Machine learning-based prediction models unleash the enhanced production of fucoxanthin in *Isochrysis galbana*

Janani Manochkumar[1], Annapurna Jonnalagadda[2], Aswani Kumar Cherukuri[3], Brigitte Vannier[4], Dao Janjaroen[5], Rajasekaran Chandrasekaran[1] and Siva Ramamoorthy [ID][1]*

[1]Laboratory of Plant Biotechnology, Department of Biotechnology, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, India, [2]School of Computer Science & Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India, [3]School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India, [4]Cell Communications and Microenvironment of Tumors Laboratory UR 24344, University of Poitiers, Poitiers, France, [5]Department of Environmental and Sustainable Engineering, Faculty of Engineering, Chulalongkom University, Bangkok, Thailand

Introduction: The marine microalga *Isochrysis galbana* is prolific producer of fucoxanthin, which is a xanthophyll carotenoid with substantial global market value boasting extensive applications in the food, nutraceutical, pharmaceutical, and cosmetic industries. This study presented a novel integrated experimental approach coupled with machine learning (ML) models to predict the fucoxanthin content in *I. galbana* by altering the type and concentration of phytohormone supplementation, thus overcoming the multiple methodological limitations of conventional fucoxanthin quantification.

Methods: A novel integrated experimental approach was developed, analyzing the effect of varying phytohormone types and concentrations on fucoxanthin production in *I. galbana*. Morphological analysis was conducted to assess changes in microalgal structure, while growth rate and fucoxanthin yield correlations were explored using statistical analysis and machine learning models. Several ML models were employed to predict fucoxanthin content, with and without hormone descriptors as variables.

Results: The findings revealed that the Random Forest (RF) model was highly significant with a high R^2 of 0.809 and *RMSE* of 0.776 when hormone descriptors were excluded, and the inclusion of hormone descriptors further improved prediction accuracy to R^2 of 0.839, making it a useful tool for predicting the fucoxanthin yield. The model that fitted the experimental data indicated methyl jasmonate (0.2 mg/L) as an effective phytohormone. The combined experimental and ML approach demonstrated rapid, reliable, and cost-efficient prediction of fucoxanthin yield.

Discussion: This study highlights the potential of machine learning models, particularly Random Forest, to optimize parameters influencing microalgal growth and fucoxanthin production. This approach offers a more efficient

alternative to conventional methods, providing valuable insights into improving fucoxanthin production in microalgal cultivation. The findings suggest that leveraging diverse ML models can enhance the predictability and efficiency of fucoxanthin production, making it a promising tool for industrial applications.

1 Introduction

Microalgae are a diverse group of photoautotrophic organisms with a promising source of bioactive compounds, including polysaccharides, fatty acids, carotenoids, phytosterols, and phenols, with beneficial applications (Lourenço-Lopes et al., 2021). Recently, extensive research has contributed to investigating the potential of microalgae to synthesize value-added metabolites owing to their simple cell organization, enhanced accumulation of lipids, rapid life cycle, steady growth rate, non-toxicity, biodegradability, and utilization of CO_2 as a carbon source for growth (Lim et al., 2024). Marine pigments thus have evolved to be an effective alternative in food, therapeutic, and cosmetic applications (Manochkumar et al., 2022). In particular, carotenoids are recommended as dietary supplements, as they possess diverse bioactivities that aid the prevention of chronic diseases including cancer, cardiovascular diseases, diabetes, and age-related macular degeneration (Anantharaman et al., 2014, 2016; Rivera-Madrid et al., 2020).

Fucoxanthin is a xanthophyll marine carotenoid that is abundantly found in the thylakoid membrane of chloroplasts in macroalgae, and its distribution varies within chloroplast in microalgae (Foo et al., 2021; Manochkumar et al., 2022). Among all the carotenoids, fucoxanthin significantly contributes to more than 10% of estimated total carotenoid production globally (Lourenço-Lopes et al., 2021). Currently, the global market value of fucoxanthin upholds an average annual growth rate of 5%, representing an increase from 199.48 million USD in 2022 to 280.7 million USD in 2029 (Market Reports World, 2023). Due to its enormous applications and valuable bioactivities, the cost of purified fucoxanthin ranges from 40,000 to 80,000 USD/kg, depending on the concentration and purity of the compound (Khaw et al., 2022). Despite its promising applications, fucoxanthin remains limited in availability, as it has not yet been fully commercialized. Indeed, synthetic production of fucoxanthin is not feasible due to its complexity; hence, microalgae were explored as an effective and reliable source for fucoxanthin production (Lourenço-Lopes et al., 2021). It is evident that fucoxanthin plays a significant role in microalgae by absorption of photons, thus regulating photosynthesis and aiding photoprotection to chlorophyll from photodamage (Miyashita et al., 2020).

The increasing demand for fucoxanthin and significant market potential in the cosmetic, nutraceutical, and pharmaceutical industries drive the need for reliable alternative production methods. Currently, fucoxanthin is extracted from macroalgae, but microalgal-based production offers a more sustainable and efficient alternative. Hence, this study investigates the potential of *Isochrysis galbana*, a microalga with high fucoxanthin content and scalability. The main objective of this work is to optimize fucoxanthin production in *I. galbana*, leveraging its biosynthetic pathway and rapid growth rate to develop a commercial-scale production process. However, variable fucoxanthin yields and suboptimal production in microalgae could hinder profitability. To address this challenge, this study aims to integrate machine learning (ML) models for predicting the fucoxanthin yield, leveraging data-driven insights to optimize production and enhance scalability. By developing prediction models, this study contributes to the development of sustainable, data-driven production methods, advancing the frontier of microalgal biotechnology.

I. galbana belongs to the class of flagellated marine microalgae that shows a higher accumulation of lipids, omega-3 polyunsaturated fatty acids, and fucoxanthin. The absence of a cell wall in this species allows for an easy extraction of fucoxanthin during downstream processing (Sun et al., 2019). The inherent capability of microalgae to produce a higher fucoxanthin yield and short life cycle, independent of seasonal variations, could be cultivated all year round, and not competing for land makes it a promising source of fucoxanthin (Yusof et al., 2022). This microalgal species was mostly studied to enhance lipid production (Cañavate et al., 2020; Chin et al., 2023; Cui et al., 2021); this study is the first of a kind to explore the impact of various phytohormones to enhance fucoxanthin production.

The common method used to estimate the pigment concentration is high-performance liquid chromatography (HPLC), which requires long extraction and column time for each run, is time- and cost-consuming, and requires highly skilled persons to maintain the equipment (Chong et al., 2023). Even though HPLC is the conventional method to determine the concentration of microalgal carotenoids, is very tedious to operate as well as time-consuming. Additionally, the use of hazardous solvents for HPLC analysis (acetonitrile and methanol) hinders the suitability of this method, making it less sustainable (Thiviyanathan et al., 2024). In contrast, the UV spectrometry-based equations could be readily used for the quantification of fucoxanthin, which could significantly reduce the delay in obtaining the microalgal fucoxanthin concentrations compared to HPLC

while retaining adequate accuracy. Hence, a high-throughput method should be simple, accurate, and reliable for the extraction and detection of pigment.
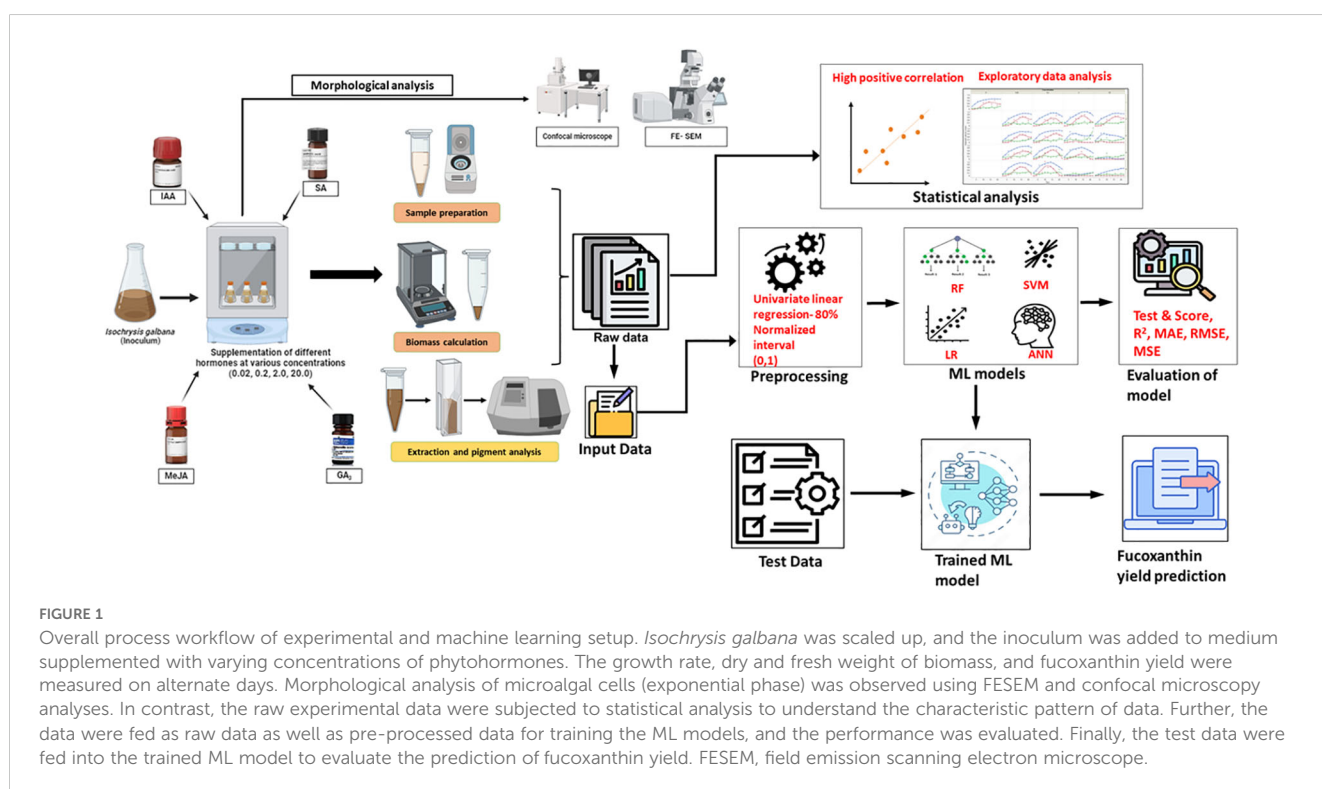
Here, we employed the equation derived by Wang et al. (2018) for UV spectrometry-based quantification of fucoxanthin. While the HPLC method requires at least 3 h to quantify the fucoxanthin, this method could detect the fucoxanthin within 5 min (Wang et al., 2018). It is to be noted that spectrophotometric analysis measures the reflectance of microalgal extract at a specific wavelength and utilizes a formula to determine the concentration of the fucoxanthin (Wang et al., 2018). This method could be flawed in a few instances, when absorption spectra of other pigments overlap with the fucoxanthin spectrum, for example, fucoxanthin with chlorophyll. Hence, the application of this spectrometric analysis for fucoxanthin yield in large-scale applications is considered to be time-consuming and limited in technological advancement (Tang et al., 2023).

Thus, ML models were implemented for the prediction of fucoxanthin yield, as they require less solvent, short analysis time, and low cost and have high accuracy and good prediction. The accuracy of the prediction of ML models depends on the input variables and the training dataset. In this study, the experimental data of *I. galbana* supplemented with different types and concentrations of phytohormones were subjected to statistical analysis followed by data preprocessing to train the ML models for fucoxanthin prediction.

The advancements in ML and artificial intelligence (AI) algorithms have profoundly contributed to the easy search for novel natural product-based drug discovery in the 21st century (Manochkumar and Ramamoorthy, 2024). Recently, numerous omics-related datasets have been developed for diverse species of marine organisms, and the need to develop and integrate ML algorithms for multi-omics studies has been extensively reviewed (Manochkumar et al., 2023). In crop breeding research, multimodal data from three sensors coupled with ML algorithms were efficiently used in a study for the estimation of the crop harvest index of faba bean and pea (Ji et al., 2024). Similarly, ML-based phenotyping combined with optical tomography was used to measure the stomatal density and improve the water use efficiency of sorghum crop (Ferguson et al., 2021).

In previous studies related to microalgae, an ML model was incorporated to derive a spectrophotometric equation for simultaneously quantifying the concentration of chlorophyll, violaxanthin, zeaxanthin, and lutein from *Chlorella vulgaris* and *Scenedesmus almeriensis* (Victor and Camarena-Bernard, 2023). Similarly, the convolutional neural network (CNN) model was used to predict the microalgal pigments including chlorophyll *a*, phycocyanin, lutein, fucoxanthin, and zeaxanthin from diatoms using experimental data obtained from water samples (Pyo et al., 2022). A hybrid ML-based approach was developed to optimize the production of biomass and phycobiliproteins in *Nostoc* sp. (Saini et al., 2021). Furthermore, Tang et al. (2023) compared linear regression with the artificial neural network (ANN) model to predict the chlorophyll concentration in *Desmodesmus* sp. and *Scenedesmus* sp. based on RGB, CYMK, and HSL color models. In this study, we constructed two ML frameworks to compare and evaluate the predictive performance of four models on fucoxanthin production from *I. galbana* by altering the input data parameters. The overall process workflow of experimental and ML setup for fucoxanthin production is depicted in Figure 1.

FIGURE 1
Overall process workflow of experimental and machine learning setup. *Isochrysis galbana* was scaled up, and the inoculum was added to medium supplemented with varying concentrations of phytohormones. The growth rate, dry and fresh weight of biomass, and fucoxanthin yield were measured on alternate days. Morphological analysis of microalgal cells (exponential phase) was observed using FESEM and confocal microscopy analyses. In contrast, the raw experimental data were subjected to statistical analysis to understand the characteristic pattern of data. Further, the data were fed as raw data as well as pre-processed data for training the ML models, and the performance was evaluated. Finally, the test data were fed into the trained ML model to evaluate the prediction of fucoxanthin yield. FESEM, field emission scanning electron microscope.

2 Materials and methods

2.1 Algal strain and culture conditions

I. galbana, a marine water microalga, was obtained from the National Repository for Microalgae and Cyanobacteria, Bharathidasan University. The seawater used for medium preparation was collected freshly from Mandapam, Tamil Nadu (9°16′17.9″N79°07′49.4″E) and filtered through a 0.22-µm membrane filtration system followed by the sterilization using autoclave for 20 min at 121°C. The salinity and pH of the sterilized seawater must be within 27 ± 1 and 8 ± 0.5, respectively. The experimental setup was maintained under controlled laboratory conditions with optimum temperature (23°C ± 2°C), light intensity (2,000 lx), and photoperiod (16-h dark:8-h light) for 30 days.

The *I. galbana* stock solution was maintained in Conway's medium for 14 days, and its density was adjusted to 2.5 mg/mL of wet biomass using sterile seawater. The algal suspension was then partitioned and added into sterile conical flasks each containing 150 mL of medium-enriched seawater. Then, the freshly prepared phytohormones were added to the flasks at specific concentrations (Supplementary Table S1). The concentration of phytohormones was based on previous studies (Chu et al., 2019; Fierli et al., 2022; Mc Gee et al., 2020). Each treatment employed three biological replicates. Cultures were cultivated in conical flasks supplemented with various phytohormones, maintaining consistent conditions of light and temperature as those used for stock culture maintenance. The culture medium without the addition of phytohormones was used as the control.

2.2 Experimental data collection

The growth rate of microalgae was monitored by measuring the optical density (OD) of the algal suspension culture every alternate day using a UV–Vis spectrophotometer (Cary 3500 Multicell, Agilent Technologies, Santa Clara, CA, USA) at 680 nm (Hawrot-Paw et al., 2019). For the spectrometry-based quantification of fucoxanthin yield, the absorbance of the cultures was measured at 750 nm on alternate days. Simultaneously, 1 mL of sample from each flask was centrifuged at 7,000 rpm, and the pellet was resuspended in 1 mL of ethanol. The absorbance of the supernatant was then measured at 445 and 663 nm within 5 min of extraction (Wang et al., 2018), and the fucoxanthin yield in cultures supplemented with phytohormones was calculated (Equation 1).

$$C_{fuc}' = 6.39(OD_{445}) - 5.18(OD_{663}) + 0.312(OD_{750}) - 5.27 \qquad (1)$$

where OD_{445}, OD_{663}, and OD_{750} are the absorbance at 445 nm, 663 nm, and 750 nm, respectively.

For the estimation of fresh weight biomass, pre-weighed Eppendorf loaded with 1 mL of harvested sample was weighed and was allowed to dry at 60°C till constant weight was obtained to determine the dry weight biomass. The fresh weight and dry weight were calculated by taking the difference between the initial and final weight. This experiment was conducted for 30 days by measuring the growth rate, biomass, and fucoxanthin yield on alternate days.

2.3 Morphological data acquisition

2.3.1 Field emission scanning electron microscopy

The cells of *I. galbana* (control and phytohormone-treated cells) at exponential phase were fixed in sterile seawater using 2% glutaraldehyde and 4% paraformaldehyde in a shaker (1 h at room temperature) followed by rinsing with Milli-Q water. Then, the cells were subjected to dehydration by sequential ethanol wash (de Haan et al., 2024). Dried cells were sputter-coated, and images were recorded using a field emission scanning electron microscope (FESEM; FEI QUANTA 250 FEG, Thermo Fisher Scientific, Waltham, MA, USA) to analyze the morphological changes in cell structure in response to phytohormone treatment.

2.3.2 Confocal laser scanning microscopy

Confocal laser scanning microscopy was employed to scrutinize the fluorescence of chlorophyll, lipids, and pigments within the cells. The cell suspensions of *I. galbana* without hormone treatment (control) and cultures treated with hormones were harvested and subsequently centrifuged at 8,000 rpm for 5 min. Pellets were resuspended in phosphate-buffered saline (PBS) buffer (Yadav et al., 2023a). Nile red (9-diethylamino-5*H*-benzo[α]phenoxazine-5-one) staining was performed 15 min before imaging to detect the presence of lipid by adding 380 µL of microalgal suspension to 20 µL of Nile red [Sigma, Darmstadt, Germany; stock solution of 0.2 mg/mL in dimethyl sulfoxide (DMSO)]. Approximately 5 µL of the algal suspension was loaded, and images were recorded using the confocal laser scanning microscope Fluoview Fv3000 at 40× objective (Olympus, Tokyo, Japan). The detection ranges were as follows: λ_{exc} = 488 nm and λ_{em} = 510–630 nm for carotenoids, λ_{exc} = 560 nm and λ_{em} = 640–750 nm for chlorophyll, and λ_{exc} = 530 nm and λ_{em} = 636 nm for Nile red (Zienkiewicz et al., 2020; Duval et al., 2023).

2.4 Machine learning-assisted fucoxanthin prediction

This study utilized four models [Random Forest (RF), Support Vector Machine (SVM), Linear Regression (LR), and ANN] to predict the optimized concentration and type of hormone for enhanced fucoxanthin productivity. The performance of these models was compared for the selection of the optimal prediction model. These models were chosen owing to their ability to analyze complex biological data (Kang et al., 2023).

RF is one of the most used ML-based ensemble-learning methods, which constructs a forest using multiple decision trees for training and predicting the samples by random extraction (Chen et al., 2023). Each decision tree generates the identification output for unknown test data. Based on the identification output of

all decision trees, the final identification output is generated for the unknown test sample. The greater the number of output times for a specific category, the more likely that the unknown test data belong to it. The process of calculation is simple and easy to understand and interpret, yet it could lead to overfitting performance. The parameters employed for computation of output by RF include the number of trees as 10 and the number of attributes considered at each spit as 6. The features utilized for RF are replicable training, and the number of features in the subset could not be less than 4.

SVM is one of the supervised learning methods of ML algorithms that work based on statistical learning theory (Pisner and Schnyer, 2020). It is effective in high-dimensional space and could be used for identification, regression, and classification tasks and could function better in conditions where the number of dimensions is higher than the number of samples. The data are effectively separated between two categories using a hyperplane for two-dimensional data followed by mapping of test points and prediction of its category depending on the side of the gap they belong to. This method could solve the computational complexity and high-dimensional issues efficiently. The major disadvantage is that it has less sensitivity to data, and hence, it is strenuous to find appropriate kernel functions for non-linear data. In this study, the parameters for SVM were given as cost (c) = 1, regression loss epsilon (ε) = 0.10, and tolerance limit = 0.0010. The radial basis function (RBF) kernel was employed in this study, and the iteration limit was set to 100.

LR falls within the realm of supervised machine learning algorithms, which operate by learning from labeled datasets and fitting the data points to optimal linear functions. These functions can then be used to predict outcomes for new datasets. It is effective in predictive analysis and provides a linear relationship between dependent and independent variables for the prediction of outcomes (Maulud and Abdulazeez, 2020). The least absolute shrinkage and selection operator (LASSO) regression (L^1 norm) was utilized for linear regression with a regularization strength of α = 0.001.

ANN could train itself for the recognition of patterns in a dataset and the prediction of non-linear relationships between input variables and output (Kumar et al., 2024). It is demonstrated to be the research hotspot in the field of artificial intelligence and is commonly referred to as a neural network (Chen et al., 2023). A multilayer fully connected feed-forward ANN was applied in this study to develop a model for the prediction of fucoxanthin yield (Supplementary Figure S1). It comprises an input layer, an output layer, and one or more hidden layers. Although the flexibility of the model could be enhanced by increasing the number of hidden layers, one hidden layer is adequate to model the microalgal growth. The process was repeated until the achieved mean squared error (MSE) was as low as possible. All the ML models and the data processing process of this study have been developed using the JMP and Orange software (Demšar et al., 2013).

2.5 Construction of models for prediction of fucoxanthin yield

The ML models thus developed were used as the driving engine and compared for the accuracy of prediction based on the data used to train the model. In this study, two ML frameworks (Case Study 1 and Case Study 2) were constructed for the inclusion and exclusion of hormone descriptors to train the model and compare its prediction accuracy.

2.5.1 Case Study 1 (without descriptors)

The ML framework is constructed in a way that when the concentration of hormones, number of days, growth rate, dry biomass, and fresh biomass are given as input, the model will be able to predict the fucoxanthin yield as output. For the initial model, no descriptors will be given for the hormones; hence, the prediction will be completely based on the input parameters.

2.5.2 Case Study 2 (with descriptors)

We constructed and developed an integrated ML framework to incorporate the characteristics of hormones; hence, descriptor values were given to the hormones in addition to the pre-processed experimental data. The input data including days, concentration, and descriptors of hormones will be given as input to the first model, which will predict the growth rate. The output of the first model (i.e., predicted growth rate) will be given as input to the second model, which will finally predict the fucoxanthin yield.

2.6 Data evaluation

2.6.1 Evaluation of model performance

To evaluate the accuracy of ML models, 70% of the sample data were selected as the training dataset, and the remaining 30% were used as the testing dataset. The ML models were trained with the experimental data obtained from supplementation of indole-3-acetic acid (IAA), salicylic acid (SA), gibberellin A3 (GA$_3$), and methyl jasmonate (MeJa) phytohormones, whereas abscisic acid hormone was used as testing data. The modeling process was repeated 200 times to minimize the errors. The prediction accuracy of the ML models was evaluated using four indicators: the coefficient of determination (R^2), root mean squared error (RMSE), MSE, and mean absolute error (MAE) (Equations 2–5, respectively). Therefore, these indicators could better measure the degree of fitness between actual and simulated values.

$$R^2 = 1 - \sum_{i=1}^{n} (Xi - \hat{X}i)^2 / (Xi - \bar{X})^2 \qquad (2)$$

$$RMSE = \sqrt{\sum_{i=1}^{n} (X_i - \hat{X}_i)/n} \qquad (3)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{X}_i)^2 \qquad (4)$$

$$MAE = \sum_{i=1}^{n} |X_i - \hat{X}_i|/n \qquad (5)$$

where n is the total number of samples; Xi and $\hat{X}_i$ are the actual measured and predicted fucoxanthin yield of the samples, respectively; and $\bar{X}$ denotes the mean of the measured fucoxanthin yield.

2.7 Statistical data analysis

Data from microalgal cultivation were processed, and exploratory data analysis was performed to understand the characteristic data pattern and correlation between the input parameters. Supervised machine learning using four ML models was carried out to explore and observe the correlational relationships between the microalgal growth rate, biomass, and fucoxanthin yield as affected by various types and concentrations of phytohormones, followed by the construction of heatmaps to visualize multidimensional data and compress and simplify the complex scientific process. All further statistical modeling and figure generation were performed using JMP® and RStudio (JMP®, 2017; R Core Team, 2017).

3 Results

3.1 Data acquisition and visualization

The proposed approach endorsed simultaneous data collection from the *I. galbana* to monitor the growth as well as cellular production of fucoxanthin and biomass. The spectrophotometric quantification of fucoxanthin showed advantages in terms of time and cost. The default experimental setup ensured that the result of the proposed method would not be affected by temperature and light. Hence, the results will be impacted by the type and concentration of hormones and number of days. I 1, I 2, I 3, and I 4 indicate the hormones IAA, SA, GA$_3$, and MeJa, respectively.
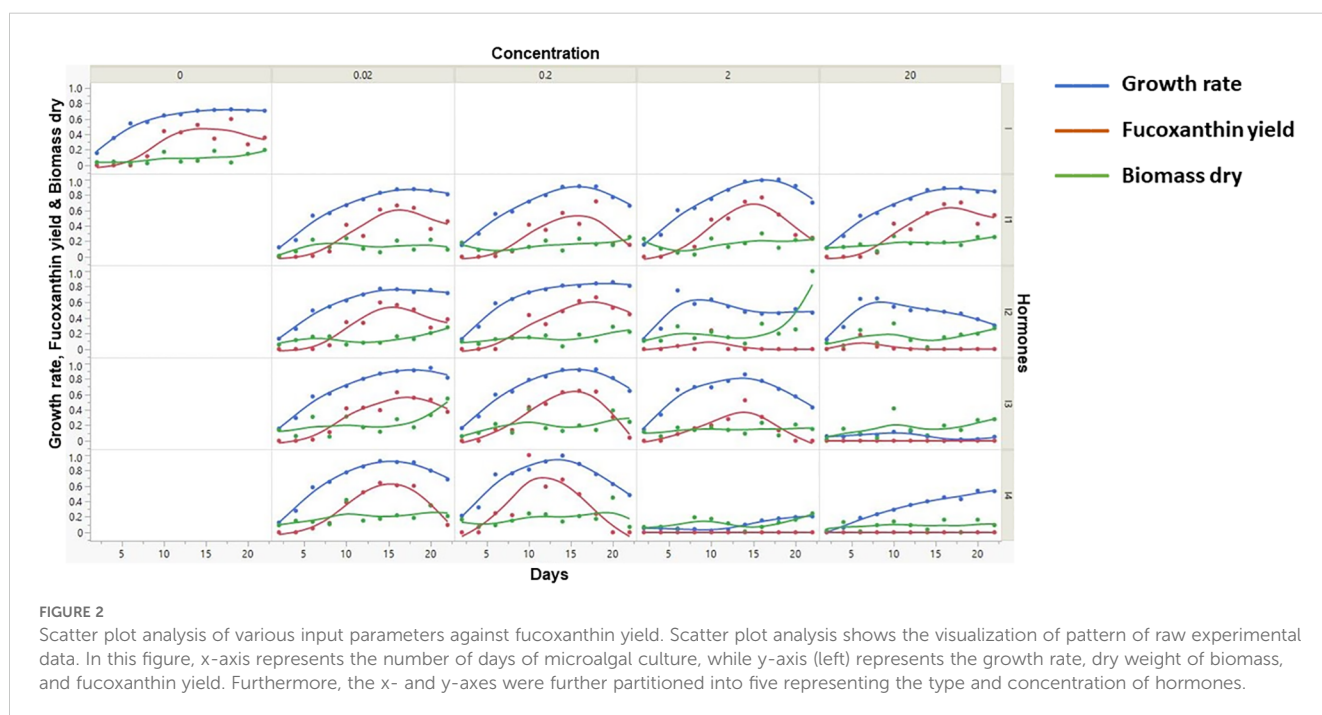
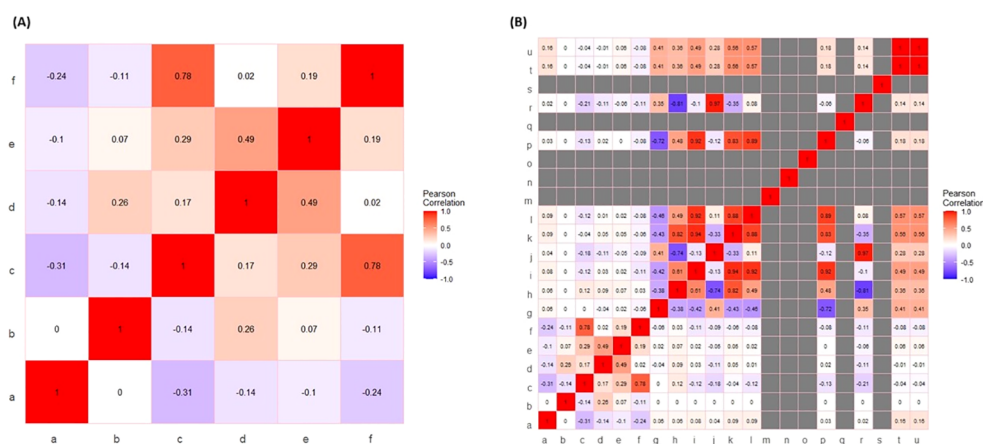## 3.2 Statistical models and correlation analysis

The data from the supplementation of four phytohormones were analyzed by statistical analysis (scatter plot) to explore and understand the data distribution across various parameters (Figure 2). Among them, the growth rate of microalgae is directly proportional to the fucoxanthin yield. Furthermore, the days at which maximum growth rate and fucoxanthin yield were attained are tabulated (Supplementary Table S2). Overall, the maximum fucoxanthin yield was achieved with 0.2 mg/L MeJa supplementation in minimal time within 10 days.

In this study, the experimental data excluding hormone descriptors and including hormone descriptors allowed Pearson's correlation coefficient analysis of fucoxanthin across the investigated complete set of input parameters (Figure 3). On a relative basis, consistent with the scatter plot analysis, the fucoxanthin yield showed maximum correlation against growth rate followed by dry weight of biomass, whereas concentration and number of days show a negative correlation in both cases (Figures 3A, B), demonstrating weak and moderate associations. When descriptors of hormones were included in the input data (Figure 3B), the fucoxanthin yield showed a minimal positive correlation with the hydrogen bond donor count depicting a moderate association.

3.3 Morphological alterations in microalgal structure

FESEM analysis revealed the morphology of *I. galbana* cells at day 12 in the absence of hormone treatment and at various

FIGURE 2
Scatter plot analysis of various input parameters against fucoxanthin yield. Scatter plot analysis shows the visualization of pattern of raw experimental data. In this figure, x-axis represents the number of days of microalgal culture, while y-axis (left) represents the growth rate, dry weight of biomass, and fucoxanthin yield. Furthermore, the x- and y-axes were further partitioned into five representing the type and concentration of hormones.

FIGURE 3
Pearson's correlation analysis of input variables (excluding and including) hormone descriptors. **(A)** Representative heatmap of Pearson's correlation of input variables excluding hormone descriptors. The notations for the figure: a, phytohormone concentration; b, days; c, growth rate; d, biomass (wet); e, biomass (Dry); f, fucoxanthin yield. **(B)** Representative heatmap of Pearson's correlation of input variables including hormone descriptors. The notations for figure: a, phytohormone concentration; b, days; c, growth rate; d, biomass (wet); e, biomass (dry); f, fucoxanthin yield; g, XLogP3; h, hydrogen bond donor count; i, hydrogen bond acceptor; j, rotatable bond count; k, topological polar surface area; l, heavy atom count; m, formal charge; n, complexity; o, isotope atom count; p, defined atom stereocenter count; q, undefined atom stereocenter count; r, covalently bonded unit count; s, defined bond stereocenter count; t, undefined bond stereocenter count; u, canonicalized compound.
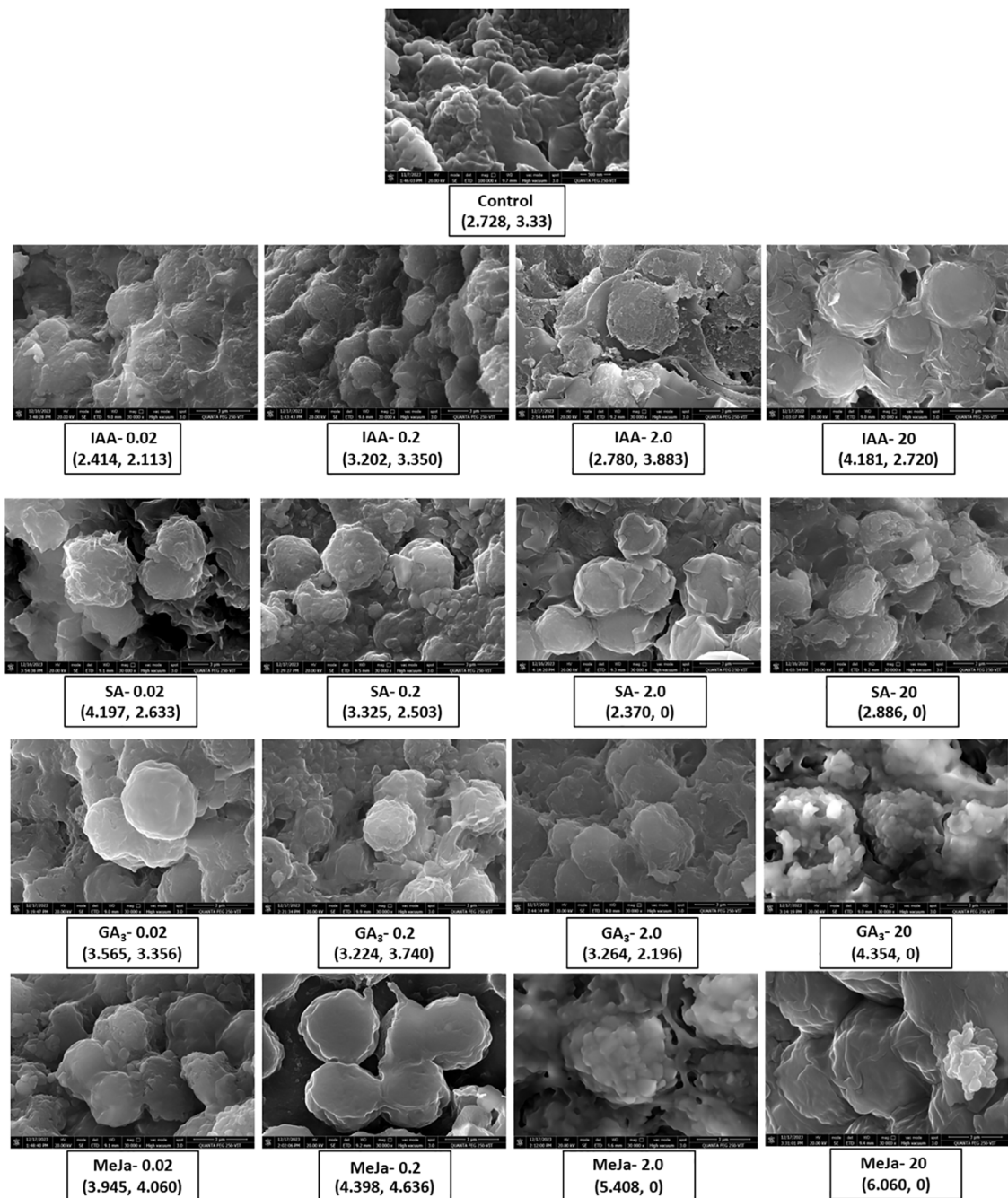
concentrations of hormone treatment (Figure 4). The size of each cell (diameter) was measured, and the shape and appearance of cells were observed to analyze the impact of phytohormone supplementation. At control, the cells appear clustered with smooth surfaces, whereas different concentrations of hormone treatment morphologically alter the structure of microalgae. The caption of each figure indicates the average diameter of the microalgal cell followed by the yield of fucoxanthin (Figure 4). For instance, when IAA was supplemented at 0.02 and 0.2 mg/L, the cells appeared clustered and enlarged, whereas higher concentrations caused the cell surface to become relatively rough with irregular grooves and increased the average cell size. In contrast, SA at 0.02 and 0.2 mg/L concentrations depicted enlarged cells with rough and distorted cell surfaces. SA of 2 and 20 mg/L made the cells appear smaller with no proper shape. The cells appeared smooth round and enlarged at 0.02 mg/L concentration of GA_3, whereas at 0.2 and 2 mg/L concentrations, they became irregularly shaped and stretchy, respectively. The supplementation of GA_3 at 20 mg/L clustered the cells with irregular grooves and protuberances on the surface. In contrast, cells appeared smooth and round at MeJa of 0.02 mg/L and swollen and enlarged with maximum production of fucoxanthin at MeJa of 0.2 mg/L. MeJa of 2 mg/L supplementation completely altered the cell with a distorted and irregular shape. At 20 mg/L, the cells appeared extremely swollen, which caused the cell to explode. Hence, the morphology of cells was altered depending on the type and concentration of the hormone supplementation to medium.

Further, the impact of phytohormone treatment on the presence of lipid, pigment, and chlorophyll content within *I. galbana* was visualized using confocal microscopy analysis. Carotenoids are lipophilic pigments present in the interior and exterior of chloroplasts and are detected as green globular forms using Nile red stain under confocal microscopy whereas chlorophyll autofluorescence as red globules. The chlorophyll autofluorescence of *I. galbana* cells in the exponential phase reveals that the type and concentration of phytohormone supplementation negatively affect the chlorophyll content. For instance, supplementation of IAA and SA at higher concentrations demonstrated higher chlorophyll content, whereas supplementation of GA_3 and MeJa at higher concentrations demonstrated degradation of chlorophyll (Figure 5A). During IAA and SA supplementation, the lipid droplets increased in size and number, whereas GA_3 and MeJa supplementation progressively decreased the size of lipids (Figure 5B). As fucoxanthin belongs to xanthophyll carotenoids, the carotenoid fluorescence emission is detected as green globules at 488-nm excitation, whereas chlorophyll was detected by red light excitation at 560 nm. Similar to lipids, the type and concentration of hormone supplementation affected the accumulation of carotenoids. MeJa of 0.2 mg/L showed the maximum carotenoid accumulation (Figure 5C). The merged fluorescence was emitted by lipids and pigment and chlorophyll autofluorescence within *I. galbana* in the absence and presence of varying concentrations of hormone supplementation (Figure 5D).

3.4 Machine learning-based fucoxanthin prediction

In this study, the experimental dataset was divided into training and testing data, and four ML models (SVM, RF, LR, and ANN) were adopted for the prediction of fucoxanthin yield, as these models are extensively used for analyzing complex biological data (Figure 6A). The reliability of the models was evaluated based on previously trained data. The parameters considered for the

FIGURE 4
FESEM analysis of *Isochrysis galbana* without and with hormone treatment. Impact of varying concentrations of phytohormones on the morphological analysis of surface and average size of cells of *I. galbana*. The representative caption of each hormone concentration indicates the average cell size and the yield of fucoxanthin. The appearance of cell clusters and their surface morphology give insights into how phytohormones affect the cellular morphology of microalgae to aid in enhanced production of fucoxanthin. FESEM, field emission scanning electron microscope.

construction of ML models and the framework of models for Case Study 1 and Case Study 2 are illustrated in Figure 6B. For Case Study 1, the models were trained using whole experimental data, whereas for Case Study 2, models were trained with restricted experimental data, and descriptors for hormones were included to incorporate the characteristics of hormones. The experimental dataset (growth rate, dry biomass, fresh biomass, number of days,

and type and concentration of hormones IAA, SA, GA$_3$, and MeJa) used for the modeling step in this study included a dataset consisting of 273 samples with six variables (excluding hormone descriptors) for the Case Study 1, whereas for Case Study 2, the experimental dataset consisted of 273 samples with 24 features (including hormone descriptors) for prediction of fucoxanthin yield.
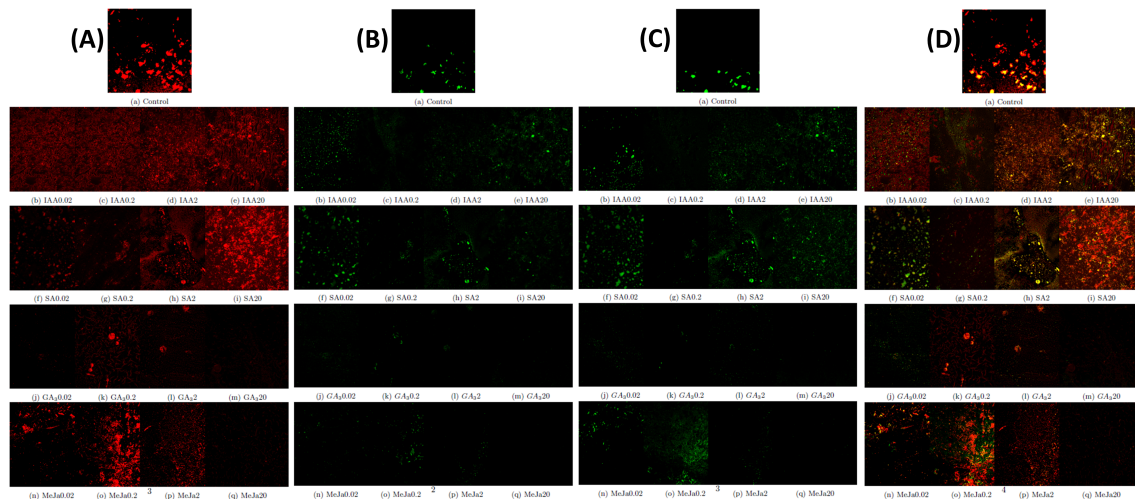
FIGURE 5

Cellular changes in chlorophyll, lipid, and pigment content of *Isochrysis galbana* in response to phytohormone supplementation. **(A)** Representative confocal microscopy images of chlorophyll autofluorescence (red) *I. galbana* cells in response to phytohormone supplementation. **(B)** Representative confocal microscopy images of Nile red stained-lipid fluorescence (green) of cells in response to phytohormone supplementation. **(C)** Representative confocal microscopy image of pigment fluorescence (green) of *I. galbana* cells in response to phytohormone supplementation. **(D)**. Shown are the merged fluorescence of all signals of chlorophyll, lipids, and pigments of *I. galbana* cells.

3.5 Performance comparison between ML models (Case Study 1—exclusion of hormone descriptors)

For Case Study 1, the RF model performance for fucoxanthin yield prediction provided higher R^2 values and lower RMSE, MSE, and MAE values ($R^2 = 0.809$, $MSE = 0.602$, $RMSE = 0.776$, and $MAE = 0.458$). For the prediction of fucoxanthin yield (Table 1), R^2 values of LR and SVM were lower than those of the RF and ANN models. Among the four models, RF provided the maximum accuracy in fucoxanthin yield prediction followed by the ANN model ($R^2 = 0.722$). Compared with other models, LR was the algorithm with the poorest performance for predicting the fucoxanthin yield ($R^2 = 0.605$, $MSE = 1.248$, $RMSE = 1.117$, and $MAE = 0.906$). The randomly selected predictions made by four ML algorithms at specified instances show that the RF and ANN models predicted the fucoxanthin yield with maximum accuracy and suggested that MeJa (0.2 mg/L) proved to synthesize maximum fucoxanthin compared to other hormones (Table 1). Therefore, the RF and ANN models were adopted as the optimized modeling methods for fucoxanthin prediction for Case Study 1.
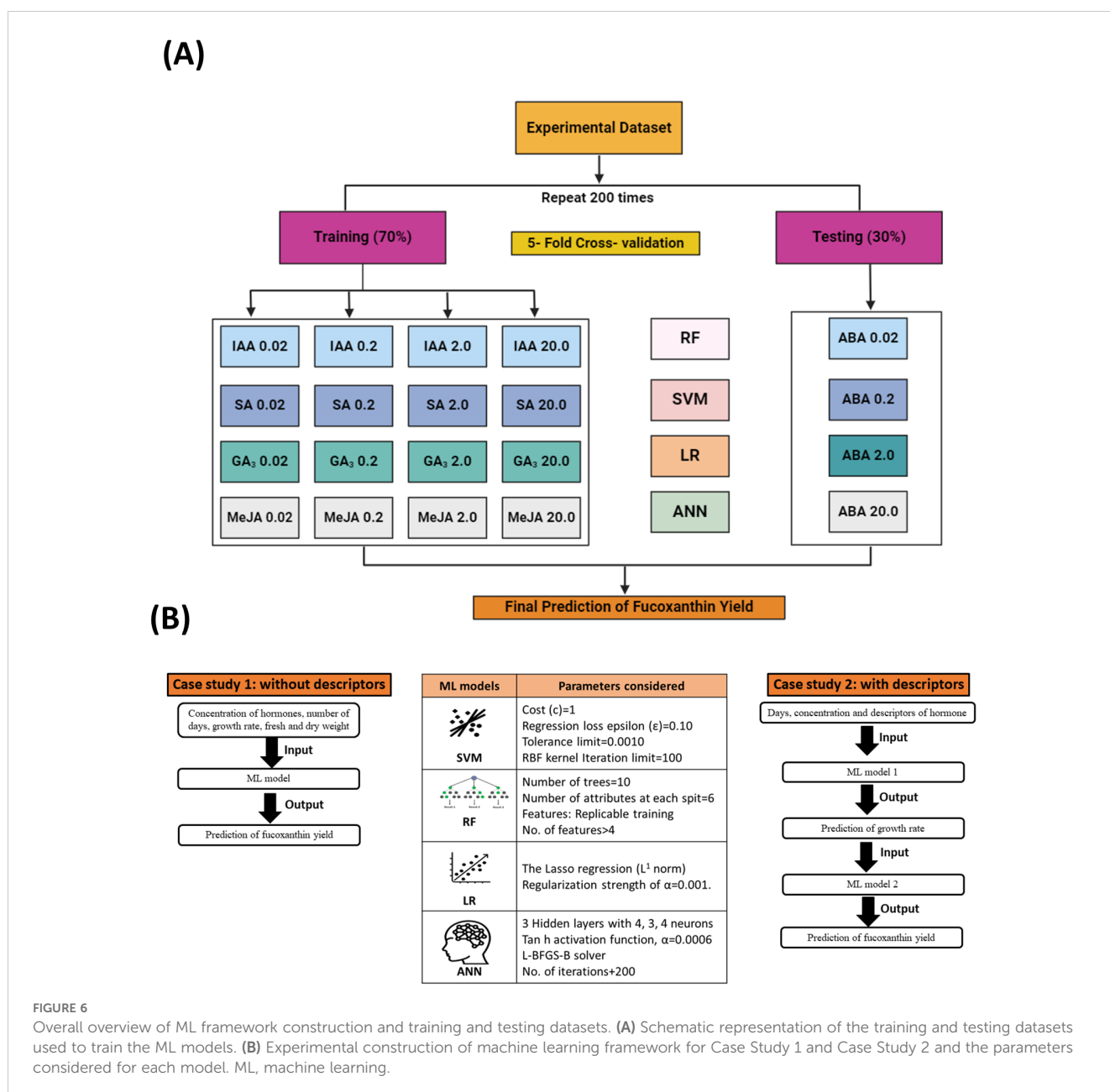
The randomly selected ML-predicted fucoxanthin yield at different days, types, and concentrations of hormones show the differences in the prediction of ML models (Supplementary Table S3). From these predictions, it can be inferred that MeJa could yield maximum fucoxanthin production at lower concentrations in a shorter time followed by GA$_3$. The hormones IAA and SA were able to produce higher fucoxanthin after 15 days. However, the results obtained by ML prediction were purely based on the training and experimental data, as the hormone descriptors have not been included. Hence, descriptors of hormones were given as an additional input to the developed model, and prediction performance was evaluated (Supplementary Table S4). Consistent with the previous results, the

RF and ANN models ($R^2 = 0.825$ and $R^2 = 0.746$) showed better predictions respectively. This result suggests that the inclusion of hormone descriptors in input data improved the prediction accuracy of fucoxanthin yield (Supplementary Table S5).

3.6 Performance comparison between ML models (Case Study 2—inclusion of hormone descriptors)

3.6.1 Growth rate prediction using pre-processed data

As the inclusion of hormone descriptors improved the prediction accuracy (Case Study 1), a generic integrated ML model was constructed exclusively to incorporate the hormone characteristics, and the predictive performance of ML models for growth rate and fucoxanthin yield was evaluated. The experimental data were pre-processed before training the ML models to avoid discrepancies. In this model, growth rate and fucoxanthin yield were predicted by varying the input data (Figure 6B). The prediction results of ML models (Table 2) showed ANN to predict growth rate with maximum accuracy ($R^2 = 0.836$) followed by RF ($R^2 = 0.82$). Additionally, the lower values of MAE, MSE, and RMSE along with higher R^2 value indicated ANN with higher prediction accuracy. These results indicate that the ANN model demonstrated better performance in the prediction of the growth rate in several instances, whereas LR showed the poorest prediction accuracy of the growth rate. However, RF failed to provide the expected maximum estimated accuracy at growth rate prediction, which was provided by ANN (Supplementary Table S6). These results demonstrate that the ANN model performed better than the other models in predicting the growth rate of *I. galbana*. Hence, these trained models will be further used to imply the actual prediction of fucoxanthin yield.

FIGURE 6
Overall overview of ML framework construction and training and testing datasets. **(A)** Schematic representation of the training and testing datasets used to train the ML models. **(B)** Experimental construction of machine learning framework for Case Study 1 and Case Study 2 and the parameters considered for each model. ML, machine learning.

3.6.2 Fucoxanthin prediction using pre-processed data

In this study, the ML models were fed with predicted growth rate as input data for the prediction of the fucoxanthin yield by combining the advantages of integration of the ML models and avoiding overfitting or overestimating. Compared to all the above models, the RF model (R^2 = 0.839) employed in this method gave the maximum accuracy for fucoxanthin yield prediction followed by the ANN model (R^2 = 0.738), whereas the predictive performance of SVM and LR was better than that of the previously developed models (Table 3). In several instances, the RF model fed with pre-processed experimental data showed a better prediction of fucoxanthin yield followed by the ANN model (Supplementary Table S7). For Case Study 2 (including descriptors), the RF and

ANN models were able to improve the generalization by integration of multiple models, thus providing a more stable prediction result. The prediction values obtained from the integration of ML models are in good agreement with the measured fucoxanthin yield from *I. galbana*, which reflects a satisfactory prediction result.

3.6.3 Prediction of growth rate using raw data

In this study, additionally, to evaluate the influence of pre-processing of experimental data on the prediction of growth rate and fucoxanthin yield, the constructed models were trained with raw data. Consistent with the results for pre-processed data, the growth rate prediction results were better with the ANN model followed by the RF model (Supplementary Table S8). The ANN model showed the maximum accuracy of growth rate predictions

TABLE 1 Test results of ML models trained with all input variables excluding hormone descriptors.

Model	MSE	RMSE	MAE	R^2
Random Forest	0.602	0.776	0.458	0.809
Linear Regression	1.248	1.117	0.906	0.605
Support Vector Machine	0.949	0.974	0.648	0.699
Artificial Neural Network	0.878	0.937	0.506	0.722

ML, machine learning; MSE, mean squared error; RMSE, root mean squared error; MAE, mean absolute error.

(R^2 = 0.846), whereas LR showed the worst growth rate prediction accuracy (R^2 = 0.156). However, the RF model failed to provide the expected estimated accuracy at growth rate prediction, which was provided by ANN (Supplementary Table S9). Hence, the artificial neural network is optimized as the best model for the prediction of growth rate for both raw and pre-processed data. Although the RF model failed to provide the best prediction accuracy for all the case studies, it achieved a more stable performance by minimizing the deviations and randomness of the other models.

3.6.4 Prediction of fucoxanthin yield using raw data

In this study, when raw data were given as input, the ANN (R^2 = 0.836) and RF models (R^2 = 0.845) achieved the maximum accuracy in the prediction of fucoxanthin yield. The predictive performance of LR remained the same as that of the model trained with pre-processed data, whereas the predictive performance of SVM decreased (Supplementary Table S10). The RF model showed better prediction similar to the measured fucoxanthin yield at several instances, whereas ANN overestimated the fucoxanthin yield at a few instances (Supplementary Table S11). Hence, these results infer that pre-processing of data shows an influence on the predictive performance of the ML models. However, contrary to the ML-based fucoxanthin prediction, the quantitative experimental values of fucoxanthin yield obtained for the *I. galbana* were lower than those obtained from ML prediction in a few instances. This discrepancy was possibly observed, as ML models at few instances overestimated the fucoxanthin production depending on the training dataset.

TABLE 2 Test results of ML models for the prediction of growth rate using pre-processed data.

Model	MSE	RMSE	MAE	R^2
Random Forest	0.011	0.105	0.083	0.820
Linear Regression	0.052	0.227	0.195	0.157
Support Vector Machine	0.021	0.144	0.116	0.661
Artificial Neural Network	0.01	0.1	0.077	0.836

ML, machine learning; MSE, mean squared error; RMSE, root mean squared error; MAE, mean absolute error.

4 Discussion

4.1 Sustainable approach: ensuring enhanced fucoxanthin and biomass production

Microalgae synthesize a wide range of bioactive metabolites including carotenoids, lipids, and polysaccharides, which makes them a sustainable source for next-generation feedstock (Foo et al., 2017). The microalgal species *I. galbana* was selected in this study because they have gained widespread application in aquaculture and animal feed due to their rapid and stable growth rates. However, compared with other microalgal species (*Phaeodactylum tricornutum* and *Chaetoceros calcitrans*), there have been fewer studies on *I. galbana* for fucoxanthin production. Recently, the impact of spermidine, a type of plant growth regulator on fucoxanthin accumulation in *Isochrysis* sp. acclimated to different light intensities, was studied. The supplementation of spermidine increased the fucoxanthin production to 6.11 mg/g under low light intensity (Bo et al., 2023). Most studies on *I. galbana* focus on the extraction of lipids (Wu et al., 2023; Yang et al., 2024).

Therefore, the current study, which focuses on predicting the fucoxanthin production of *I. galbana* through the UV spectroscopic method coupled with high-throughput ML studies in the research field, is of great significance for the future development of commercial production of microalgal fucoxanthin. The integration of ML models with biotechnological tools (UV spectrometry-based measurement of fucoxanthin yield) allows for the rapid and accurate prediction of fucoxanthin yield, which can aid in understanding the influence of different types and concentrations of hormones on the microalgal growth, biomass, and response to elicitor supplementation. By predicting the fucoxanthin production of *I. galbana*, this study can provide valuable insights into their enhanced yield potential and optimized type and concentration of hormones, aiding in the improved cultivation strategies as well as commercial fucoxanthin production strategies.

Figure 1 represents the experimental workflow of UV-based measurement of fucoxanthin coupled with ML-based fucoxanthin prediction, whereas Figure 2 represents the scatter plot analysis of experimental data, which shows that the type and concentration of hormone supplementation have an influence on the fucoxanthin yield. For control cultures, the maximum yield of fucoxanthin was achieved only on day 18, whereas the 0.02 mg/L concentration of I 1 and I 2 showed maximum fucoxanthin yield on days 10 to 15. At I 1 (0.2 mg/L, 2 mg/L, and 2 mg/L), I 2 (0.2 mg/L), and I 3 (0.02 and 0.2 mg/L), the maximum yield was obtained from days 15 to 20. The supplementation of I 4 (0.02 mg/L) could give maximum fucoxanthin yield at days 12 to 18, whereas I 4 (0.2 mg/L) concentration could attain maximum yield within 10 days. For hormones I 2, I 3, and I 4 (2 and 20 mg/L), there was negligible or minimum yield of fucoxanthin.

These results were consistent with the previous findings on phytohormone supplementation. Mc Gee et al. (2020) showed that fucoxanthin content in *Stauroneis* sp. and *Phaeothamnion* sp.

TABLE 3 Test results of generic ML models for the prediction of fucoxanthin yield using pre-processed data.

Model	MSE	RMSE	MAE	R^2
Random Forest	0.507	0.712	0.357	0.839
Linear Regression	1.27	1.127	0.919	0.598
Support Vector Machine	1.332	1.154	0.781	0.578
Neural Network	0.826	0.909	0.51	0.738

ML, machine learning; MSE, mean squared error; RMSE, root mean squared error; MAE, mean absolute error.

increased owing to the addition of MeJa (10 and 100 μM). It was also reported that MeJa (2.2 mg/L) supplementation enhanced the biosynthesis of fucoxanthin in *Stauroneis* sp (Mc Gee et al., 2021). Similar results were obtained for *P. tricornutum* cultivated with GA₃. The supplementation of SA also boosted the synthesis of carotenoids in *Nitzschia*, leading to a 1.7-fold increase in fucoxanthin content. In contrast, MeJa supplementation at 0.5 mg/L has a negligible impact on fucoxanthin yield (Fierli et al., 2022).

Additionally, Fierli et al. (2023) studied the effect of the combined application of exogenous phytohormones along with blue light in *P. tricornutum*. When GA₃ was supplemented separately, the fucoxanthin yield increased by 30%. The combined supplementation of GA₃ and ABA was demonstrated to be more effective. Therefore, supplementation of phytohormones provides a promising strategy to enhance fucoxanthin production due to their intrinsic role in promoting microalgal growth. A similar pattern of results was obtained in studies predicting the fucoxanthin production. Gao et al. (2021) studied the effect of light on biomass and fucoxanthin production in *P. tricornutum* and *Tisochrysis lutea*. The prediction models developed using fluorescence spectroscopy showed a positive correlation between biomass and fucoxanthin yield (Gao et al., 2021). However, the impact on fucoxanthin production depends on the type of hormone, concentration, and the microalgal species.

4.2 Pearson's correlation coefficient analysis

Pearson's correlation coefficient uses a correlation coefficient (R) ranging from −1 to +1 to evaluate the linear relationship between the variables X and Y. The ideal positive and negative relationships between the variables are indicated by R values of 1 and −1, respectively. The absolute magnitude of R represents the strength of correlation such that a higher absolute value indicates a greater correlation. An absolute value of $R > 0.6$ is considered a robust correlation. We detected a positive correlation between growth and fucoxanthin yield in both the cases of exclusion and inclusion of hormone descriptors in input data (Figures 3A, B). In this study, the growth rate of *I. galbana* shows a higher positive correlation with fucoxanthin yield ($R = 0.78$). This is consistent with the findings of previous studies showing a positive correlation of

fucoxanthin yield with microalgal growth rate followed by biomass (Li et al., 2020; Ishika et al., 2019). Recently, Sequeira et al. (2021) reviewed the positive influence of hydrogen bond donor chemicals on the yield of fucoxanthin from macroalgae as well as microalgae. Hence, consistent with theoretical expectations and prior observations, when hormone descriptors are included in input, fucoxanthin yield shows a higher positive correlation with growth rate followed by hydrogen bond donor count.

4.3 Morphological alterations in the microalgal structure in response to phytohormones

In this study, FESEM analysis of *I. galbana* at day 12 revealed that the type and concentration of hormones alter the morphological structure of microalgal cells (Figure 4). Consistent with this result, similar changes in the morphology of microalgae were observed when the concentration of nutrient supplementation was varied. For instance, variations in the nutrient composition of the culture medium morphologically altered the cell wall and structure of *Amphiprora* sp (Jayakumar et al., 2021).

The presence of lipid, pigment, and chlorophyll within *I. galbana* was visualized using confocal microscopy analysis. The chlorophyll autofluorescence-based detection method has revealed immense potential as an on-site tool to assess microalgal vitality (Li et al., 2022). However, there are very limited data on the impact of phytohormones on the presence of chlorophyll within the microalgal structure. In this study, the effects of different phytohormones with four concentrations (0.02 mg/L, 0.2 mg/L, 2 mg/L, and 20 mg/L) on the chlorophyll autofluorescence in cells of *I. galbana* were investigated by red light excitation at 560 nm (Figure 5A). Experimental results showed that both the type and concentration of hormones were major factors that caused the degradation of chlorophyll.

There are several reports on the enhanced accumulation of lipids in Nile red-stained microalgal cells grown under nutrient-stress conditions and phytohormone supplementation. In this study, the lipid accumulation was higher within the microalgal cells supplemented with IAA and SA hormones, whereas supplementation of GA₃ and MeJa at higher concentrations degraded the lipids (Figure 5B). These results were consistent with the findings of previous studies (Ahamed et al., 2022; Duval et al., 2023; Zienkiewicz et al., 2020).

Additionally, spectral analysis using a confocal laser scanning microscope was performed to investigate the alterations in fluorescence emission of endogenous pigments in *I. galbana* cells. The microalgal pigments when excited by specific wavelengths of UV–visible laser light will produce a specific emission spectrum. The fluorescence emission of carotenoids is detected in the green-yellow spectral region, whereas chlorophyll is typically detected in the red spectral region (Zienkiewicz et al., 2020). In this study, when blue light at 488-nm excitation was given to hormone-treated *I. galbana* cells, a change in spectral characteristic occurred owing to an increased carotenoid pigment (Figure 5C).

4.4 Performance of machine learning models for fucoxanthin prediction

Few previous studies have demonstrated the feasibility of the UV spectroscopic method, and the fusion of ML models to analyze the data from multiple treatment parameters could provide a better prediction of chlorophyll and other pigments. However, fucoxanthin prediction based on the UV-spectroscopic method has not been previously investigated.

4.4.1 Differences in fucoxanthin prediction metrics (Case Study 1)

It can be observed from the prediction results of ML models trained with whole input data excluding descriptors (Table 1) that the RF model is the most stable and showed higher accuracy with less error rate ($R^2 = 0.809$, $MSE = 0.602$, $RMSE = 0.776$, and $MAE = 0.458$) followed by the ANN model ($R^2 = 0.722$, $MSE = 0.878$, $RMSE = 0.937$, and $MAE = 0.506$). Consistent with previous results, ML models trained with whole data including hormone descriptors (Supplementary Table S4), and the RF and ANN models showed the maximum prediction accuracy. The major advantage of the RF model over the other ML models is that it utilizes an integrated learning algorithm to generate multiple decision trees for learning and prediction. The average of each decision tree was used to attain the final prediction. Thus, this assures robust training and decreases the chances of overfitting and the influence of noised data. In contrast, LR and SVM models enable single training from the input dataset without statistical average and bootstrap sampling. Hence, compared with other models, RF models show better performance as per previous studies (Chen et al., 2023; Lei et al., 2019). ML models could effectively capture the influences of parameters by altering the growth and biomass concentrations of microalgae compared to conventional mathematical models owing to their complexity (Yu et al., 2024). For instance, Raj et al. (2021) investigated response surface methodology (RSM) and ANN for the optimization of factors involving biodiesel production from *Nannochloropsis salina*, which proved the ANN model to be the optimized model ($R^2 = 0.957$). Similarly, Tang et al. (2023) utilized linear regression and the ANN model for the prediction of chlorophyll content in microalgae compared to the conventional spectrophotometric method, which showed ANN to be an effective prediction model. Hence, ML models demonstrate their ability to accurately predict the fucoxanthin yield in *I. galbana* in this study. However, the predictive results depend on the input data and training process and are independent of the biological process behind modeling. Hence, the training data were further modified, and the models were further evaluated for their fucoxanthin predictive performance.

4.4.2.1 Differences in growth rate prediction metrics (Case Study 2)

As in Case Study 1, the inclusion of hormone descriptors in the basic model improved the prediction accuracy of fucoxanthin yield; we constructed an integrated ML model framework exclusively for the inclusion of hormone descriptors and pre-processed the

experimental data to avoid further discrepancies. It can be observed that the construction of the ML model (Figure 6) and the inclusion of hormone descriptors in pre-processed input data enhanced the prediction accuracy compared to Case Study 1 (Table 2; Supplementary Table S6). However, the ANN model showed maximum accuracy in the prediction of the growth rate ($R^2 = 0.836$), whereas the RF model showed maximum accuracy in the prediction of fucoxanthin yield using pre-processed data ($R^2 = 0.839$) (Table 3). Furthermore, these results are in strong accordance with ML models trained with raw data (Supplementary Tables S8, S9), as the ANN model showed the maximum growth rate prediction accuracy ($R^2 = 0.846$). As far as the ANN model is considered, the number of nodes in the hidden layer plays a vital role in the performance; hence, the models should be carefully selected based on the dimensions of the input parameters and output as well as the number of samples being trained. A higher number of nodes could lead to overfitting performance, whereas inadequate nodes could relatively suppress the generalization capability (Fiorentini et al., 2023). Similarly, the ANN model depicted a validation of ($R^2 = 0.98$) in predicting the growth rate of *Synechocystis* sp. at different light regions (Yu et al., 2024). Consistently, in this study, the ANN model proved to be effective in predicting the growth rate of microalgae.

4.4.2.2 Differences in fucoxanthin yield prediction metrics (Case Study 2)

When the constructed model was trained with previously predicted growth rate as input (Table 3), the RF model showed the maximum fucoxanthin prediction accuracy ($R^2 = 0.839$) followed by the ANN model ($R^2 = 0.738$). The prediction results of fucoxanthin yield by generic integrated ML model trained with growth rate from previous model showed RF to be the best model (Supplementary Table S7).

Consistent with the above results, ML models trained with raw data (Supplementary Tables S10, S11) gave the best fucoxanthin prediction results with the ANN model ($R^2 = 0.836$) and the RF model ($R^2 = 0.845$). The predictive performance of LR and SVM was better than that of the previously developed models. These results are in strong accordance with the concept that neural network effectively processes the non-linear characteristics of data when enough data and neurons are given. For an ANN model to be ideal, it requires three vital functions to operate (Otálora et al., 2021). The major requirement is that the data should be adequate for training and validation of the model. The second vital function is the construction and structure of the neural network, which includes the selection of the type, size, and choice of layers based on the problem addressed, input type, amount of data, and complexity of the model to be developed. The final part of developing an ideal model lies in the process of training, which is defined by the calculation frequency of input parameters, duration of the training, type of data used for training, and the stop factors (Hudson et al., 1992). In contrast, as SVM and LR process only the linear characteristics of data, they demonstrated poor performance in both Case Study 1 and Case Study 2.

4.4.3 Performance of ML models (test data) in fucoxanthin prediction

Marine biotechnological research is progressing swiftly, with a burgeoning interest in utilizing multi-omics approaches and machine learning techniques to analyze marine metabolite datasets (Manochkumar et al., 2023). The developed integrated ML model harnesses the complementary strengths of the basic models to minimize the occurrence of random errors, thereby enhancing the reliability of its predictions. When abscisic acid phytohormone (predictions and actual measured values are highlighted in red) was used as testing data, the RF and ANN networks showed the maximum prediction accuracy (Supplementary Tables S3, S5–S7, S9, S11). Even though RF showed stable prediction, the ML models overestimated the fucoxanthin yield of abscisic acid in several instances when compared to the actual measured values. This requires the need to train the ML model with increased sample size and different phytohormones.

Yadav et al. (2023b) investigated the impact of the ANN-GA model and statistical RSM-based model to optimize the process parameters and elevate the production of isoprene in engineered *Synechococcus elongatus* UTEX 2973. The ANN-GA model combined with the metabolic pathway inhibition strategy performed better than the statistical model and achieved a 29.52-fold higher isoprene yield. Similarly, Kang et al. (2023) demonstrated the ML-guided prediction of engineered *Deinococcus radiodurans* R1 for enhanced lycopene production. The multilayer perceptron models combined with the genetic algorithm predicted the potential overexpression targets from 2,047 combinations of key genes. This model achieved a threefold increased lycopene production from glycerol and a sixfold increased lycopene yield. Yeh et al. (2023) investigated the use of ML models for the modeling and growth monitoring of microalgae. The performance of Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) was compared for outdoor cultivation of *P. tricornutum* in flat-panel airlift photo-bioreactors. The LSTM model outperformed the SVR model and showed its potential ability to capture the acclimation effects of light on microalgal growth. Recently, data-enhanced interpretable ML was used to predict the biochar characteristics. Data enhancement significantly improved the model accuracy from 5.8% to 15.8%. Compared to the ANN and SVM models, the optimal RF model showed a maximum accuracy of 94.89% (Chen et al., 2023).

Consistent with previous studies, in our study, the addition of hormone descriptors and pre-processing of data to the constructed generic integrated model enhanced the performance of the RF optimal model to 83.9%. Therefore, the production of fucoxanthin depends on the type and concentration of hormone supplementation and number of days of cultivation. In addition, the growth rate of microalgae was directly proportional to the fucoxanthin production. Machine learning models predicted that supplementation of MeJa (0.02 and 0.2 mg/L) contributed to maximum fucoxanthin production in shorter time intervals, whereas IAA supplementation showed maximum fucoxanthin production on day 18. The created generic model was found to be more effective in predicting the fucoxanthin yield, as this is the first study to employ ML models to predict the fucoxanthin yield from microalgae.

Testing the potential combination of phytohormones to forecast the synergetic effect on fucoxanthin production and dynamics of microalgal growth will constitute a significant aspect of the upcoming research endeavors in this field. It will be intriguing to contrast various deep-learning models with the ML models employed in this study for the enhancement of fucoxanthin production. Overall, the fucoxanthin production from *I. galbana* was validated and verified by the construction of different ML models. These constructed models were only applicable in the determination of fucoxanthin yield using spectrometry-based data acquisition. This study highlights the superior potential of ML models in predicting and optimizing fucoxanthin production, outperforming the conventional quantification methods. By leveraging ML models, there is a significant increase in prediction accuracy (R^2 = 0.839) with the inclusion of hormone descriptors. This data-driven approach reduced the experimentation time as well as minimized the utilization of resources, making fucoxanthin production more sustainable and cost-effective. Also, ML provided insights into the correlation between phytohormones, growth rate, and fucoxanthin yield. Therefore, ML models could be applied as a prediction tool for the commercial production of fucoxanthin by tracking the growth rate as well as determining the fucoxanthin yield for industrial purposes. In contrast, conventional methods often rely on trial-and-error approaches, which are time-consuming and result in suboptimal fucoxanthin production. However, an ML-based approach can aid in saving time, costs, and manpower associated with optimizing the process parameters, underscoring the scalability of ML models in biotechnology applications. Our study demonstrates the exclusive potential of an ML-based approach in fucoxanthin production, paving the way for efficient, sustainable, and data-driven pigment production.

4.5 Future improvements

Although the current results are satisfactory, there are still areas for improvement that should be addressed in future research. To further enhance the prediction accuracy of fucoxanthin using machine learning models, expanding the dataset size could prove beneficial, considering that the sample size utilized in this study has certain limitations. For instance, as we trained the model with limited data using four phytohormone supplementations and four concentrations, the ML model could only capture and program as per the characteristics of the trained hormones. Future studies should include more types and concentrations of hormones to test the applicability and robustness of developed ML models. Recent studies have shown that deep learning models can effectively harness large datasets. Therefore, the incorporation of

deep learning should be considered to explore the potential applicability of UV-based fucoxanthin detection in marine research.

5 Conclusion

This study unequivocally demonstrates the potential of integrating UV-based fucoxanthin estimation with ML models as a reliable predictive tool that enhances yield accuracy and accelerates production. Findings offer insights into MeJa (0.2 mg/L) as an effective phytohormone in enhancing fucoxanthin yield to 7.83 µg/mL in a shorter time interval of less than 10 days. Compared with the basic ML models of Case Study 1, the integrated ML model (Case Study 2) contributed to higher prediction accuracy in most cases. ANN showed maximum accuracy in the growth rate prediction, whereas RF showed maximum accuracy in the fucoxanthin prediction. Moreover, the critical role of data pre-processing and hormone descriptors in enhancing prediction accuracy streamlines the optimization process. These findings open up new avenues for exploring phytohormone-mediated fucoxanthin optimization and provide a scalable, adaptable framework for predicting fucoxanthin yield. The implications of this research are that through the adoption of an integrated approach, industries could reduce the time-to-market, increase yield predictability, and minimize production risks. Future studies should focus on expanding sample and raw dataset size and exploring additional phytohormones, implementing advanced deep learning models to further solidify these findings. Ultimately, this research sets a new avenue in fucoxanthin production, characterized by sustainability, efficiency, and data-driven innovation.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author contributions

JM: Writing – original draft, Methodology, Data curation, Conceptualization. AJ: Writing – review & editing, Validation, Supervision, Software. AC: Writing – review & editing, Validation, Supervision, Software. BV: Writing – review & editing. DJ: Writing – review & editing. RC: Writing – review & editing. SR: Writing – review & editing, Validation, Supervision, Conceptualization.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1461610/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Schematic representation of Artificial neural network ML model. The input layer comprises four neurons corresponding to the four input variables (type of hormone, concentration of hormones, growth rate and biomass) and the output layer has one neuron corresponding to fucoxanthin yield. Three hidden layers with 4, 3, and 4 neurons respectively were found to give good performance. The tan h activation function was utilized in this study for the computation of output. The regularization value of $\alpha=0.0006$ and L-BFGS-B solver was used for the ANN model. The neurons between adjacent layers are fully interconnected and the number of iterations was set to 200 for the training algorithm to reduce the error between the actual and predicted output.

References

Ahamed, T. S., Brindhadevi, K., Krishnan, R., Phuong, T. N., Alharbi, S. A., Chinnathambi, A., et al (2022). *In vivo* detection of triacylglycerols through Nile red staining and quantification of fatty acids in hyper lipid producer *Nannochloropsis sp.* cultured under adequate nitrogen and deficient nitrogen condition. *Fuel.* 322, 124179. doi: 10.1016/j.fuel.2022.124179

Anantharaman, A., Hemachandran, H., Mohan, S., Ayyathan, D. M., and Siva, R. (2016). Induction of apoptosis by apocarotenoids in B16 melanoma cells through ROS-mediated mitochondrial-dependent pathway. *J. Funct. Foods* 20, 346–357. doi: 10.1016/j.jff.2015.11.019

Anantharaman, A., Subramanian, B., Chandrasekaran, R., Seenivasan, R., and Siva, R. (2014). Colorants and cancer: a review. *Ind. Crop Prod.* 53, 167–186. doi: 10.1016/j.indcrop.2013.12.025

Bo, Y., Wang, S., Ma, F., Manyakhin, A. Y., Zhang, G., Li, X., et al. (2023). The influence of spermidine on the build-up of fucoxanthin in Isochrysis sp. Acclimated to varying light intensities. *Bioresour. Technol.* 387, 129688. doi: 10.1016/j.biortech.2023.129688

Cañavate, J. P., Hachero-Cruzado, I., Pérez-Gavilán, C., and Fernández-Díaz, C. (2020). Lipid dynamics and nutritional value of the estuarine strain Isochrysis galbana VLP grown from hypo to hyper salinity. *J. Appl. Phycol.* 32, 3749–3766. doi: 10.1007/s10811-020-02258-2

Chen, C., Wang, Z., Ge, Y., Liang, R., Hou, D., Tao, J., et al. (2023). Characteristics prediction of hydrothermal biochar using data enhanced interpretable machine learning. *Bioresour. Technol.* 377, 128893. doi: 10.1016/j.biortech.2023.128893

Chin, G. J. W. L., Andrew, A. R., Abdul-Sani, E. R., Yong, W. T. L., Misson, M., and Anton, A. (2023). The effects of light intensity and nitrogen concentration to enhance lipid production in four tropical microalgae. *Biocatal. Agric. Biotechnol.* 48, 102660. doi: 10.1016/j.bcab.2023.102660

Chong, J. W. R., Tang, D. Y. Y., Leong, H. Y., Khoo, K. S., Show, P. L., and Chew, K. W. (2023). Bridging artificial intelligence and fucoxanthin for the recovery and quantification from microalgae. *Bioengineered* 14 (1), 2244232. doi: 10.1080/21655979.2023.2244232

Chu, J., Li, Y., Cui, Y., and Qin, S. (2019). The influences of phytohormones on triacylglycerol accumulation in an oleaginous marine diatom Phaeodactylum tricornutum. *J. Appl. Phycol.* 31, 1009–1019. doi: 10.1007/s10811-018-1623-y

Cui, H., Yang, F., and Li, Y. (2021). Exogenous methyl jasmonate enhances lipid production in Isochrysis galbana under nitrogen deprivation and high light. *Algal Res.* 58, 102406. doi: 10.1016/j.algal.2021.102406

de Haan, D., Ramos, N. H., Meng, Y. F., Rotkopf, R., Addadi, Y., Rosenhek-Goldian, I., et al. (2024). Decoupling cell size homeostasis in diatoms from the geometrical constraints of the silica cell wall. *New Phytol.* 243, 258–270. doi: 10.1111/nph.19743

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., et al. (2013). Orange: data mining toolbox in Python. *J. Mach. Learn. Res.* 14, 2349–2353.

Duval, J. F., Razafitianamaharavo, A., Bihannic, I., Offroy, M., Lesniewska, N., Sohm, B., et al. (2022). New insights into the effects of growth phase and enzymatic treatment on the cell-wall properties of Chlorella vulgaris microalgae. *Algal Res.* 69, 102955. doi: 10.1016/j.algal.2022.102955

Ferguson, J. N., Fernandes, S. B., Monier, B., Miller, N. D., Allen, D., Dmitrieva, A., et al. (2021). Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. *Plant Physiol.* 187, 1481–1500. doi: 10.1093/plphys/kiab346

Fierli, D., Aranyos, A., Barone, M. E., Parkes, R., and Touzet, N. (2022). Influence of exogenous phytohormone supplementation on the pigment and fatty acid content of three marine diatoms. *Appl. Microbiol. Biotechnol.* 106, 6195–6207. doi: 10.1007/s00253-022-12140-5

Fierli, D., Barone, M. E., Mc Donnell, A., Conlon, T., and Touzet, N. (2023). Combined application of exogenous phytohormones and blue light illumination to the marine diatom Phaeodactylum tricornutum. *Algal Res.* 71, 103052. doi: 10.1016/j.algal.2023.103052

Fiorentini, N., Pellegrini, D., and Losa, M. (2023). Overfitting prevention in accident prediction models: Bayesian regularization of artificial neural networks. *Transp. Res. Rec.* 2677, 1455–1470. doi: 10.1177/036119812211113

Foo, S. C., Khoo, K. S., Ooi, C. W., Show, P. L., Khong, N. M., and Yusoff, F. M. (2021). Meeting sustainable development goals: Alternative extraction processes for fucoxanthin in algae. *Front. Bioeng. Biotechnol.* 8. doi: 10.3389/fbioe.2020.546067

Foo, S. C., Yusoff, F. M., Ismail, M., Basri, M., Yau, S. K., Khong, N., et al. (2017). Antioxidant capacities of fucoxanthin-producing algae as influenced by their carotenoid and phenolic contents. *J. Biotechnol.* 241, 175–183. doi: 10.1016/j.jbiotec.2016.11.026

Gao, F., Sá, M., Teles, I., Wijffels, R. H., and Barbosa, M. J. (2021). Production and monitoring of biomass and fucoxanthin with brown microalgae under outdoor conditions. *Biotechnol. Bioeng.* 118, 1355–1365. doi: 10.1002/bit.27657

Hawrot-Paw, M., Koniuszy, A., Gałczyńska, M., Zając, G., and Szyszlak-Bargłowicz, J. (2019). Production of microalgal biomass using aquaculture wastewater as growth medium. *Water* 12, 106. doi: 10.3390/w12010106

Hudson, M., Martin, B., Hagan, T., and Demuth, H. B. (1992). Deep learning toolbox™ User's guide. *Tech. Rep.* 1992, 61.

Ishika, T., Laird, D. W., Bahri, P. A., and Moheimani, N. R. (2019). Co-cultivation and stepwise cultivation of Chaetoceros muelleri and Amphora sp. for fucoxanthin production under gradual salinity increase. *J. Appl. Phycol.* 31, 1535–1544. doi: 10.1007/s10811-018-1718-5

Jayakumar, S., Bhuyar, P., Pugazhendhi, A., Rahim, M. H. A., Maniam, G. P., and Govindan, N. (2021). Effects of light intensity and nutrients on the lipid content of marine microalga (diatom) Amphiprora sp. for promising biodiesel production. *Sci. Total Environ.* 768, 145471. doi: 10.1016/j.scitotenv.2021.145471

Ji, Y., Liu, Z., Cui, Y., Liu, R., Chen, Z., Zong, X., et al. (2024). Faba bean and pea harvest index estimations using aerial-based multimodal data and machine learning algorithms. *Plant Physiol.* 194, 1512–1526. doi: 10.1093/plphys/kiad577

JMP® (2017). *statistical software* (Cary, NC: SAS Institute Inc.), 1989–2023. Available at: https://www.jmp.com/en_in/home.html.

Kang, C. K., Shin, J., Cha, Y., Kim, M. S., Choi, M. S., Kim, T., et al. (2023). Machine learning-guided prediction of potential engineering targets for microbial production of lycopene. *Bioresour. Technol.* 369, 128455. doi: 10.1016/j.biortech.2022.128455

Khaw, Y. S., Yusoff, F. M., Tan, H. T., Noor Mazli, N. A. I., Nazarudin, M. F., Shaharuddin, N. A., et al. (2022). Fucoxanthin production of microalgae under different culture factors: a systematic review. *Mar. Drugs* 20, 592. doi: 10.3390/md20100592

Kumar, R. R., Sarkar, D., and Sen, R. (2024). Simultaneously maximizing microalgal biomass and lipid productivities by machine learning driven modeling, global sensitivity analysis and multi-objective optimization for sustainable biodiesel production. *App. Energy.* 358, 122597. doi: 10.1016/j.apenergy.2023.122597

Lei, C., Deng, J., Cao, K., Xiao, Y., Ma, L., Wang, W., et al. (2019). A comparison of random forest and support vector machine approaches to predict coal spontaneous combustion in gob. *Fuel* 239, 297–311. doi: 10.1016/j.fuel.2018.11.006

Li, N., Liu, Z., Wang, P., Suman, K., Zhang, J., and Song, Y. (2022). Effects of sodium hypochlorite treatment on the chlorophyll fluorescence in photosystem II of microalgae. *Sci. Total Environ.* 833, 155192. doi: 10.1016/j.scitotenv.2022.155192

Li, Y., Li, R., and Yi, X. (2020). Effects of light quality on growth rates and pigments of Chaetoceros gracilis (Bacillariophyceae). *J. Oceanol. Limnol.* 38, 795–801. doi: 10.1007/s00343-019-9171-0

Lim, Y. A., Ilankoon, I. M. S. K., Khong, N. M., Priyawardana, S. D., Ooi, K. R., Chong, M. N., et al. (2024). Biochemical trade-offs and opportunities of commercialized microalgae cultivation under increasing carbon dioxide. *Bioresour. Technol.* 393, 129898. doi: 10.1016/j.biortech.2023.129898

Lourenço-Lopes, C., Fraga-Corral, M., Jimenez-Lopez, C., Carpena, M., Pereira, A. G., García-Oliveira,, et al. (2021). Biological action mechanisms of fucoxanthin extracted from algae for application in food and cosmetic industries. *Trends Food Sci.* 117, 163–181. doi: 10.1016/j.tifs.2021.03.012

Manochkumar, J., Cherukuri, A. K., Kumar, R. S., Almansour, A. I., Ramamoorthy, S., and Efferth, T. (2023). A critical review of machine-learning for "multi-omics" marine metabolite datasets. *Comput. Biol. Med.* 165, 107425. doi: 10.1016/j.compbiomed.2023.107425

Manochkumar, J., Doss, C. G. P., Efferth, T., and Ramamoorthy, S. (2022). Tumor preventive properties of selected marine pigments against colon and breast cancer. *Algal Res.* 61, 102594. doi: 10.1016/j.algal.2021.102594

Manochkumar, J., and Ramamoorthy, S. (2024). Artificial intelligence in the 21st century: the treasure hunt for systematic mining of natural products. *Curr. Sci.* 126, 19–35. doi: 10.18520/cs/v126/i1/19-35

Market Reports World (2023). Global fucoxanthin market report – market reports world. Available online at: https://www.marketreportsworld.com/global-fucoxanthin-market-19888189. (accessed March 20, 2024).

Maulud, D., and Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *J. App. Sci. Tech. Trends* 1 (2), 140–147. doi: 10.38094/jastt1457

Mc Gee, D., Archer, L., Fleming, G. T., Gillespie, E., and Touzet, N. (2020). The effect of nutrient and phytohormone supplementation on the growth, pigment yields and biochemical composition of newly isolated microalgae. *Process Biochem.* 92, 61–68. doi: 10.1016/j.procbio.2020.03.001

Mc Gee, D., Archer, L., Parkes, R., Fleming, G. T., Santos, H. M., and Touzet, N. (2021). The role of methyl jasmonate in enhancing biomass yields and bioactive metabolites in Stauroneis sp. (Bacillariophyceae) revealed by proteome and biochemical profiling. *J. Proteomics* 249, 104381. doi: 10.1016/j.jprot.2021.104381

Miyashita, K., Beppu, F., Hosokawa, M., Liu, X., and Wang, S. (2020). Bioactive significance of fucoxanthin and its effective extraction. *Biocatal. Agric. Biotechnol.* 26, 101639. doi: 10.1016/j.bcab.2020.101639

Otálora, P., Guzmán, J. L., Acién, F. G., Berenguel, M., and Reul, A. (2021). Microalgae classification based on machine learning techniques. *Algal Res.* 55, 102256. doi: 10.1016/j.algal.2021.102256

Pisner, D. A., and Schnyer, D. M. (2020). "Support vector machine," in *Machine learning*, Ed. A. Mechelli and S. Vieira (Academic Press), 101–121.

Pyo, J., Hong, S. M., Jang, J., Park, S., Park, J., Noh, J. H., et al. (2022). Drone-borne sensing of major and accessory pigments in algae using deep learning modeling. *GISci. Remote Sens.* 59, 310–332. doi: 10.1080/15481603.2022.2027120

R Core Team. (2017). *R: A language and environment for Statistical computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/.

Raj, J. V. A., Kumar, R. P., Vijayakumar, B., Gnansounou, E., and Bharathiraja, B. (2021). Modelling and process optimization for biodiesel production from Nannochloropsis salina using artificial neural network. *Bioresour. Technol.* 329, 124872. doi: 10.1016/j.biortech.2021.124872

Rivera-Madrid, R., Carballo-Uicab, V. M., Cárdenas-Conejo, Y., Aguilar-Espinosa, M., and Siva, R. (2020). "Overview of carotenoids and beneficial effects on human health," in *Carotenoids: properties, processing and applications*, 1st ed. Ed. C. M. Galanakis (London, UK: Academic Press), 1–40.

Saini, D. K., Rai, A., Devi, A., Pabbi, S., Chhabra, D., Chang, J. S., et al. (2021). A multi-objective hybrid machine learning approach-based optimization for enhanced biomass and bioactive phycobiliproteins production in Nostoc sp. CCC-403. *Bioresour. Technol.* 329, 124908. doi: 10.1016/j.biortech.2021.124908

Sequeira, R. A., Mondal, D., and Prasad, K. (2021). Neoteric solvent-based blue biorefinery: for chemicals, functional materials and fuels from oceanic biomass. *Green Chem.* 23, 8821–8847. doi: 10.1039/D1GC03184H

Sun, Z., Wang, X., and Liu, J. (2019). Screening of Isochrysis strains for simultaneous production of docosahexaenoic acid and fucoxanthin. *Algal Res.* 41, 101545. doi: 10.1016/j.algal.2019.101545

Tang, D. Y. Y., Chew, K. W., Ting, H. Y., Sia, Y. H., Gentili, F. G., Park, Y. K., et al. (2023). Application of regression and artificial neural network analysis of Red-Green-Blue image components in prediction of chlorophyll content in microalgae. *Bioresour. Technol.* 370, 128503. doi: 10.1016/j.biortech.2022.128503

Thiviyanathan, V. A., Ker, P. J., Tang, S. G. H., Amin, E. P., Yee, W., Hannan, M. A., et al. (2024). Microalgae biomass and biomolecule quantification: optical techniques, challenges and prospects. *Renew. Sustain. Energy Rev.* 189, 113926. doi: 10.1016/j.rser.2023.113926

Victor, P., and Camarena-Bernard, C. (2023). Lutein, violaxanthin, and zeaxanthin spectrophotometric quantification: A machine learning approach. *J. Appl. Phycol.* 35, 73–84. doi: 10.1007/s10811-022-02855-3

Wang, L. J., Fan, Y., Parsons, R. L., Hu, G. R., Zhang, P. Y., and Li, F. L. (2018). A rapid method for the determination of fucoxanthin in diatom. *Mar. Drugs* 16, 33. doi: 10.3390/md16010033

Wu, M., Cao, J., Xu, Y., Xu, S., Zhou, Z., Shao, M., et al. (2023). Impact of three phycospheric bacterial strains on the growth and fatty acid composition of Isochrysis galbana. *Algal Res.* 74, 103183. doi: 10.1016/j.algal.2023.103183

Yadav, K., Nikalje, G. C., Pramanik, D., Suprasanna, P., and Rai, M. P. (2023a). Screening of the most effective media for bioprospecting three indigenous freshwater microalgae species. *Int. J. Plant Biol.* 14, 558–570. doi: 10.3390/ijpb14030044

Yadav, I., Rautela, A., Gangwar, A., Wagadre, L., Rawat, S., and Kumar, S. (2023b). Enhancement of isoprene production in engineered Synechococcus elongatus UTEX 2973 by metabolic pathway inhibition and machine learning-based optimization strategy. *Bioresour. Technol.* 387, 129677. doi: 10.1016/j.biortech.2023.129677

Yang, F., Liu, W., Li, Y., Che, X., and Liu, S. (2024). Glycerol changes the growth and lipid profile of the marine microalga Isochrysis galbana via the regulation of photosynthetic and respiratory metabolic pathways. *J. Appl. Phycol.* 36, 57–71. doi: 10.1007/s10811-023-03148-z

Yeh, Y. C., Syed, T., Brinitzer, G., Frick, K., Schmid-Staiger, U., Haasdonk, B., et al. (2023). Improving microalgae growth modeling of outdoor cultivation with light history data using machine learning models: A comparative study. *Bioresour. Technol.* 390, 129882. doi: 10.1016/j.biortech.2023.129882

Yu, T., Fan, F., Huang, L., Wang, W., Wan, M., and Li, Y. (2024). Artificial neural networks prediction and optimization based on four light regions for light utilization from Synechocystis sp. PCC 6803. *Bioresour. Technol.* 394, 130166. doi: 10.1016/j.biortech.2023.130166

Yusof, Z., Khong, N. M., Choo, W. S., and Foo, S. C. (2022). Opportunities for the marine carotenoid value chain from the perspective of fucoxanthin degradation. *Food Chem.* 383, 132394. doi: 10.1016/j.foodchem.2022.132394

Zienkiewicz, A., Zienkiewicz, K., Poliner, E., Pulman, J. A., Du, Z. Y., Stefano, G., et al. (2020). The microalga Nannochloropsis during transition from quiescence to autotrophy in response to nitrogen availability. *Plant Physiol.* 182, 819–839. doi: 10.1104/pp.19.00854