysis. The image processing approach used for evaluating microalgae color intensity is shown in Figure 8, which displays four sample images in order of green pixel intensity from lowest to highest. For images 1 to 4, the average green intensity values were 139.23, 140.08, 145.38, and 145.54, respectively.

The 13-point averaged green pixel intensity and algae growth shows an inverse relationship, signifying:

Table 2. Comparison of Algae Color Intensity over a 5 Day Span

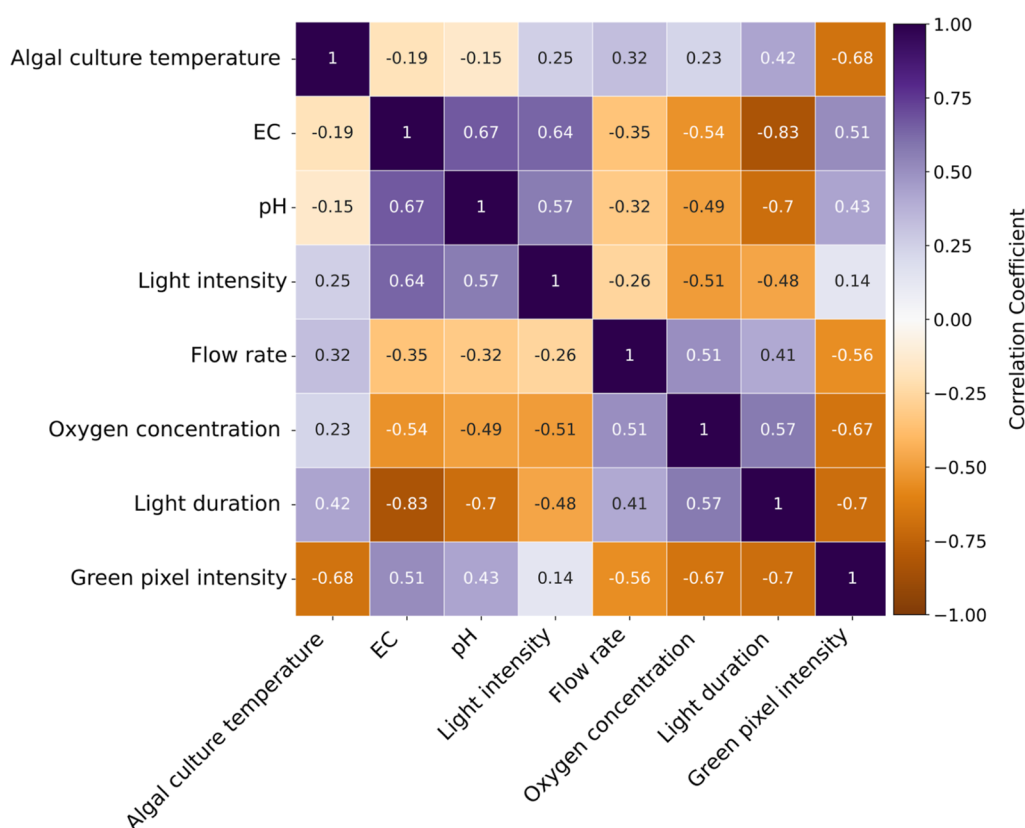| Day	Color image	Average green pixel intensity (scale: 0-255)
1		146.24
6		136.23

Figure 9. Correlation matrix of physicochemical parameters in microalgal cultivation.

- low green pixel intensity (images on the left in Figure 8) implies high algae concentration or considerable growth.
- High green pixel intensity (images on the right in Figure 8) suggests low algae concentration or slight growth.

The results of image analysis for *Chlorella* sp. growth over 5 days are given in Table 2. Showing a reduction of 4.79%, the average green pixel intensity reduced from 146.24 to 136.23 units (scale: 0−255). Indicating enhanced photosynthetic efficiency, likely because of optimal light distribution in the PBR, this reduction in green pixel intensity implies an increase in biomass and chlorophyll content inside *Chlorella* sp. cells. The consistent color shift across the picture indicates efficient mass transfer and mixing throughout the system, guaranteeing uniform nutrient delivery. Important for advancing adaptive control strategies to preserve ideal growth conditions for *Chlorella* sp., this image analysis approach delivers noninvasive, instantaneous quantitative data. Additionally, precise growth rates and the effectiveness of the culture system can be determined using this information. The research shows how image processing can be included with conventional cultivation monitoring techniques to improve the general effectiveness of microalgae production systems.

3.4. Predictive Performance of ML Models. *3.4.1. Data Set and Descriptive Statistics.* Comprising 301 samples for a 12 h light cycle and 301 samples for a 24 h light cycle, the data set utilized 602 data points, equally dispersed between two light exposure durations, for training and testing the ML models. The data points contained seven input features comprising algal culture temperature, EC, pH, light intensity, flow rate, oxygen concentration, and light duration. Performing as a substitute for algal growth, the singular output variable was the green pixel intensity of *Chlorella* sp. The descriptive statistics for this data set, explaining the range, central tendency, and dispersion of each feature across the experimental conditions, are given in Table 3.

3.4.2. Hyperparameter Tuning. Significant variances in the optimal learning structures for each algorithm (Table 4) were observed in the hyperparameter tuning for RF and XGBoost models. XGBoost showed peak performance with a moderate learning rate (eta = 0.4) and limited tree depth (max_depth = 6), while using all data and features in each boosting round (subsample = colsample_bytree = 1) with a higher number of estimators (n_estimators = 120). Conversely, RF used much deeper trees (max_depth = 80) but with fewer estimators (n_estimators = 60), only applying the square root of the total number of features for each node split (max_features = sqrt).

Alongside XGBoost aiming to build a larger number of less complex decision trees, which helps to reduce overfitting and enhance predictive capability, the differences indicate the separate learning mechanisms of each algorithm. Conversely, RF uses feature randomization to preserve diversity among trees while depending on deeper trees to depict complex data relationships. This divergent tuning emphasizes the need to find the most suitable hyperparameters for each model to attain ideal performance in calculating *Chlorella* sp. growth under diverse cultivation settings.

3.4.3. Correlation Analysis of Physicochemical Parameters. Using Pearson correlation coefficient (PCC) calculations, the correlation matrix shown in the heat map in Figure 9 explains the complex relationships between various physicochemical parameters in the algal cultivation system. Significant links between light intensity with pH, and EC (r = 0.57 and 0.64, respectively), all of which establish strong positive relationships, are indicated by the wide-ranging analysis. The
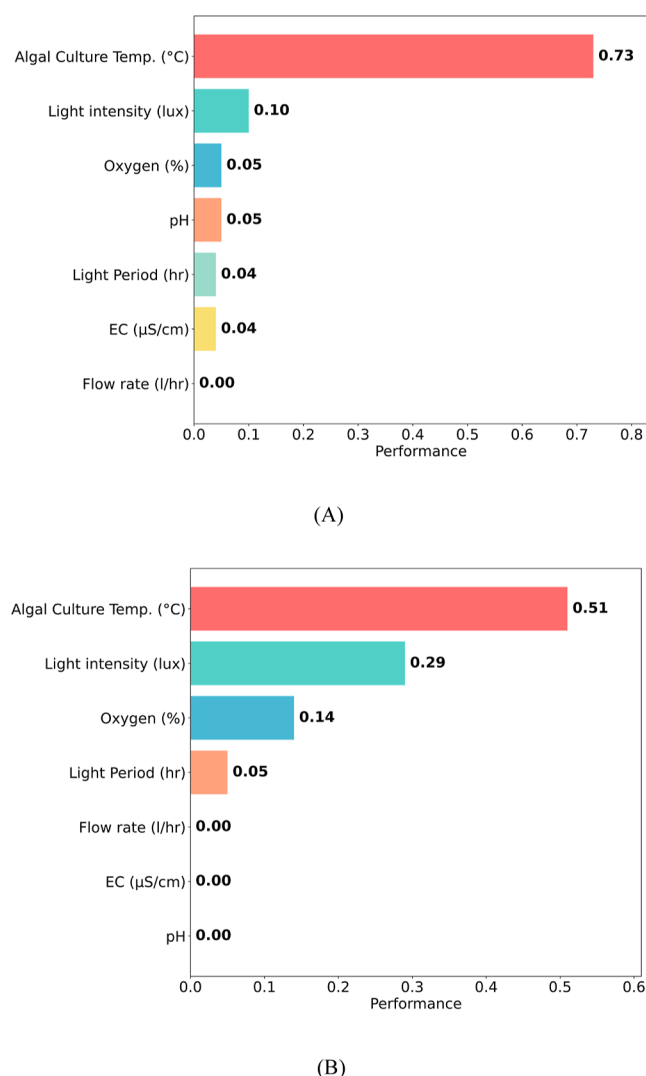
(A)

(B)

Figure 10. Feature significance using (A) RF and (B) XGBoost.

significant interplay of these factors in the algal growth environment emphasize these strong associations.

The duration of light reveals strong negative links with pH and EC ($r = -0.7$ and -0.83, respectively), showing considerable chemical shifts in the cultivation medium through extended exposure to light. This outcome emphasizes the dynamic nature of the algal culture environment and why meticulous light regime adjustment is necessary.

Inverse relationships with most other parameters, especially pH and EC ($r = -0.15$ and -0.19, respectively), are shown by the culture temperature of algae. This inverse relationship highlights the significance of temperature changes on the chemical balance of the system and, subsequently, the physiology of algae.

Exhibiting strong positive links with flow rate ($r = 0.51$), oxygen concentration develops as a key factor. This result

emphasizes the importance of having enough water circulation to encourage the production of oxygen by the system, which is essential for the most beneficial algal photosynthesis.

The green pixel intensity, possibly suggestive of algal biomass or chlorophyll content, shows positive associations with light intensity, pH, and EC ($r = 0.14$, 0.43, and 0.51, respectively). On the other hand, it shows a potent negative link with the culture temperature of algae ($r = -0.68$). These connections imply that though higher temperatures could have a negative impact on algal development, increasing light and specific chemical conditions may enhance the growth of algae.

3.4.4. Feature Importance. Algal culture temperature is the most significant factor in both models, with importance values of 0.73 for RF (Figure 10a) and 0.51 for XGBoost (Figure 10b), as shown by the feature importance evaluation for *Chlorella* sp. growth using RF and XGBoost (Figure 10). Though its weights varied greatly (0.10 for RF and 0.29 for XGBoost), light intensity is the second most critical component in both scenarios. The two models exhibit differing levels of relevance for other factors such as oxygen content, pH, light period, and EC, with RF assigning greater weight to these variables than XGBoost. In both cases, flow rate always appears as the least significant element. In feature ranking, the disparities between the models are an indication of their different learning strategies and abilities to identify intricate links in the data.

3.4.5. Model Comparison. Various evaluation metrics, including MSE, RMSE, and R^2, were used to evaluate the operation of the RF and XGBoost models, as shown by the results in Table 5.

Both models showed high R^2 values when assessed on the training and testing data sets, implying compelling predictive abilities. Attaining 0.9997 and 0.9680 for the training and test sets, respectively, XGBoost slightly surpassed RF in R^2 metrics evaluated against RF's 0.9969 and 0.9652, suggesting XGBoost can capture data variance with minor superiority.

Mild performance variances were shown by the evaluation of error metrics. XGBoost revealed lower error rates (MSE: 0.0063, RMSE: 0.0792) compared to RF (MSE: 0.0604, RMSE: 0.2458) in the training set, suggesting tighter fitting of data. On the other hand, the results of the test set exhibited a more intricate pattern. RF displayed a slightly lower RMSE (0.5806 vs 0.7313), while XGBoost sustained a lower MSE (0.5348 vs 0.7619), suggesting that RF possibly presents more consistent predictions across different test samples, though XGBoost may have average squared errors that are lower.

An illustration of the performance of the model, weighing expected and actual values for the training and test data sets, is shown in Figure 11. Scatter plots for the XGBoost and RF models both demonstrate a strong linear relationship between the actual and projected values, supporting the high R^2 values that were identified. Suitable predictive accuracy for both models is shown by the tight clustering of points down the diagonal line in both plots, though slight variances can be seen. Specifically in the test data set (blue points), there is a slightly

Table 5. Performance Comparison of RF and XGBoost Models

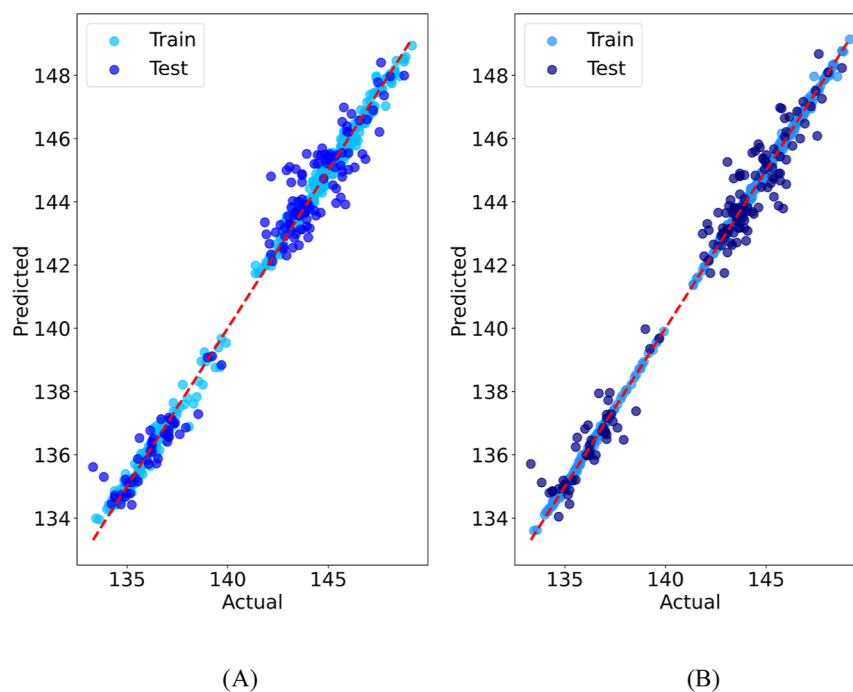| models	R^2		MSE		RMSE	
	train	test	train	test	train	test
RF	0.9969	0.9652	0.0604	0.7619	0.2458	0.5806
XGBoost	0.9997	0.9680	0.0063	0.5348	0.0792	0.7313

Figure 11. Comparison of predicted vs actual values for train and test data sets using (A) RF and (B) XGBoost.

wider spread of points shown in Figure 11A, aligning with its slightly lower R^2 value. There is a stronger clustering of points, particularly in the training set (light blue points), in accordance with its higher R^2 and lower MSE values, as shown in Figure 11B.

As shown by the performance gap between training and test sets, the assessment of model overview disclosed that both models show some amount of overfitting. Remarkably, the RF model presented a minor relative surge in RMSE from training to test evaluated against XGBoost, implying possibly better generalization to unseen data. As shown by the spread of test points (dark blue) seeming more consistent between RF and XGBoost, even with the firmer training set allocation for XGBoost, this is perceptively revealed in Figure 11.

3.5. Comparison of Light Exposure Durations. Two distinctive light exposure durations, comprising continuous illumination (24 h) and a 12 h light, were used to examine the progression of *Chlorella* sp. in the tube PBR.

3.5.1. Temporal Dynamics of Green Pixel Intensity. Using 24 and 12 h light conditions, as illustrated in Figure 12A, this study contrasted the growth of *Chlorella* sp. using green pixel intensity evaluation of images captured for 7 days, at 30 min intervals, as a gauge of growth; declining values (approaching 0) for green intensity suggest enhanced biomass and chlorophyll content. First, photoacclimation[32] and rapid cell division likely caused both light conditions to exhibit rapid exponential growth, especially in the initial 2 days. Contrasting to the 12 h cycle, the cultures that received 24 h light showed a steadier exponential drop in green intensity as growth developed, suggesting persistent biomass accumulation. Especially after day 5, growth reduced the cultures attained a stationary phase after the exponential phase. Continuous 24 h light promoted a higher long-term biomass increase, as shown by the logarithmic trend lines, proving exponential biomass growth and chlorophyll accumulation can benefit from continuous exposure to light. When conditions such as light are abundant, the exponential growth phase is typical in

cultures of microalgae, in line with the Beer−Lambert law utilized in related models.[33]

3.5.2. Daily Average Green Pixel Intensity and Biomass Accumulation. Regarding further analysis, a comparison of the daily average green pixel intensity between 12 and 24 h light cycles over the 7 day experimental period is shown in Figure 12B. Further perceptions into the growth dynamics of *Chlorella* sp. under different light conditions are offered by the bar chart.

The average green intensity values for the 12 h cycle at 146.44 and the 24 h cycle at 145.05 on the first day of the experiment constituted similar starting points for both cultures. Conversely, a more noticeable variance began to appear from day 2 onward. Green intensity values decreased steadily and more quickly during the 24 h light cycle, falling from 144.18 on day 2 to 134.62 on day 7, suggesting a quicker rise in biomass and/or chlorophyll content. Establishing continuous growth under continuous illumination, this reduction was steady throughout the period of the experiment. On the other hand, the 12 h light cycle showed a gradual and more stable reduction in green intensity values, dropping from 145.54 on day 2 to 143.38 on day 7. The rate was more gradual compared to the 24 h cycle, perhaps representing slower growth but possibly higher chlorophyll accumulation per cell, even though there was a reduction in green intensity. Reinforcing earlier analyses that continuous illumination can cause higher biomass accumulation, the disparity between the green intensity values of the two conditions expanded with time; the 24 h cycle exhibited considerably lower values than the 12 h cycle by the last day of the test (134.62 vs 143.38).

3.5.3. Investigation of Changes in Green Pixel Intensity and Biomass Accumulation Dynamics. As shown by variations in green pixel intensity, which reduces as biomass increases, the experimental results show a considerable distinction in biomass accumulation between the 12 h and 24 h light exposure cycles over 7 days.

The green pixel intensity fell from 146.44 to 143.38 in the 12 h light cycle, implying a 2.09% boost in biomass. Conversely,
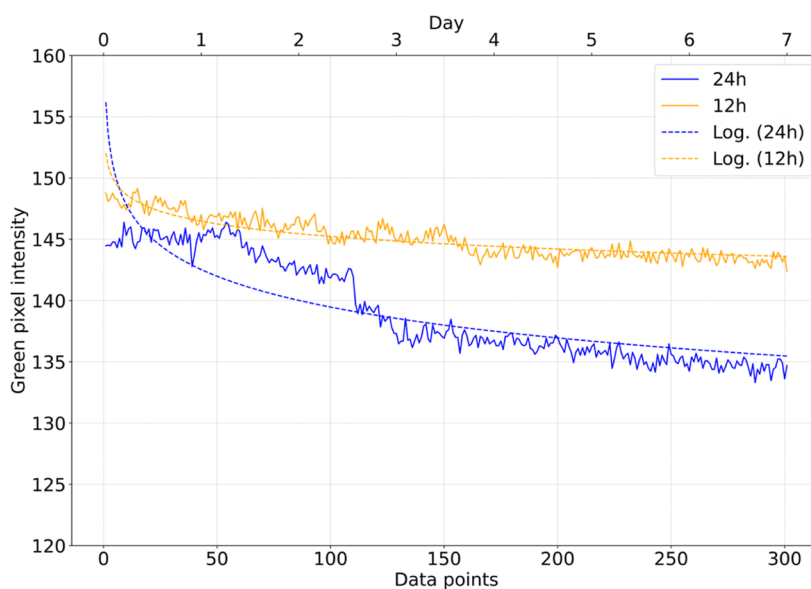
(A)
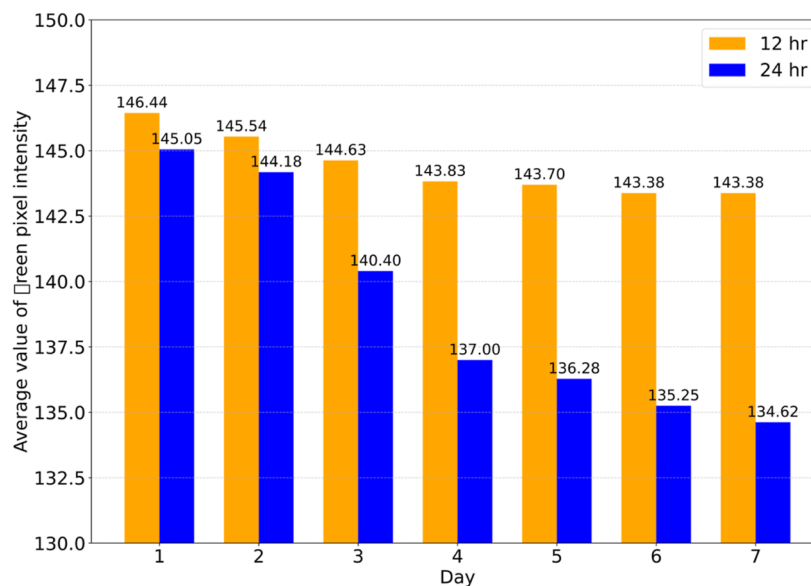
(B)

Figure 12. (A) Green pixel intensity over time under 12 and 24 h light cycles; (B) Daily average green pixel intensity comparison between 12 and 24 h cycles.

Table 6. Comparison of Green Pixel Intensity Changes under Different Light Exposure Durations over 7 days

light exposure duration (h)	initial green pixel intensity	final green pixel intensity	percentage change
12	146.44	143.38	2.09%
24	145.05	134.62	7.19%

the 24 h light cycle revealed a more extensive surge in biomass. Green pixel intensity reduced from 145.05 to 134.62, signifying a 7.19% increase in biomass.

Continuous 24 h illumination can cause considerably higher biomass accumulation when evaluated against the 12 h light cycle, as suggested by the results, possibly due to the ability of

algae to uninterruptedly undertake photosynthesis. Nonetheless, other factors such as biomass quality, energy efficacy, and production goals should be taken into account when selecting the duration for light exposure. Regarding the conservation of energy and preserving the long-term biological balance of algae, the 12 h cycle appears to present benefits (Table 6).

4. CONCLUSIONS

Using the combination of cutting-edge technologies, comprising a CFD-designed tubular PBR, IoT sensors, AI-driven control, and noninvasive image processing, a state-of-the-art strategy resulted in the effective establishment of an intelligent cultivation system for *Chlorella* sp. microalgae. The primary

control board for data management and operational functions of the system used an ESP8266, while simultaneous observation and analysis of green intensity to assess algal density was done with an ESP32 camera, specifying an increase in biomass and chlorophyll concentration. While observation showed a 2.09% rise in the 12 h cycle, the results revealed that continuous lighting for 24 h boosted green pixel intensity by 7.19%. In keeping with the photosynthetic and metabolic processes of algae, feature importance analysis showed that temperature, light intensity, and oxygen content were the most substantial elements affecting growth. Showing an R^2 value of 0.9997, the XGBoost model exhibited remarkable effectiveness in forecasting algae proliferation. Owing to its Gradient Boosting methodology, improved overfitting mitigation approaches, and ability to manage complex, nonlinear data, making it particularly suitable for biological systems, XGBoost outperformed RF.

A wide-ranging assessment of the link between biomass production and green intensity, including an analysis of the chemical composition of biomass at various growth stages to establish more accurate monitoring and management systems should be the focus of future research. Moreover, creating AI models that can adapt to variable environmental conditions and exploring the integration of renewable energy sources to improve sustainability is required.

■ AUTHOR INFORMATION

Corresponding Author

Somboon Sukpancharoen − *Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand;* ⬤ orcid.org/0000-0002-6974-8661; Phone: +66(043)009700 ext. 50638; Email: sombsuk@kku.ac.th

Authors

Thananop Tummawai − *Department of Mechanical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand*

Thongchai Rohitatisha Srinophakun − *Department of Chemical Engineering, Faculty of Engineering, Kasetsart University, Bangkok 10900, Thailand*

Surapol Padungthon − *Department of Environmental Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.4c05971

Author Contributions

Thananop Tummawai: Writing − original draft, Resources, Methodology, Formal analysis, Data curation, Investigation, Validation. **Thongchai Rohitatisha Srinophakun**: Writing − review & editing, Software, Methodology, Supervision. **Surapol Padungthon**: Writing − review & editing, Formal analysis. **Somboon Sukpancharoen**: Writing − original draft, Writing − review & editing, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Project administration, Validation, Visualization.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Scott, S. A.; Davey, M. P.; Dennis, J. S.; Horst, I.; Howe, C. J.; Lea-Smith, D. J.; Smith, A. G. Biodiesel from Algae: Challenges and Prospects. *Curr. Opin. Biotechnol.* **2010**, *21* (3), 277−286.

(2) Spolaore, P.; Joannis-Cassan, C.; Duran, E.; Isambert, A. Commercial Applications of Microalgae. *J. Biosci. Bioeng.* **2006**, *101* (2), 87−96.

(3) Ruiz, J.; Olivieri, G.; de Vree, J.; Bosma, R.; Willems, P.; Reith, J. H.; Eppink, M. H. M.; Kleingris, D. M. M.; Wijffels, R. H.; Barbosa, M. J. Towards Industrial Products from Microalgae. *Energy Environ. Sci.* **2016**, *9* (10), 3036−3043.

(4) Grobbelaar, J. U. Microalgal Biomass Production: Challenges and Realities. *Photosynth. Res.* **2010**, *106* (1−2), 135−144.

(5) Mondal, P. P.; Galodha, A.; Verma, V. K.; Singh, V.; Show, P. L.; Awasthi, M. K.; Lall, B.; Anees, S.; Pollmann, K.; Jain, R. Review on Machine Learning-Based Bioprocess Optimization, Monitoring, and Control Systems. *Bioresour. Technol.* **2023**, *370*, 128523.

(6) Ravindran, B.; Kurade, M. B.; Kabra, A. N.; Jeon, B.-H.; Gupta, S. K. Recent Advances and Future Prospects of Microalgal Lipid Biotechnology. In *Algal Biofuels: Recent Advances and Future Prospects*; Springer: Cham, 2017; pp 1−37.

(7) Cao, J.; Yuan, H.; Li, B.; Yang, J. Significance Evaluation of the Effects of Environmental Factors on the Lipid Accumulation of *Chlorella* Minutissima UTEX 2341 under Low-Nutrition Heterotrophic Condition. *Bioresour. Technol.* **2014**, *152*, 177−184.

(8) Hulatt, C. J.; Thomas, D. N. Productivity, Carbon Dioxide Uptake and Net Energy Return of Microalgal Bubble Column Photobioreactors. *Bioresour. Technol.* **2011**, *102* (10), 5775−5787.

(9) Arnon, D. I. Copper Enzymes in Isolated Chloroplasts. Polyphenoloxidase in Beta Vulgaris. *Plant Physiol.* **1949**, *24* (1), 1−15.

(10) Narala, R. R.; Garg, S.; Sharma, K. K.; Thomas-Hall, S. R.; Deme, M.; Li, Y.; Schenk, P. M. Comparison of Microalgae Cultivation in Photobioreactor, Open Raceway Pond, and a Two-Stage Hybrid System. *Front. Energy Res.* **2016**, *4*, 29.

(11) Mazzelli, A.; Cicci, A.; Franceschini, G.; Di Caprio, F.; Iaquaniello, G.; Altimari, P.; Toro, L.; Pagnanelli, F. Investigation of Effects of Nutrients and External Parameters on Kinetic Growth of Outdoor Microalgal Cultivation. *Chem. Eng. Prog.* **2018**, *64*, 691−696.

(12) Lam, M. K.; Lee, K. T.; Mohamed, A. R. Current Status and Challenges on Microalgae-Based Carbon Capture. *Int. J. Greenhouse Gas Control* **2012**, *10*, 456−469.

(13) Skjånes, K.; Rebours, C.; Lindblad, P. Potential for Green Microalgae to Produce Hydrogen, Pharmaceuticals, and Other High-Value Products in a Combined Process. *Crit. Rev. Biotechnol.* **2013**, *33* (2), 172−215.

(14) Wijffels, R. H.; Barbosa, M. J. An Outlook on Microalgal Biofuels. *Science* **2010**, *329* (5993), 796−799.

(15) Brennan, L.; Owende, P. Biofuels from Microalgae—A Review of Technologies for Production, Processing, and Extractions of Biofuels and Co-Products. *Renewable Sustainable Energy Rev.* **2010**, *14* (2), 557−577.

(16) Sukpancharoen, S.; Katongtung, T.; Rattanachoung, N.; Tippayawong, N. Unlocking the potential of transesterification catalysts for biodiesel production through machine learning approach. *Bioresour. Technol.* **2023**, *378*, 128961.

(17) Sukpancharoen, S.; Wijakmatee, T.; Katongtung, T.; Ponhan, K.; Rattanachoung, N.; Khojitmate, S. Data-Driven Prediction of Electrospun Nanofiber Diameter Using Machine Learning: A Comprehensive Study and Web-Based Tool Development. *Results Eng.* **2024**, *24*, 102826.

(18) Tuntiwongwat, T.; Thammawiset, S.; Srinophakun, T. R.; Ngamcharussrivichai, C.; Sukpancharoen, S. BCLH2Pro: A Novel

Computational Tools Approach for Hydrogen Production Prediction via Machine Learning in Biomass Chemical Looping Processes. *Energy AI* **2024**, *18*, 100414.

(19) Chong, J. W. R.; Khoo, K. S.; Chew, K. W.; Ting, H. Y.; Iwamoto, K.; Ruan, R.; Ma, Z.; Show, P. L. Artificial Intelligence-Driven Microalgae Autotrophic Batch Cultivation: A Comparative Study of Machine and Deep Learning-Based Image Classification Models. *Algal Res.* **2024**, *79*, 103400.

(20) Lakshmi, S.; Sivakumar, R. Chlorella Algae Image Analysis Using Artificial Neural Network and Deep Learning. *Biologically Rationalized Computing Techniques For Image Processing Applications*; Springer Cham, 2018; Vol. *25*, pp 215−248..

(21) Jung, S. K.; Lee, S. B. Image analysis of light distribution in a photobioreactor. *Biotechnol. Bioeng.* **2003**, *84* (3), 394−397.

(22) Chen, X.; Goh, Q. Y.; Tan, W.; Hossain, I.; Chen, W. N.; Lau, R. Lumostatic Strategy for Microalgae Cultivation Utilizing Image Analysis and Chlorophyll a Content as Design Parameters. *Bioresour. Technol.* **2011**, *102* (10), 6005−6012.

(23) Peter, A. P.; Chew, K. W.; Pandey, A.; Lau, S. Y.; Rajendran, S.; Ting, H. Y.; Munawaroh, H. S. H.; Phuong, N. V.; Show, P. L. Artificial Intelligence Model for Monitoring Biomass Growth in Semi-Batch Chlorella Vulgaris Cultivation. *Fuel* **2023**, *333*, 126438.

(24) Castaldello, C.; Gubert, A.; Sforza, E.; Facco, P.; Bezzo, F. Microalgae Monitoring in Microscale Photobioreactors via Multi-variate Image Analysis. *ChemEngineering* **2021**, *5* (3), 49.

(25) Concepcion, R.; Saavedra, M. J.; Alejandrino, J.; Palconit, M. G. Chlorella Vulgaris Surface-Mount Photobioreactor with Vision-Based Growth Signature Prediction Optimized by Electromagnetism-Like Mechanism. *Int. J. Recent Technol. Eng.* **2020**, *4* (4), 378.

(26) Teng, S. Y.; Yew, G. Y.; Sukačová, K.; Show, P. L.; Máša, V.; Chang, J. S. Microalgae with Artificial Intelligence: A Digitalized Perspective on Genetics, Systems and Products. *Biotechnol. Adv.* **2020**, *44*, 107631.

(27) Mowbray, M.; Savage, T.; Wu, C.; Song, Z.; Cho, B. A.; Del Rio-Chanona, E. A.; Zhang, D. Machine Learning for Biochemical Engineering: A Review. *Biochem. Eng. J.* **2021**, *172*, 108054.

(28) Williams, T.; McCullough, K.; Lauterbach, J. A. Enabling Catalyst Discovery through Machine Learning and High-Throughput Experimentation. *Chem. Mater.* **2020**, *32* (1), 157−165.

(29) Onsree, T.; Tippayawong, N. Machine Learning Application to Predict Yields of Solid Products from Biomass Torrefaction. *Renewable Energy* **2021**, *167*, 425−432.

(30) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5−32.

(31) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, 2016, pp 785−794.

(32) Van Vooren, G.; Le Grand, F.; Legrand, J.; Cuiné, S.; Peltier, G.; Pruvost, J. Investigation of Fatty Acids Accumulation in *Nannochloropsis oculata* for Biodiesel Application. *Bioresour. Technol.* **2012**, *124*, 421−432.

(33) Huesemann, M. H.; Van Wagenen, J.; Miller, T.; Chavis, A.; Hobbs, S.; Crowe, B. A Screening Model to Predict Microalgae Biomass Growth in Photobioreactors and Raceway Ponds. *Biotechnol. Bioeng.* **2013**, *110* (6), 1583−1594.