https://doi.org/10.31436/iiumej.v27i3.4333

Regression-Based Optical Intensity and Image Processing Techniques for Diatom Algae Content Estimation

RAHMAT MAGHROBI ALWAFI[1], AGUS INDRA GUNAWAN[1*], SETIAWARDHANA[2]

[1]*Department of Electrical Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia*
[2]*Department of Informatics and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia*

**Corresponding author: agus_ig@pens.ac.id*

(Received: 13 April 2026; Accepted: 13 August 2026; Published online: 11 September 2026)

ABSTRACT: Microalgae are responsible for oxygen production, pH stabilization, and nitrogen waste absorption in shrimp pond ecosystems, serving as essential biological indicators of water quality. Within the unicellular microalgae group, diatoms such as *Skeletonema* sp. and *Thalassiosira* sp. confer major ecological benefits, excessive growth can disturb pond environments. Despite their significance, monitoring these diatoms, particularly in Indonesian aquaculture systems, remains constrained by time-consuming manual counting. This study proposes a non-invasive integrated framework combining direct raw optical intensity (OI) sensing and digital image processing (IP) techniques to estimate the concentration of *Skeletonema* sp. and *Thalassiosira* sp. across varying laboratory-grown concentration ranges. The proposed regression models demonstrated high predictive performance across all concentration ranges, with test set coefficients of determination exceeding 0.95 ($R^2 > 0.95$). The results indicate that biomass estimations derived from image processing are more accurate than those obtained from direct optical intensity, with data suggesting that species-specific optical properties are modulated by cellular pigment responses. Overall, this non-destructive RGB regression framework offers reliable microalgae biomass estimation for precision aquaculture.

ABSTRAK: Mikroalga bertanggungjawab menghasilkan oksigen, menstabilkan pH, dan menyerap sisa nitrogen dalam ekosistem kolam udang, berfungsi sebagai penunjuk biologi penting kualiti air. Dalam kumpulan mikroalga bersel tunggal, diatom seperti *Skeletonema* sp. dan *Thalassiosira* sp. memberikan manfaat ekologi besar, namun pertumbuhan berlebihan boleh mengganggu persekitaran kolam. Walaupun kepentingannya, pemantauan diatom ini, terutamanya dalam sistem akuakultur Indonesia, masih terhad oleh pengiraan manual yang memakan masa. Kajian ini mencadangkan rangka kerja terintegrasi tidak invasif yang menggabungkan pengesanan intensiti optik mentah secara langsung (OI) dan teknik pemprosesan imej digital (IP) untuk menganggarkan kepekatan Skeletonema sp. dan Thalassiosira sp. merentasi julat kepekatan yang ditanam di makmal. Model regresi yang dicadangkan menunjukkan prestasi peramalan yang tinggi merentasi semua julat kepekatan, dengan pekali determinan set ujian melebihi 0.95 ($R^2 > 0.95$). Keputusan menunjukkan bahawa anggaran biomassa yang diperoleh daripada pemprosesan imej adalah lebih tepat berbanding yang diperoleh daripada intensiti optik langsung, dengan data mencadangkan bahawa sifat optik spesies tertentu dimodulasi oleh tindak balas pigmen selular. Secara keseluruhannya, rangka kerja regresi RGB bukan merosakkan ini menawarkan anggaran biomassa mikroalga yang boleh dipercayai untuk akuakultur tepat.

KEYWORDS: *Microalgae, RGB, Optical Intensity, Image Processing, Regression*

https://doi.org/10.31436/iiumej.v27i3.4333

1. INTRODUCTION

Shrimp aquaculture is an important part of global aquaculture production. However, maintaining water quality at the level needed for long-term productivity is difficult [1,2]. Water quality parameters, such as dissolved oxygen (DO), pH, temperature, salinity, and nitrogen content, must be maintained within specific ranges to prevent stress and mortality in cultured shrimp [3,4]. Microalgae are important for managing the aquaculture ecosystem because they produce oxygen through photosynthesis, absorb CO_2 to help stabilize pH, and take up nitrogenous compounds from shrimp waste and feed leftovers [5]. The visual characteristics of pond water color indicate microalgal diversity and abundance; green coloration indicates healthy phytoplankton dominance, brown suggests diatom prevalence, and variations in color intensity reflect changes in microalgal biomass concentration, which directly correlates with stability in water quality parameters [6,7,8]. However, uncontrolled microalgal growth can result in harmful blooms that destabilize water quality parameters, particularly when dissolved oxygen levels decrease and pH values fluctuate significantly [9].

That situation requires access to accurate, rapid, and non-invasive methods for measuring microalgal concentration. Currently applied approaches primarily combine optical density (OD) measurements and digital image processing (IP). Traditionally, OD measurements quantify cell density using absorbance at particular wavelengths and standard logarithmic calculations [10]. While these conventional studies establish the core physical principle that light attenuation correlates with biomass concentration, implementing standard logarithmic OD transformations in real-time embedded systems limits direct, efficient implementation on constrained edge devices due to the overhead of floating-point operations. To address this limitation, recent paradigms in multi-sensing monitoring systems suggest that raw transmittance light intensity can serve as a direct, lightweight proxy for biomass, improving resource efficiency on embedded platforms [11]. Concurrently, digital image processing utilizes feature extraction such as average RGB values, color intensity, and pixel distribution to examine the characteristics of visual culture and detect complex, species-specific pigment variations [12]. Therefore, this study adapts the fundamental light-attenuation principle by using a direct, raw RGB optical intensity approach alongside digital image processing to achieve highly accurate, computationally optimized biomass estimation.

Previous studies have demonstrated the effectiveness of the optical method for microalgae monitoring. Cvjetinovic et al. (2023) developed specialized optical systems that use LED light at 505 nm to focus on changes in the concentration of different diatom cultures, achieving coefficient of determination values up to 0.999 with 17.67 mm^3 sensor volumes [13]. In a related study, Castaldello et al. (2021) used an image-based method focused on chlorophyll-a, the main pigment involved in photosynthesis associated with microalgal biomes containing diatoms, where the biomass prediction model achieved an average relative error of about 18%, comparable to laboratory measurements [14]. However, research specifically focused on Skeletonema sp. and Thalassiosira sp. primarily depends on manual counting using hemocytometers, which poses a considerable limitation in aquaculture settings, especially in Indonesia [15]. Although advanced narrow-band optical and multivariate image regression techniques offer high precision, their reliance on specialized hardware or intensive computation limits real-time deployment on low-cost embedded platforms, thereby restricting their application as automated alternatives to manual hemocytometer counting for *Skeletonema* sp. and *Thalassiosira* sp. in Indonesian aquaculture. Therefore, this study aims to develop and validate an integrated, non-invasive methodology that combines direct raw optical intensity readings and image-processing techniques with regression models to accurately and rapidly estimate the biomass concentration of *Skeletonema* sp. and *Thalassiosira* sp.

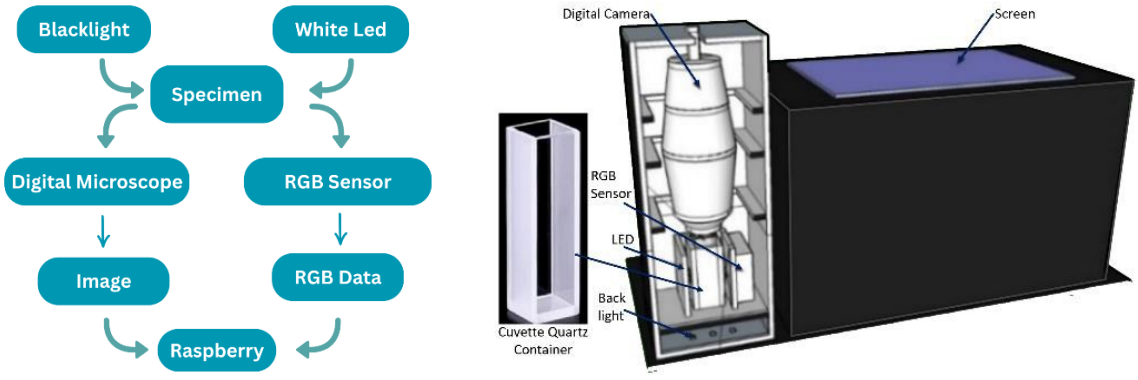
https://doi.org/10.31436/iiumej.v27i3.4333

This study builds upon and expands the methodology introduced in the previous prototype study [16]. While the baseline system utilized optical density to characterize raw sensor outputs, the present work refines this physical framework by formally defining these variables as transmittance-based raw optical intensity features. This terminology adjustment aligns more closely with the system's direct RGB readings, ensuring strict methodological clarity between polychromatic light attenuation and classical logarithmic absorbance metrics.

Beyond this conceptual clarification, this study's principal technical contributions are organized into three main advancements. First, the research focuses exclusively on *Skeletonema* sp. and *Thalassiosira* sp., leveraging their distinct pigment profiles to generate species-specific RGB responses and improve taxonomic specificity over the initial system. Second, expanding the baseline single-modality approach, it evaluates multiple linear regression and second-order polynomial regression models using raw optical data and image-processing features, with concentration-based train–test splits and 5-fold cross-validation. Third, the study provides a comprehensive biological interpretation linking pigment composition to modeling performance. The resulting models reached high predictive accuracy ($R^2 > 0.95$) with minimal errors. The remainder of this paper is organized as follows: Section 2 details the methods and experimental procedures, Section 3 presents the measurement results, and Section 4 offers the study's conclusions.

2. RESEARCH METHODS

This study used a non-destructive measurement method. As shown in Fig. 1(a), the system design describes the measurement workflow, in which a white LED backlight illuminates the specimen. A digital microscope and an RGB sensor recorded data from the specimen simultaneously, producing an image and RGB intensity. The results were then processed using a Raspberry Pi for further analysis. As shown in Fig. 1(b), the prototype consisted of a Class A enclosure equipped with a quartz glass cell unit, a digital microscope, an RGB sensor, and LED lighting integrated into a compact enclosure to minimize external light interference and ensure consistent measurement conditions.



(a) Flow Diagram (b) Prototype

Figure 1. System Design

2.1. Methods and System

To achieve comprehensive and robust biomass estimation, this study employs an integrated dual-modality framework combining direct optical intensity (OI) and image processing (IP). As illustrated in the system architecture in Fig. 1(b), the optical sensing block features a TCS34725 RGB sensor array aligned with a light source. The microalgae sample cuvette was systematically positioned between the light-emitting diode (LED) backlight and

the sensor photodiodes. A white LED was selected to provide a broad, stable polychromatic spectrum, ideal for exciting and analyzing the distinct color and pigment characteristics of the diatom cultures [17]. Short-term exposure during monitoring cycles ensures no impact on microalgae growth [18]. The TCS34725 sensor measured the light intensity incident on the paper in RGB and transmitted the data to the Raspberry Pi via the microcontroller.

Optical measurement relies on polychromatic light attenuation, where microalgae cells absorb and scatter photons, reducing transmitted intensity. In this study, I bypassed classical logarithmic optical density calculations. Because the incident light intensity and the optical path length remained strictly constant, the raw transmitted readings I retained a predictable, monotonic inverse relationship with biomass concentration. Therefore, the direct optical Intensity (OI) framework was operationally formulated as Eq (1):

$$OI(R,G,B) = I(R,G,B) \tag{1}$$

where $I(R,G,B)$ represents the raw digital values extracted directly from the red, green, and blue sensor channels. Bypassing floating-point logarithmic transformations minimizes computational overhead and optimizes real-time processing within the resource-constrained embedded platform. These raw intensity values are referred to as R_{OI}, G_{OI}, and B_{OI} for the red, green, and blue channels, respectively, and represent transmittance-based RGB intensity features rather than classical logarithmic OD values.

IP was also performed by extracting color histograms from digital images of microalgae samples obtained using an integrated camera system [19]. A digital microscope was used as a camera to capture color features rather than for microscopic observation, as shown in Fig. 1(b). The samples were placed between a backlight and a digital microscope. As the measurement space was dark, a white backlight was required to emphasize the intensity captured by the camera from above. To determine the color average parameters for each RGB channel, Eq (2), (3), and (4) were used:

$$R_{ip} = \frac{\sum_{i=1}^{n} R_i}{n} \tag{2}$$

$$G_{ip} = \frac{\sum_{i=1}^{n} G_i}{n} \tag{3}$$

$$B_{ip} = \frac{\sum_{i=1}^{n} B_i}{n} \tag{4}$$

where n is the total number of pixels in the image, and R_i, G_i, and B_i are the intensity values of the i-th pixel in each channel.

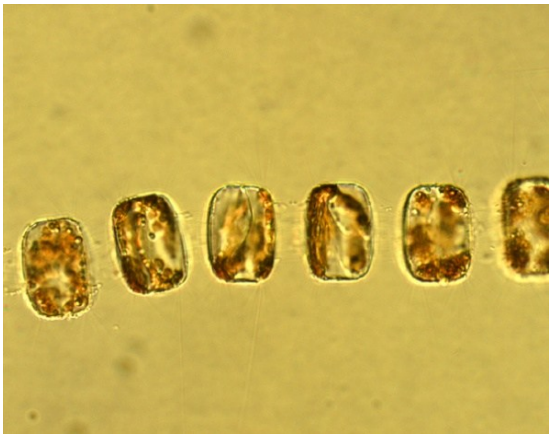
2.2. Microalgae Phycology

Skeletonema sp. is a filamentous diatom that forms long chains of box-shaped cells, as shown in Fig. 2(a) [20]. It does not move by itself, and its color can be yellowish-green or brown because of chlorophyll and carotenoid pigments. The color changes depending on the nutrition, salinity, light intensity, and growth stage. The culture gets darker as biomass increases because cell density and carotenoid levels rise [15]. *Thalassiosira* sp. is a planktonic diatom that comes in cylindrical or box shapes, as shown in Fig. 2(b) [21]. Fish farmers often use it as natural feed. The pigments fucoxanthin and chlorophyll a and c can make it yellow-green or golden brown [22]. This behavior of the three channels suggests that the *Skeletonema* sp. cultures are largely yellowish-green to light brown in color, in agreement with the ever-present, more dominant chlorophyll pigment culture components [23]. *Thalassiosira* sp. and *Skeletonema* sp. have very different amounts of pigments. *Thalassiosira* sp. has more fucoxanthin, which makes it a darker brown color, while *Skeletonema* sp. has more chlorophyll

https://doi.org/10.31436/iiumej.v27i3.4333

c, which makes it a lighter or greenish color [23]. This color difference results from a complex mix of pigments, shapes, and cell densities. This is clear from the color and physical traits of each microalga as it grows [24].

(a) Skeletonema Sp.[20]



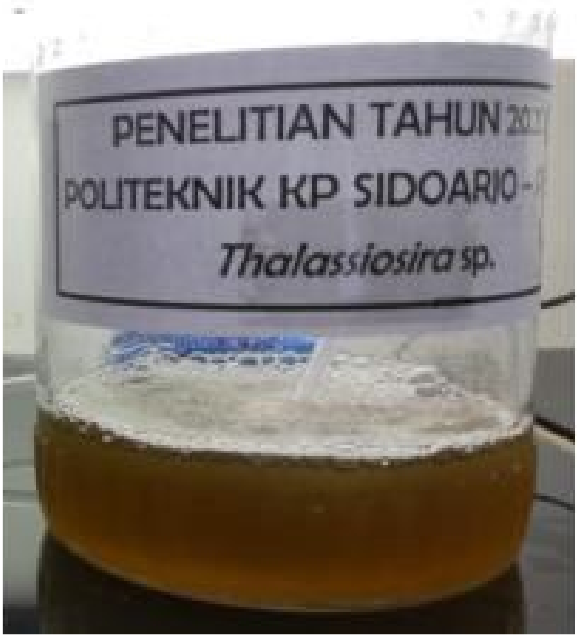
(b) Thalassiosira Sp.[21]

Figure 2. Microalgae

2.3. Sample Preparation

The Department of Fisheries Pathology at Sidoarjo Marine and Fisheries Polytechnic prepared and cultivated samples of *Thalassiosira* Sp. and *Skeletonema* Sp. obtained from the Jepara Brackish Water Aquaculture Research Center (BBPBAP). Fig. 3 shows 43×10^4 cells/mL of Skeletonema sp. and 216×10^4 cells/mL of Thalassiosira sp. For each sample, a 4 mL cuvette was used to measure concentrations ranging from 0% to 100%. To obtain samples at different concentration levels, each microalgae species was diluted with pure water, producing 10% concentration increments for each species. This cuvette measures 12.5 × 12.5 × 45 mm and has two clear glass sides and two non-clear glass sides. The materials and measuring tools were cleaned in an autoclave to kill any germs that could slow down growth.

(a) Skeletonema Sp.



(b)Thalassiosira Sp

Figure 3. Alga Sample

https://doi.org/10.31436/iiumej.v27i3.4333

The samples were placed in cuvette containers and placed in the measuring chamber, as shown in Fig. 1(b). For the OI method, white LEDs and RGB sensors were placed on the left and right sides, facing the clear glass. When the sample was exposed to white light, the RGB sensor on the other side measured the optical intensity response, which revealed the sample's properties. The IP technique is similar, using a digital sensor to capture the reaction and a backlight to illuminate the sample from below.

2.4. Data Collection and Model Validation

For dataset eleven, we prepared concentrations for each microalga from 0% to 100% in 10% increments. At each concentration, we obtained 100 replicate measurements using both the OI and IP methods. We averaged 100 measurements at each concentration to obtain a single representative value for each of the red, green, and blue channels.

We implemented a concentration-based data split to evaluate model performance on previously unseen concentrations. The training set comprised averaged values from nine concentration levels (0%, 20%, 30%, 40%, 60%, 70%, 80%, 90%, and 100%), while the testing set included the remaining two levels (10% and 50%). This partitioning enabled the assessment of the models' interpolation capability at both low (10%) and mid-range (50%) concentrations. Tables 1 and 2 explicitly report the test concentrations, showing actual concentration (AC) versus predicted concentration (PC). All regression modeling and performance evaluation, including coefficient of determination (R^2) and mean absolute error (*MAE*), were conducted using Python's scikit-learn library.

To more rigorously assess model robustness and address potential overfitting concerns, we performed 5-fold cross-validation on the training dataset with nine concentration levels. For each fold, we trained the model on four subsets and evaluated it on the remaining subset. To capture both predictive stability and variance, we report the cross-validation metrics as the mean ± standard deviation (SD) of R2 and MAE across the five validation folds. This approach provides a more reliable estimate of generalization performance than a single train–test split, particularly given the limited number of available concentration levels.

3. Results and Discussion

This section presents the results of measurements of *Skeletonema* sp. and *Thalassiosira* sp. using OI and IP techniques. We estimated biomass concentration using both simple and multiple regression models. We used both methods to examine how including multiple factors affects biomass-estimation accuracy, because conventional regression and multiple linear regression differ in the number of predictors considered. Individual plots illustrate the relationship between each predictor and biomass concentration to improve visualization. We evaluated the performance of each regression method for the OI and IP approaches using R^2 and MAE, where R^2 indicates the model's ability to explain variation in the data, and *MAE* reflects the average prediction error.

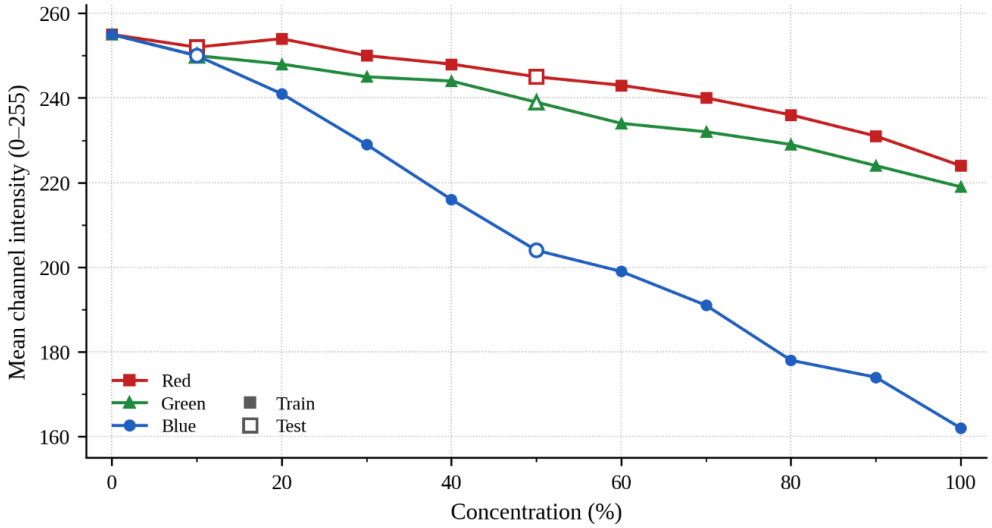
Both the image-based average intensity channels (R_{IP}, G_{IP}, and B_{IP}) and channels derived from optical intensity (R_{OI}, G_{OI}, and B_{OI}) served as predictors in a multiple linear regression (MLR) model formulated as Concentration = $\beta_0 + \beta_1 R + \beta_2 G + \beta_3 B + \varepsilon$. Additionally, a second-order polynomial regression (PR) approach was implemented, in which individual polynomial models were fitted for each feature, and the final concentration was determined by averaging the predictions from all features. A multivariate polynomial regression model was examined, but it did not yield satisfactory results. The model's complexity proved excessive for the dataset's limited size, leading to unstable solutions. Strong correlations among the RGB

https://doi.org/10.31436/iiumej.v27i3.4333

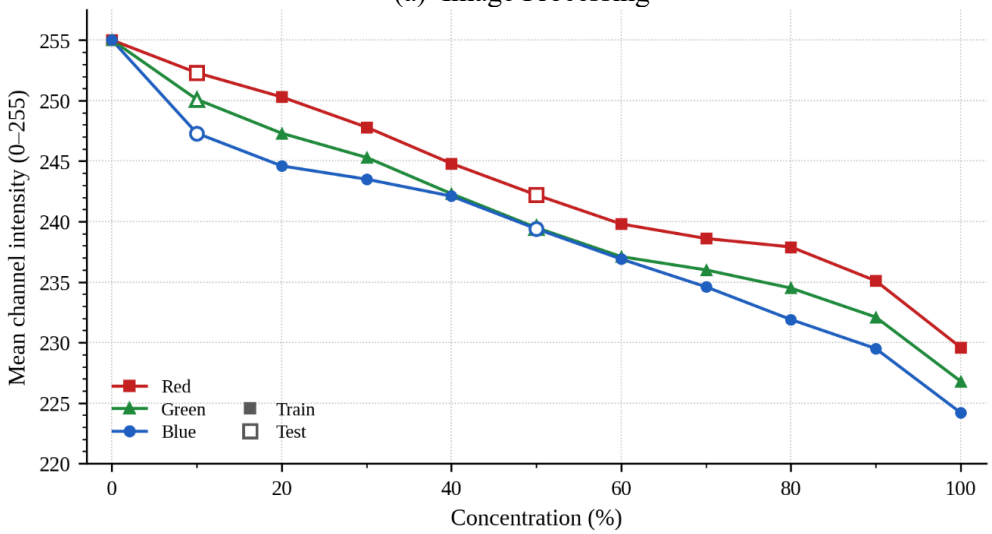
channels led to overfitting, in which the model learned the training data too well but failed to generalize to the test data. As a result, because of complementary patterns in the training data, the multivariate polynomial regression model did not generalize well to unseen data, even compared with individual polynomials and MLR methods.

A custom Python program using the scikit-learn library estimated parameters via least-squares fitting on the training dataset to produce all regression coefficients. The distribution of RGB intensities, changes in OI, and the chosen concentration levels all affect the final coefficients. With the same data and preprocessing circumstances, this formulation allows a direct comparison of linear and nonlinear modeling approaches.

3.1. Skeletonema Sp.



(a) Image Processing



(b) Optical Intensity

Figure 4. Skeletonema Sp. measurement results (a) IP and (b) OI

As shown in Fig. 4, the red and green channels were the dominant color channels during both the IP and OI analyses for *Skeletonema* sp. In general, the blue channel was the most diminished as biomass concentration increased. The image data-based method showed greater separation among the three (red, green, and blue) color channels, indicating that mean RGB values from the image data could potentially be used with greater sensitivity to detect changes

https://doi.org/10.31436/iiumej.v27i3.4333

in the culture's pigment presence. The curves for each OI channel showed a steady, almost linear trend. This behavior across the three channels suggests that Skeletonema sp. cultures are largely yellowish-green to light brown, consistent with the ever-present, more dominant chlorophyll pigment components [23]. The increased coverage of the green color by the fucoxanthin pigment could explain the increase in blue-channel capture. The variation in response between the three channels was much sharper in the IP, implying that this method preserves more color information than the OI method, which provides a more direct optical response to biomass change. Based on the relationships between RGB channel responses and biomass concentration shown in these trends, MLR and second-order PR were acquired for *Skeletonema* sp. for both OI and IP. Eq (6) and (7) provide the descriptive equations for the relationship of the RGB channels to concentration (Y), and Eq (8) to (13) give the individual fit equations for each RGB channel with the average response, which are then used to estimate the overall concentration.

$$Y_{IP} = 473.871827 + 74.817793*R_{IP} - 398.203066*G_{IP} - 150.075396*B_{IP} \quad (5)$$

$$Y_{OI} = 1053.113088 - 1296.950445*R_{OI} + 640.025413*G_{OI} - 396.188056*B_{OI} \quad (6)$$

$$YR_{IP} = -3504.9130 + 8391.1507*R_{IP} - 4882.7901*R_{IP}^2 \quad (7)$$

$$YG_{IP} = -66.0257 + 1004.2964*G_{IP} - 940.5092*G_{IP}^2 \quad (8)$$

$$YB_{IP} = 291.2321 - 320.0015*B_{IP} + 31.6565*B_{IP}^2 \quad (9)$$

$$YR_{OI} = 143.8409 + 875.3660*R_{OI} - 1021.4505*R_{OI}^2 \quad (10)$$

$$YG_{OI} = 2137.1253 - 3427.0015*G_{OI} + 1286.5958*G_{OI}^2 \quad (11)$$

$$YB_{OI} = 1706.5261 - 2609.8157*B_{OI} + 899.0882*B_{OI}^2 \quad (12)$$

Table 1. Skeletonema Sp. Evaluation

Methods	Type	R^2 Train	R^2 Test	AC	PC	MAE	
Multiple Linear Regression	Optical Intensity	0.9892	0.9947	10 50	12.05464 50.21728	2.054638 0.217281	1.136
	Image Processing	0.9992	0.9954	10 50	11.37673 51.32571	1.376728 1.32571	1.3512
2nd Polynomial Regression	Optical Intensity	0.9854	0.9683	10 50	14.86567 51.29772	4.865673 1.297719	3.0817
	Image Processing	0.9993	0.9969	10 50	8.80026 51.02743	1.199742 1.027428	1.1136

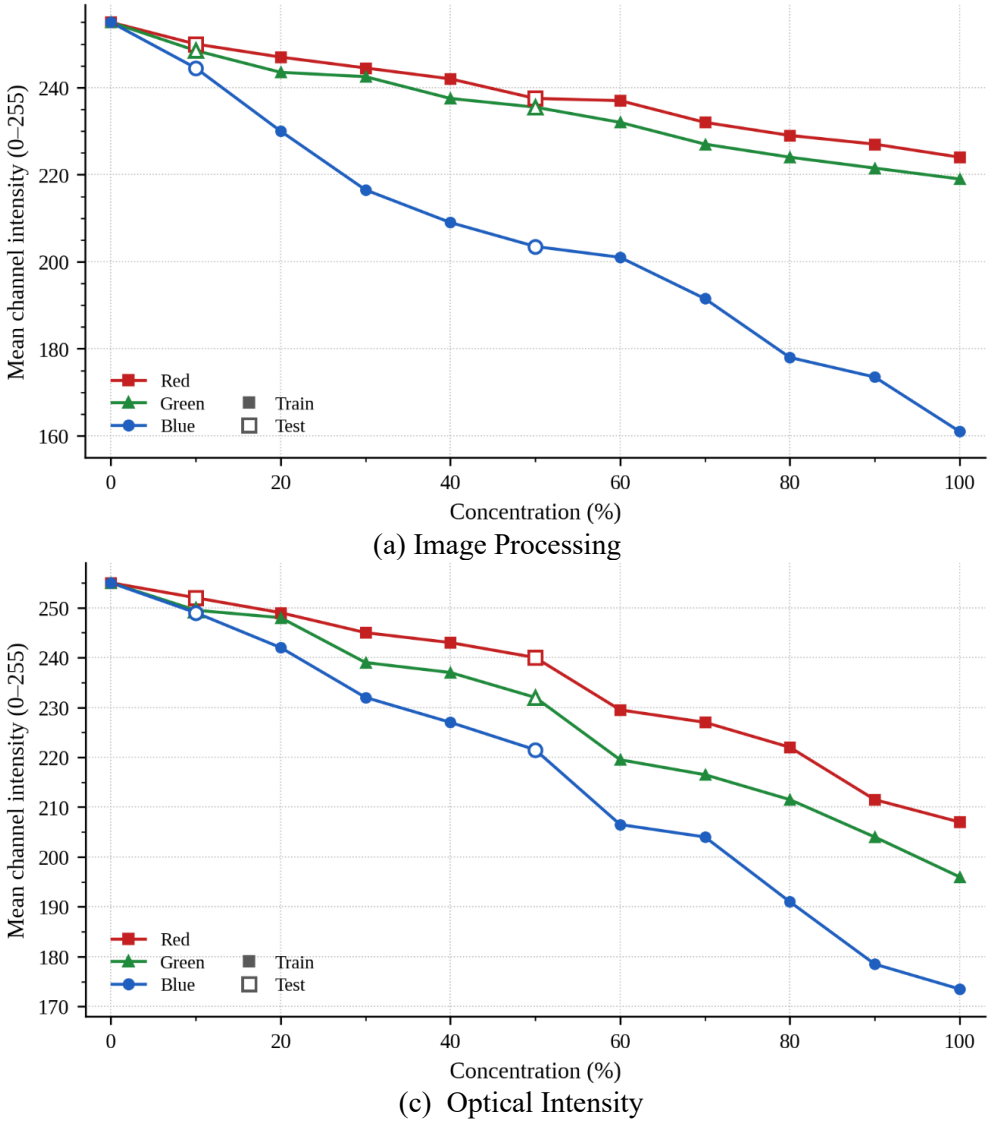
Based on Table 1, IP models outperformed OI models for *Skeletonema* sp. Among all approaches, the second-order PR using IP features delivered the highest predictive accuracy, achieving a test R^2 of 0.9969 and an *MAE* of 1.1136. The MLR model on IP data also performed strongly (test R^2 = 0.9954, *MAE* = 1.3512). OI models were comparatively less accurate, especially the second-order PR variant (*MAE* = 3.0817). These results indicate that RGB features from images better reflect the pigment variations in *Skeletonema* sp. cultures than transmittance-based measurements, and that second-order PR is more suitable for modeling the nonlinear relationships involved. Therefore, the second-order PR model based on IP is recommended as the most reliable method for estimating this species' concentration.

The MLR's strength lies in its ability to use all data in a single regression equation, allowing predictions to be much closer to the actual values, providing consistent performance across the dataset, and reducing the risk of overfitting. Second-order PR is better suited for data

https://doi.org/10.31436/iiumej.v27i3.4333

with complex nonlinear patterns because it can capture the data variation more completely. Using R^2 and *MAE* together provides a more complete analysis of model performance. While R^2 measures the model's overall explanation of variation in the data, MAE assesses individual prediction accuracy.

3.2. Thalassiosira Sp.



(a) Image Processing

(c) Optical Intensity

Figure 5. Thalassiosira Sp. measurement result

For *Thalassiosira* sp., an all-channel decrease with increasing concentration can be observed in Fig. 5(a) for IP and Fig. 5(b) for OI, with the blue channel changing at a faster rate than the red or green channels. This led to greater channel differentiation with increasing biomass, as indicated by the shapes of the curves on the optical intensity plot, providing a strong optical response to concentration. The extracted mean RGBs for the cultures in the IP method showed a similar decreasing behavior, although there was less distinction among the different channels. Compared with Skeletonema sp., Thalassiosira sp. generally exhibited lower RGB values, indicating a darker sample appearance. This behavior results from the specific pigment profile of Thalassiosira sp., where fucoxanthin is the dominant pigment, giving the cultures a rich brown to golden-brown hue [23]. The different responses of the channels indicate that the pigments may influence the color channels unequally, providing a

basis for quantitative concentration modeling. MLR equations, Eq (14) and (15), and second-order PR Eq (16) - (21) were created using OI and IP data to predict the biomass concentration (Y) from the three channel values, with a mean over the channels equaling the estimated concentration:

$$Y_{IP} = 765.643697 - 1047.030808*R_{IP} + 299.449323*G_{IP} - 23.515404*B_{IP} \quad (14)$$

$$Y_{OI} = 265.198939 + 281.705902*R_{OI} - 262.164290*G_{OI} - 283.011439*B_{OI} \quad (15)$$

$$Y_{RIP} = 2160.9818 - 3752.3202*R_{IP} + 1589.9766*R_{IP}^2 \quad (16)$$

$$Y_{GIP} = 2381.4556 - 4317.3729*G_{IP} + 1935.2007*G_{IP}^2 \quad (17)$$

$$Y_{BIP} = 373.5593 - 516.4265*B_{IP} + 140.2109*B_{IP}^2 \quad (18)$$

$$Y_{ROI} = -534.2934 + 1819.5588*R_{OI} - 1282.4392*R_{OI}^2 \quad (19)$$

$$Y_{GOI} = 171.3751 + 161.4931*G_{OI} - 330.2908*G_{OI}^2 \quad (20)$$

$$Y_{BOI} = 166.3622 + 37.1178*B_{OI} - 202.9260*B_{OI}^2 \quad (21)$$

Table 2. Thalassiosira Sp evaluation

Methods	Type	R^2 Train	R^2 Test	AC	PC	MAE	
Multiple Linear Regression	Optical Intensity	0.9949	0.9713	10 50	9.81557 45.21303	0.184433 4.786969	2.4857
	Image Processing	0.989	0.9893	10 50	7.44456 48.5861	2.55544 1.413901	1.9847
2nd Polynomial Regression	Optical Intensity	0.9947	0.9589	10 50	10.14191 44.26954	0.14191 5.730462	2.9362
	Image Processing	0.9936	0.9945	10 50	10.72131 48.02474	0.721315 1.975263	1.3483

Based on Table 2, IP models outperformed OI models for *Thalassiosira* sp. The second-order PR model trained on IP data achieved the best overall performance, with the highest coefficient of determination on the test set (R^2 = 0.9945) and the lowest mean absolute error (*MAE* = 1.3483). The MLR model using IP data also performed very well (test R^2 = 0.9893, *MAE* = 1.9847). In contrast, OI-based models were less accurate, particularly the second-order PR using OI features (*MAE* = 2.9362). These findings suggest that RGB values extracted from images more effectively capture the pigment variations of *Thalassiosira* sp. than transmittance measurements. Furthermore, the second-order PR better captured the nonlinear relationship between image features and biomass concentration than MLR.

To evaluate model durability and generalization more systematically beyond a single train–test split, we conducted 5-fold cross-validation on the training dataset consisting of nine concentration levels, as presented in Table 3. For *Skeletonema* sp., the second-order PR combined with IP features delivered the highest predictive accuracy and stability, achieving a mean R^2 of 0.9932 ± 0.0081 and a mean *MAE* of 1.1738 ± 0.6896. The MLR model trained on IP data also performed well, whereas the OI models showed lower accuracy and higher variance across validation folds. For *Thalassiosira* sp., the second-order PR applied to the OI data performed best, with a mean R^2 of 0.9440 ± 0.0746 and a mean *MAE* of 3.3455 ± 1.1081. The mean R^2 values for the MLR models were similar between the OI and IP approaches, averaging approximately 0.93. However, the OI framework showed greater variability, with an R^2 standard deviation of 0.1038 for the OI model compared to 0.0273 for the IP model. In contrast, the second-order PR on IP data for *Thalassiosira* sp. produced a lower mean R^2

https://doi.org/10.31436/iiumej.v27i3.4333

(0.8891 ± 0.1136) and higher variability, indicating less stable cross-validation performance for this specific configuration.

Table 3. 5-fold cross-validation performance (mean ± standard deviation) for Skeletonema sp. and Thalassiosira sp.

Microalgae	Methods	Type	R^2 (mean ± *SD*)	MAE (mean ± *SD*)
Skeletonema Sp.	Multiple Linear Regression	Optical Intensity	0.9575 ± 0.0208	5.3555 ± 2.0538
		Image Processing	0.9841 ± 0.0167	2.1859 ± 0.6299
	2nd Polynomial Regression	Optical Intensity	0.8529 ± 0.1173	8.0305 ± 4.2258
		Image Processing	0.9932 ± 0.0081	1.1738 ± 0.6896
Thalassiosira Sp.	Multiple Linear Regression	Optical Intensity	0.9283 ± 0.1038	3.3033 ± 1.5831
		Image Processing	0.9309 ± 0.0273	4.6568 ± 2.6311
	2nd Polynomial Regression	Optical Intensity	0.9440 ± 0.0746	3.3455 ± 1.1081
		Image Processing	0.8891 ± 0.1136	4.1481 ± 3.2165

These cross-validation dynamics indicate that optimal modeling depends heavily on the microalgae species and modality. Direct OI favors MLR because of the nearly linear inverse relationship between transmittance and cell concentration under stable illumination. Conversely, IP captures complex color shifts from pigments by aggregating spatial RGB values, making second-order PR more suitable. However, these nonlinearities can induce isolated prediction outliers across validation folds. Because R^2 relies on squared errors, it heavily penalizes these anomalies, leading to a lower mean R^2 with increased variance (e.g., *Thalassiosira's* second-order PR model on IP features at 0.8891 ± 0.1136, whereas *MAE* remains more stable due to its linearity. This localized instability aligns with a profile dominated by fucoxanthin, which induces darker, nonlinear absorption. Conversely, the low SD in *Skeletonema* sp. models confirms excellent structural stability, as this chlorophyll-rich species produces a brighter culture with a more straightforward concentration gradient.

Cross-referencing Table 3 with Tables 1 and 2 confirms a highly consistent performance hierarchy. The top models, namely second-order PR on OI data for *Thalassiosira* sp. (mean R^2 = 0.9440) and second-order PR on IP features for *Skeletonema* sp. (mean R^2 = 0.9932), closely align with the baseline tests. Minor numerical variations between the single-split and k-fold validation are statistically expected, representing the transition from a single random partition to an iterated mean. The narrow SD (±0.0746 and ±0.0081, respectively) proves that these slight fluctuations reflect realistic generalization boundaries rather than model instability. This close convergence demonstrates that integrating raw optical intensities and digital image features provides a highly stable, reproducible biomass estimation framework free from partition-dependent bias.

4. CONCLUSION

This study demonstrates that integrating direct raw OI sensing and digital IP provides a reliable, non-destructive framework for estimating the biomass of *Skeletonema* sp. and *Thalassiosira* sp. Bypassing classical logarithmic transformations successfully eliminates unnecessary computational steps, ensuring that the system is highly optimized for real-time applications. The evaluation confirms high accuracy, where the optimal model selection depends on the microalgae species and the analytical method used. The second-order PR achieved the best overall stability and performance, yielding a mean R^2 of 0.9440 on direct OI data for *Thalassiosira* sp. and 0.9932 on IP features for *Skeletonema* sp. The direct OI method

aligns effectively with MLR because of the near-linear relationship established under stable illumination. Conversely, the IP method aggregates color information across numerous visual coordinates to capture complex pigment variations, making the second-order PR more suitable. The narrow standard deviations across the validation subsets prove that these models are stable and free from partition-dependent bias. Ultimately, this dual-modality framework provides a robust, reproducible technological solution for automated microalgae biomass monitoring in precision aquaculture systems.

ACKNOWLEDGMENT

This research was supported by the Directorate of Academic Affairs for Vocational Higher Education (DAPTV) through the 2025 Master's Thesis Research Grant scheme. The authors also thank the Electrical Engineering for Aquaculture Technology Research Group, Politeknik Elektronika Negeri Surabaya, for providing the necessary laboratory facilities and equipment.

REFERENCES

[1] Yusoff FM, Umi WAD, Ramli NM, Harun R (2024) Water quality management in aquaculture. Cambridge Prism. Water, 2:e8. https://doi.org/10.1017/wat.2024.6

[2] Terkula B, Azman N (2021) Recent advances in shrimp aquaculture wastewater management. Heliyon, 7:e08283. https://doi.org/10.1016/j.heliyon.2021.e08283

[3] Moayedi A, Yargholi B, Pazira E, Babazadeh H (2021) Investigation of bio-desalination potential algae and their effect on water quality. Desalin. Water Treat., 212:78-86. https://doi.org/10.5004/dwt.2021.26638

[4] Yin S, Jin W, Xi T, Zhou X, He Z, Meng X, Naushad M, Jiang G, Li X (2024) Factors affect the oxygen production of Chlorella pyrenoidosa in a bacterial-algal symbiotic system: Light intensity, temperature, pH and static magnetic field. Process Saf. Environ. Prot., 184:492-501. https://doi.org/10.1016/j.psep.2024.02.004

[5] Ende S, Henjes J, Spiller M, Elshobary M, Hanelt D (2024) Recent advances in recirculating aquaculture systems and role of microalgae to close system loop. Bioresour. Technol., 407:131107. https://doi.org/10.1016/j.biortech.2024.131107

[6] Winata HN, Nasution MA, Ahamed T, Noguchi R (2021) Prediction of concentration for microalgae using image analysis. Multimed. Tools Appl., 80(6):8541-8561. https://doi.org/10.1007/s11042-020-10052-y

[7] Yang W, Zheng Z, Lu K, Zheng C, Du Y, Wang J, Zhu J (2020) Manipulating the phytoplankton community has the potential to create a stable bacterioplankton community in a shrimp rearing environment. Aquaculture, 520:734789. https://doi.org/10.1016/j.aquaculture.2019.734789

[8] M (1997) Phytoplankton dynamics in shrimp ponds. Aquac. Res., 28(5):351-360. https://doi.org/10.1046/j.1365-2109.1997.00865.x

[9] Xiao X, Peng Y, Zhang W, Yang X, Zhang Z, Ren B, Zhu G, Zhou S (2024) Current status and prospects of algal bloom early warning technologies: A review. J. Environ. Manage., 349:119510. https://doi.org/10.1016/j.jenvman.2023.119510

[10] Hotos GN, Avramidou D, Bekiari V (2020) Calibration curves of culture density assessed by spectrophotometer for three microalgae (Nephroselmis sp., Amphidinium carterae and Phormidium sp.). Eur. J. Biol. Biotechnol., 1(6):e132. https://doi.org/10.24018/ejbio.2020.1.6.132

[11] Dębowski M, Kazimierowicz J, Zieliński M (2025) Multi-sensing monitoring of the microalgae biomass cultivation systems for biofuels and added value products synthesis—challenges and opportunities. Appl. Sci., 15(13):7324. https://doi.org/10.3390/app15137324

[12] Sunoj S, Hammed A, Igathinathane C, Eshkabilov S, Simsek H (2021) Identification, quantification, and growth profiling of eight different microalgae species using image analysis. Algal Res., 60:102487. https://doi.org/10.1016/j.algal.2021.102487

[13] Cvjetinovic J, Perkov SA, Kurochkin MA, Sergeev IS, German SV, Bedoshvili YD, Davidovich NA, Korsunsky AM, Gorin DA (2023) Concentration dependence of optical transmission and extinction of different diatom cultures. J. Biomed. Photonics Eng., 9(1):010303. https://doi.org/10.18287/JBPE23.09.010303

[14] Castaldello C, Gubert A, Sforza E, Facco P, Bezzo F (2021) Microalgae monitoring in microscale photobioreactors via multivariate image analysis. ChemEngineering, 5(3):49. https://doi.org/10.3390/chemengineering5030049

[15] Badraeni, Trijuno DD, Eriswandi I (2021) Use of bioassay CG (Colour Graduation) to determine density of Skeletonema sp. at hatchery. IOP Conf. Ser. Earth Environ. Sci., 763(1):012039. https://doi.org/10.1088/1755-1315/763/1/012039

[16] Kamaluddin MW, Gunawan AI, Setiawardhana, Dewantara BSB, Insivitawati E, Asmarany A, Pratama AE (2024) Algae content estimation utilizing optical density and image processing method. Int. J. Electr. Comput. Eng., 14(6):6248-6257. https://doi.org/10.11591/ijece.v14i6.pp6248-6257

[17] Kondzior P, Tyniecki D, Butarewicz A (2019) Influence of color temperature of white LED diodes and illumination intensity on the content of photosynthetic pigments in Chlorella vulgaris algae cells. MDPI AG, pp. 46.

[18] Al-amshawee S, Bin Mohd Yunus MY (2020) Influence of light emitting diode (LED) on microalgae. J. Chem. Eng. Ind. Biotechnol., 5(2):9-16. https://doi.org/10.15282/jceib.v5i2.3771

[19] Wood NJ, Baker A, Quinnell RJ, Camargo-Valero MA (2020) A simple and non-destructive method for chlorophyll quantification of Chlamydomonas cultures using digital image analysis. Front. Bioeng. Biotechnol., 8:00746. https://doi.org/10.3389/fbioe.2020.00746

[20] Phycokey - Skeletonema images. Available: https://cfb.unh.edu/phycokey/Choices/Bacillariophyceae/Centric/Centric_Filaments/SKELETONEMA/Skeletonema_Image_page.html

[21] Phycokey - Thalassiosira. Available: https://cfb.unh.edu/phycokey/Choices/Bacillariophyceae/Centric/Centric_Filaments/THALASSIOSIRA/Thalassiosira_key.html

[22] Stenger-Kovács C, Béres VB, Buczkó K, Al-Imari JT, Lázár D, Padisák J, Lengyel E (2023) Review of phenotypic response of diatoms to salinization with biotechnological relevance. Hydrobiologia, 850(20):4665-4688. https://doi.org/10.1007/s10750-023-05194-7

[23] Mishra B, Tiwari A (2021) Spirulina platensis for the production of C-Phycocyanin, C-Phycoerythrin and Thalassiosira, Skeletonema, Chaetoceros for fucoxanthin. Syst. Microbiol. Biomanufacturing. https://doi.org/10.1007/s43393-020-00020-w

[24] Sharma A, Singh P, Srivastava P (2023) Photosynthetic pigments in diatoms. In: Insights into the World of Diatoms: From Essentials to Applications. Srivastava P, Khan AS, Verma J, Dhyani S (Eds.). Springer Nature Singapore, Singapore; pp. 1-20.