Author(s): Pääkkönen, Salli; Pölönen, Ilkka; Calderini, Marco; Yli-Tuomola, Aliisa; Ruokolainen, Visa; Vihinen-Ranta, Maija; Salmi, Pauliina

Title: Lipid monitoring of Chlorella vulgaris using non-invasive near-infrared spectral imaging

Year: 2024

Version: Published version

Please cite the original version:

Pääkkönen, S., Pölönen, I., Calderini, M., Yli-Tuomola, A., Ruokolainen, V., Vihinen-Ranta, M., & Salmi, P. (2024). Lipid monitoring of Chlorella vulgaris using non-invasive near-infrared spectral imaging. Journal of Applied Phycology, Early online. https://doi.org/10.1007/s10811-024-03397-6

Journal of Applied Phycology
https://doi.org/10.1007/s10811-024-03397-6

RESEARCH

Lipid monitoring of *Chlorella vulgaris* using non-invasive near-infrared spectral imaging

Salli Pääkkönen[1] · Ilkka Pölönen[1] · Marco Calderini[1] · Aliisa Yli-Tuomola[1] · Visa Ruokolainen[2] · Maija Vihinen-Ranta[2] · Pauliina Salmi[1]

Received: 13 September 2024 / Revised: 15 November 2024 / Accepted: 18 November 2024

Abstract

Microalgal lipids are molecules of biotechnological interest for their application in sustainable food and energy production. However, lipid production is challenged by the time-consuming and laborious monitoring of lipid content in microalgae. This study aimed to predict the lipid content of *Chlorella vulgaris* cultivations based on non-invasively collected near-infrared (NIR) range hyperspectral data. A gravimetric analysis of total lipids was used as reference data (between 2 and 22% per dry weight) to compare three different models to determining the lipid content. A one-dimensional convolutional neural network and partial least squares models performed at a similar level. Both models could predict the lipid content of *Chlorella* dry weight with an error of 4%pt (root mean squared error). The index-based linear regression model performed the weakest of the three models, with the error of the prediction being 6%pt. Nile Red staining was used to visualise lipids on a microscope and lipid class analysis to resolve the lipid classes that explained most of the increase in lipids in *Chlorella*. A SHAP algorithm (SHapley Additive exPlanations) was used to analyse the wavebands of NIR spectra that were important for predicting the total lipid content. The results show that spectral data, when combined with an adequate algorithm, could be used to monitor microalgae lipids non-invasively in a closed system, in a way that has not previously been demonstrated with an imaging system.

Keywords Hyperspectral imaging · Microalgae · Lipid content · Machine learning · Nile Red staining

Introduction

Microalgae are single-celled organisms that can fix carbon dioxide and produce significant amounts of molecules important for consumers, such as lipids (Müller-Navarra et al. 2000; Patel et al. 2022). Lipids are hydrophobic molecules that include triglycerides, glycolipids, phospholipids and sterols, which play multiple roles in cell metabolism, storage, structure, and other functions. As environmentally responsible technologies advance, the cultivation of microalgae for commercial purposes is becoming more and more popular (Bellou et al. 2014; Fernández et al. 2021). The lipids produced by microalgae have been exploited both in human food, such as in food supplements, and in animal feed as, for example, part of fish feed (Ansari et al. 2021; De Bhowmick et al. 2023). Microalgae have also attracted interest from a bioenergy perspective, as the lipids produced by microalgae could be used as raw materials for biofuels (Fernández et al. 2021). Compared to other raw materials of biofuels, such as soybean and oil palm, microalgae cultivation requires less land area (Lam and Lee 2012). Considering the ability of microalgae to fix carbon dioxide, microalgal biotechnology is a promising area for food production and sustainable energy development.

To ensure the quality of microalgal products, it is important to monitor algae cultivations (Havlik et al. 2022). Some of the most commonly used methods to quantify the lipid content of microalgae are gravimetric methods (Challagulla et al. 2017). They are accurate and frequently used, but they are also time-consuming and require physical manipulation of the cultures, trained personnel, and laboratory conditions (Challagulla et al. 2017). In addition, the solvents used in the protocols are toxic. Another commonly used method to

✉ Salli Pääkkönen
salli.k.paakkonen@jyu.fi

[1] Faculty of Information Technology, University of Jyväskylä, Mattilanniemi 2, P.O. Box 35, 40014 Jyväskylä, Finland

[2] Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

Published online: 04 December 2024

determine the lipid content of microalgae is Nile Red staining. Nile Red is a lipophilic dye that can be used to quantify especially the neutral lipid fraction in cells when it is combined with fluorescence detection (Chen et al. 2009; Zheng et al. 2022). The challenge with this method is that it does not work with all microalgal species due to the presence of thick cell walls, which has led to the use of different solvents being tested (Storms et al. 2014). Although solvents facilitate the staining of lipids and enhance their signal, this multi-step method is time-consuming and allows only a small volume of the cultivation to be viewed. More efficient monitoring methods have been called for to meet the needs of industrial algae cultivation, which requires reliable and rapid monitoring in various volumetric scales (Liu et al. 2020; Havlik et al. 2022; Solovchenko 2023).

Spectroscopic methods utilising absorbance or reflectance signals on the near-infrared (NIR; 780–2500 nm) portion of the light spectra have been proposed for detecting the lipid content of microalgae, because lipids contain functional groups (e.g. CH) with signals in the NIR region caused by molecular vibrations (dos Santos et al. 2013). Spectroscopic methods have increased in popularity because of their increased practicality, portability, and sensitivity. For example, Liu et al. (2015) analysed the total fatty acids of microalgae cultivation by detecting the diffusively reflected radiation of liquid samples with a portable NIRS spectrometer. A challenge for NIR signal-based monitoring, especially in liquid cultures, is posed by water, which absorbs strongly in the NIR region. Brown et al. (2014) avoided the absorption property of water by determining the lipid content from filtered microalgae samples using a portable NIR reflectance spectrometer.

A hyperspectral imager produces a stack of images, also called a data cube, where the *x*- and *y*-axes contain spatial information, and the *z*-axis displays spectral information on hundreds of different wavebands, enabling the collection of detailed spectral information. Spectral imagers have been developed to detect information from the visual range (380–700 nm) but also NIR ranges. Li et al. (2020) investigated the use of a spectral imager, that is, imaging spectroscopy, in lipid monitoring. They aimed to reduce the need for sample processing and to image liquid samples despite the absorption of water in the NIR region. They were able to predict the production, as well as spatial accumulation, of microalgal lipids from living liquid samples by imaging algae with an NIR hyperspectral imager using a transmission setup. They overcame the water's strong absorption by correcting each image with spectra of the growth medium.

Previous studies on lipid monitoring of microalgae using spectroscopic methods have focused on the analysis of filtered and liquid microalgal samples (Brown et al. 2014; Liu et al. 2015; Li et al. 2020). This study explores the potential of a spectral imager to monitor the lipids of microalgae cultivation non-invasively, that is, without sampling, on a laboratory scale. Spectral imaging is a potential method with many features that can be useful on a larger scale. In addition to its non-invasiveness, the advantages of spectral imaging include the spatial information it provides, its low cost, and operating speed (Dierssen et al. 2021). For the above reasons, the development of spectral imaging–based microalgae monitoring methods could meet the requirement for fast and efficient monitoring in large-scale microalgae cultivation.

A challenge when imaging an algae cultivation is that light propagation is affected by the arrangement of microalgal cell wall composition and microalgal growth style, as well as the growth media (Bricaud et al. 1988; Mehrubeoglu et al. 2013). Therefore, linear calibration models might not be the most expedient for monitoring biological phenomena. In particular, lipid accumulation can be non-linear concerning the growth of microalgae. This is because the accumulation of lipids in microalgae is usually highest at the point when carbon is not used for growth (i.e., cell division). This could be under unfavourable or stressful conditions (Hu et al. 2008). The challenge with non-invasive imaging is that NIR radiation penetrates the sample to a maximum depth of a few centimetres (Beć et al. 2021) and high biomass can further block the passage of light through the culture. In this case, the measurement of absorbance, often used for example in a spectrophotometer where light passes through the sample (Beć et al. 2021), is not possible. An alternative is to image the culture in a reflectance setup (Pääkkönen et al. 2024). Furthermore, in NIR spectroscopy, interpretation of the spectra is challenging as the signals from the lipid functional groups have broad spectral shapes and may overlap with each other (Beć et al. 2021).

Vegetation indices, the ratio between two or more wavelengths, are a simple way of distinguishing certain features in the spectrum. In spectral applications, the indices can be fixed, such as in the Normalised Difference Vegetation Index (NDVI) based on NIR and red wavebands, which is a common method for monitoring vegetation (Huang et al. 2021). The multiple wavelengths of the spectral imager also make it possible to find a more specific index to correlate with the desired factor (Salmi et al. 2021), which can be used for modelling. Such a method has been used, for example, in biomass prediction (Pääkkönen et al. 2024). Partial least squares (PLS) regression has also been successfully used in spectroscopic studies (Li et al. 2020; Liu et al. 2020), as spectra have variables that often correlate with each other. PLS first looks for factors that are not correlated with each other but are related to the outcome of the analysis. The analysis is then performed with these components. PLS regression has been used, for example, to predict lipid content (Brown et al. 2014), as well as biomass content (Martínez-Guijarro et al. 2018) based on spectral data.

The increase in computing capacity has enabled the generalisation of various machine learning algorithms for the analysis of spectral data. The algorithms can predict an outcome once they have been trained with separate training data. Convolutional neural networks (CNN) have become more popular in recent years as they can be trained to solve complex non-linear problems. The model filters out factors that are irrelevant to the prediction and emphasises the factors it considers most important if trained successfully. CNN can solve both classification and regression problems: it has been used to classify microalgal species from microscopic images (Pant et al. 2020; Yadav et al. 2020) but also to classify species and predict their biomass content in laboratory cultivations (Pääkkönen et al. 2024). Neural networks have been criticised as black boxes, because it is not known on what basis they form a prediction (Sheu 2020). Nowadays, explanatory algorithms have gained popularity, helping to identify which factors have had the greatest impact on the model's output.

In this study, the aim was to predict the total lipid content of microalgal cultivation from non-invasively collected NIR range reflectance-like spectra (see details in methods) using a hyperspectral imager. This study compared the ability of three algorithms – index-based model, PLS, and CNN – for lipid monitoring. Their ability to predict the total lipid content (% per dry weight) of microalgae despite non-linearity introduced by varying biomass concentrations was tested. In addition, the most important areas of the NIR spectrum for lipid prediction were investigated.

Methods

Chlorella vulgaris (CCAP 211/11B) was chosen for the study because it is a valuable microalga for mass cultivation: it is fast growing, can produce lipids efficiently (Sakarika and Kornaros 2017; Wong et al. 2017), and has a high ability to fix carbon dioxide (Senatore et al. 2021). *Chlorella* is also a widely used species in feed and food production (Ahmad et al. 2020; Coronado-Reyes et al. 2020). To encourage *Chlorella* to produce different lipid concentrations in various biomass concentrations, three different experiments were conducted. In addition to autotrophic cultivation, the cultures were encouraged to produce lipids by limiting the amount of nitrogen (N) or sulphur (S) and by varying the light intensity (Fig. 1), because the limitation of the above factors has been shown to affect the storage lipid production of *Chlorella* (Sakarika and Kornaros 2016, 2017). Lipid production was also encouraged by mixotrophic growth following glucose addition in some cultures as shown in Sakarika and Kornaros (2017).

The biomass concentration of the cultivations was assessed to analyse the cultivation density. Gravimetric quantification of total lipids was used as the ground truth assessment for NIR imaging. To visualise the algal lipids, Nile Red staining was used on a subset of cultures, followed by imaging with a confocal fluorescence microscope. Lipid class analysis was done to provide additional information on the lipids present in the cultivations. Index-based LR, PLS, and 1D CNN models were tested for lipid monitoring. Lastly, a SHAP algorithm (SHapley Additive exPlanations) was used to analyse which areas of the spectra were important for the model prediction.

Cultivation

The objective of the multiple cultivations was to obtain a wide range of lipid concentrations in different biomass concentrations and hence not all cultures were treated equally. *Chlorella* was cultivated in batches of 3 to 4 replicates for 3 to 15 days using 650-mL tissue culture flasks with filter caps (VWR international) with a starting volume of 400 mL or 1L Erlenmeyer with a volume of 500 to 1000 mL. In the standard experiment, cultivation occurred in autotrophic conditions using MWC (Modified Wright's Cryptophyte) media (Guillard and Lorenzen 1972). The cultivations were maintained in FH-130 growth chambers (HiPoint, Taiwan) at 22 ± 1 °C under continuous illumination using fluorescent lamps with a light intensity of 115 to 145 µmol photons m^{-2} s^{-1} measured with a quantum sensor (HiPoint, Taiwan). To provide sufficient inorganic carbon, 1 mL of inorganic carbon $NaHCO_3$ was added to each flask. In the nitrogen experiment, *Chlorella* was cultivated in the standard conditions (control treatment) and in N-depleted MWC (N-treatment). In the sulphur experiment, the light intensity of the growth chambers was reduced to 12–18 µmol photons m^{-2} s^{-1}, and a glucose solution was added to the MWC, the final concentration being 4.8 g L^{-1} (control treatment). For the sulphur treatment, S-depleted MWC was used and a glucose solution was added, with the final concentration being 10 g L^{-1} (Sakarika and Kornaros 2017). The flasks were covered with aluminium foil.

For the assessments of wet biomass, gravimetric lipid assessment, Nile red staining, and lipid class analysis, 50-mL samples from all cultivations were taken aseptically inside a laminar flow hood at the beginning and/or at the end of the cultivation period, right before the spectral imaging.

Wet biomass assessment

Wet biomass was assessed with an electronic cell counter (CASY, Omni Life Sciences, Germany) to obtain additional information from the cultivation growth. A 60-µm capillary was used with *Chlorella* based on its cell size. Depending on the culture density, 10 to 40 µL of the sample was added to 10 mL of CasyTon buffer. The results were processed

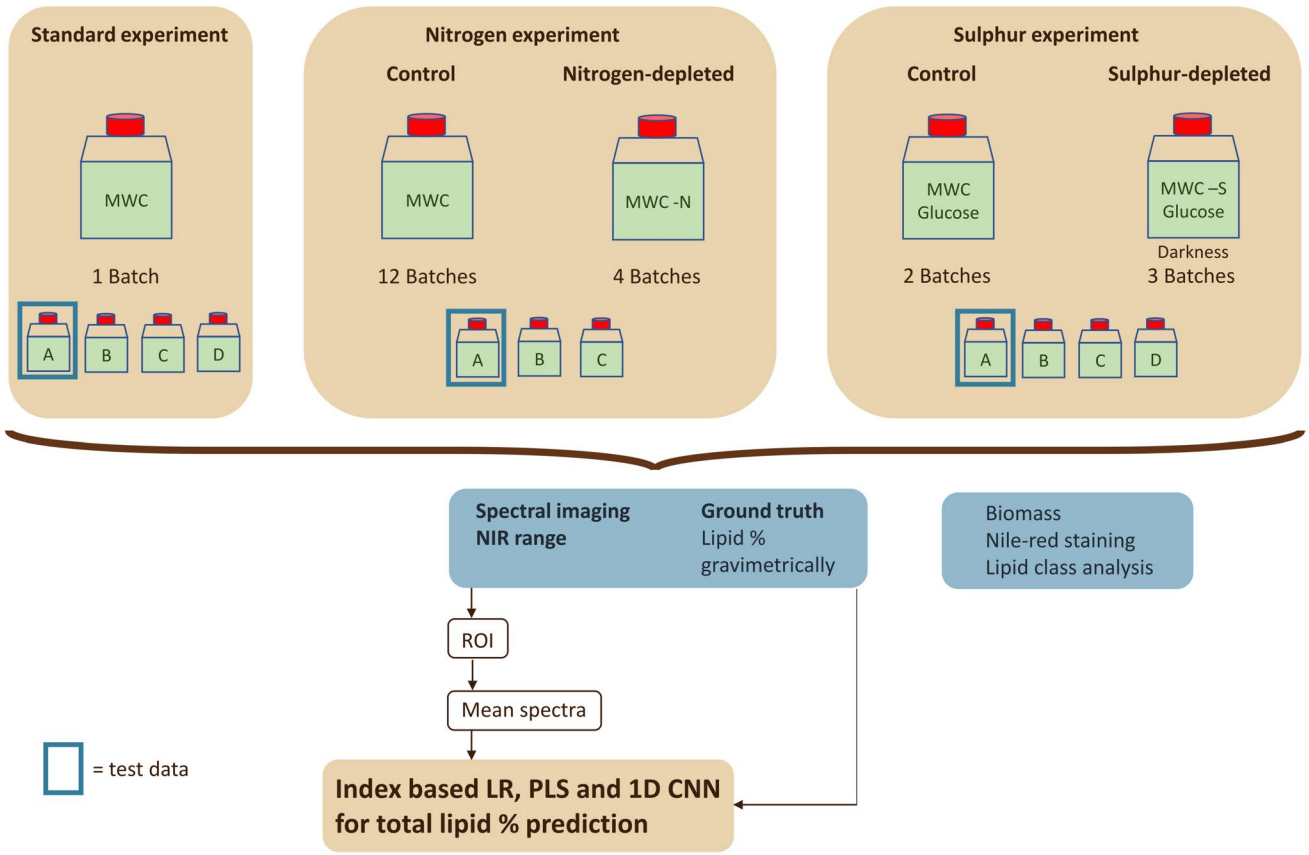

Fig. 1 Up-to-down workflow of the study. *Chlorella vulgaris* was cultivated in three different experiments to test the total lipid monitoring with a near-infrared (NIR) spectral imager. ROI stands for region of interest in the image. Index-based linear regression (LR), partial least squares (PLS) and one-dimensional convolutional neural network (1D CNN) models were tested for total lipid monitoring

with the CASY workX 1.26 macro (Omni Life Sciences) for Microsoft Excel. Microalgae biovolumes (fL mL^{-1}) were converted to wet biomass (mg mL^{-1}) by assuming that the cells are isopycnic to water. Wet biomasses were not used for lipid content analysis, which was done using dry biomasses, as described below.

Lipid ground truth

Between 40 and 50 mL of culture were centrifuged at $4696 \times g$ for 10 min. Supernatants were discarded and pellets were stored at –80 °C. Algal biomass was freeze-dried overnight, after which 0.15 to 1.5 mg were used for total lipids extraction. Extractions were carried out with chloroform:methanol:water (4:2:1) using sonication (10 min) in a KIMAX glass tube. Centrifugation (2000 RPM for 5 min) was then used to facilitate phase separation. The lipid-rich fraction was transferred to a new tube and solvents were evaporated at room temperature under an N_2 stream. The lipid extract was resuspended in 100 µL of chloroform and transferred to a pre-weighted smooth-wall tin cup (D4057 Elemental Microanalysis). To ensure no lipids were left in the KIMAX glass tube, 100 µL of methanol was added, tubes were vortexed, and methanol was transferred to the same smooth-wall tin cup. Once the tin cups were completely dry, they were weighed and the difference from the initial weighing was used to calculate total lipid mass (mg). Lipid content (%) was calculated as the fraction of lipid mass to algal dried biomass. If the cultivation biomass was too low (sample dry weight less than 0.8 mg), replicates from a given batch were pooled, or the whole batch was discarded and not included in the study. The final number of samples was 104.

Spectral imaging

All *Chlorella* cultivation flasks were imaged with an NIR hyperspectral imager Specim FX17 (Specim, Finland) to test the monitoring of the microalgae lipid content. Apart from the imager model and waveband range, the imaging setup is identical to Pääkkönen et al. (2024). The FX17 imager used here has a spectral range of 950 to 1700 nm and a spectral resolution FWHM (full width at half maximum) of 8 nm

(mean). NIR images were taken at the beginning and/or at the end of each cultivation period.

The imaging setup consisted of a broad-band halogen light source (3 bulbs of DECOSTAR 51 ALU 20W 12 V 36 deg GU5.3 halogen) and a white reference (PTFE diffuse reflector sheet PMR10P1, Thorlabs) (Fig. 2). The imager and the light source were placed on a motorised scanner (LabScanner 40 × 20, Specim, Finland). Laser shields with square engravings were placed to adjust focus and scanning speed. The scanner and imager were operated through Lumo Scanner software by Specim (Finland). The selected scanning speed was 2.5 mm s^{-1}, with a frame rate of 20 fps and an exposure time of 15 s. For simplicity, the term *reflectance* will be used in this article. The distance between the imager's lens and the target was fixed at 16 cm, while the distance between the imager's lens and the white reference was 25 cm. Both white and black references were imaged for each image separately. The black reference was taken automatically, with the mechanical shutter of the imager closing before each image. The images were taken in a darkroom to avoid any specular reflections or stray light. This setup is a reflectance-like setup because when the white reference is placed behind the target algae, light is also reflected from the white reference, and therefore the phenomenon to be measured is more complex.

The reflectance (R) of the images were calculated using Eq. 1:

$$R = \frac{I - I_d}{I_w - I_d} \quad (1)$$

where I is the irradiance from the cultivation, I_d the black reference, and I_w irradiance of the white reference, both of which were taken in conjunction with each image.

The water in the cultivation media affects the absorption characteristics of the algal culture, especially in the NIR range. The plastic of the culture flasks also has strong absorption properties in the NIR region. Therefore, before each batch was imaged, an image of the flask containing only the growth media (MWC) was also taken. These images were used to normalise the cultivation images to extract the algal signal. The volume of the MWC flask was adjusted to the volume of the imaged algae flask.

In the sulphur experiment in this study, the cultures reached the highest biomass. To test whether the model could interpret the higher biomass signal as indicating more lipids, one batch of the sulphur experiment was diluted first 1:1 by replacing 200 mL of the culture with phosphate-buffered saline (PBS). This would provide data from cultivations with low biomass and high lipid content (See supplementary Fig. 1 for mean spectra of the dilutions). A second dilution was made in the same way, resulting in 1:4 dilution from the original concentration. The flasks were imaged between the dilutions. The total lipid content (% per dry weight) of the cells was expected to stay the same between the dilutions because the dilutions were imaged within two hours.

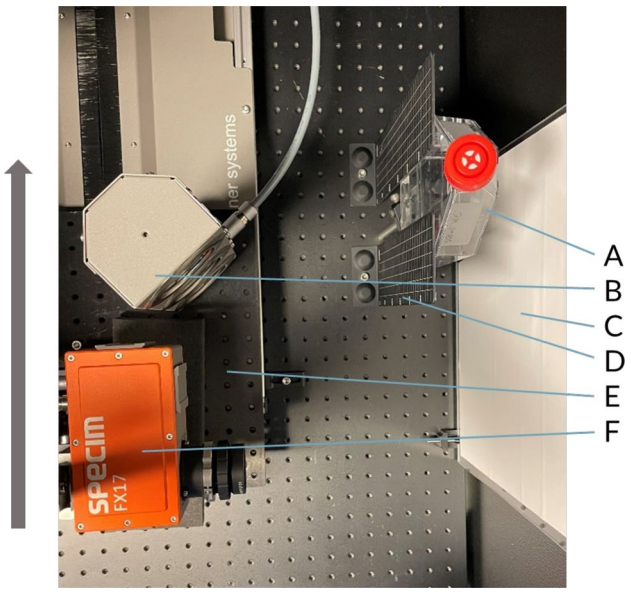

Fig. 2 Imaging setup, top view. **A**: target, **B**: light source, **C**: white reference, **D**: laser shields, **E**: computer-guided scanner mounting the imager and the light source, **F**: spectral imager. The black arrow outside the panel marks the direction of movement of the scanner. The setup is identical to Pääkkönen et al. (2024) with a different imager

Confocal microscopy and lipid staining

Storage lipids of *Chlorella* cells were visualised from representative samples of each experiment with Nile Red and imaged with a confocal fluorescence microscope. The target for the cell density was 10^8–10^7 cells mL^{-1} for imaging. A mixture of 10 µL algae, 10 µL Nile Red stain (10 µg mL^{-1}), and 80 µL ethanol (30% v/v) was used for imaged samples (Storms et al. 2014). Negative control images were taken using 10 µL algae and 90 µL alcohol (30%). Aliquots of 100 µL were pipetted on microscopy plates (1 µ + Slide 8 well ibiTrear, Ibidi, Germany) and cells were allowed to settle in the dark, at room temperature, for 45 min before imaging.

The algae were imaged using a Nikon A1R confocal microscope and NIS-Elements AR version 5.21.03. The algae were observed through a Plan Apo 60X 1.2 NA water immersion objective using 561 excitation wavelength and 595/50 emission filter and DIC contrast. This procedure produced 512 × 512 pixel images, with a pixel size of 110 nm (x, y). From each sample well, 4 to 6 randomly selected areas were imaged with the aim of visualising the overall fluorescence intensity. The imaging of different samples was made using the same laser intensity and detector settings for intensity-comparable data. The presented images were

processed in FIJI using the same linear intensity function adjustments to retain the intensity comparability between the samples.

Lipid class analysis

To scrutinise which lipid classes were mainly responsible for the variation in the lipid content of *Chlorella* in the different treatments, lipid classes were analysed from 5 to 12 samples of each treatment by VTT Technical Research Centre of Finland. The analysis was conducted in line with the study by Kotapati and Bates (2020).

Models for lipid monitoring

Spectral data processing

NIR images were cropped to 200 × 200 pixel ROIs. The spectra were truncated between 974 and 1378 nm (channels 10–125) to cut off the flat and noisy tails of the spectra likely caused by a low signal-to-noise ratio resulting from the high absorbance by water. All spectral images were corrected by dividing them with the corresponding MWC image to overcome the strong absorption characteristics of water and to separate the microalgae signal. The first replicates of each batch were used to test the model ($n = 34$) and the rest of the replicates to train the model ($n = 70$). Mean spectra were calculated from the test and training data images, and each spectrum was min–max normalised by the spectra's minimum and maximum values (spectral-wise) to highlight differences in spectral shape. All spectral data were processed using Python Version 3.9.7 in Jupyter Notebook.

Index-based linear regression

A vegetation index-based model was tested to predict the lipid content of microalgae. The best index between two wavelengths was formed with training data, in line with Pääkkönen et al. (2024). The vegetation index model was the only model in which the spectra were not min–max normalised (Supplementary Fig. 2A), as this would have resulted in a division by zero. The best index was the one with the highest Pearson's correlation with the total lipid content (% per dry weight) used as ground truth. A linear regression (LR) model was fitted between the best index and ground truth in Microsoft Excel version 2403 (See LR model in Supplementary Fig. 3). The vegetation index-based LR model was used to predict the total lipid content of separate test data.

Partial least squares

Partial least squares (PLS) regression was tested for the microalgae lipid content monitoring. The input for the PLS was min–max normalised mean spectra (Supplementary Fig. 2B) with corresponding lipid content (% per dry weight), rounded to an integer as labels for each spectrum. The PLS algorithm automatically creates a new dataset from the factors that do not correlate with each other but which are related to the outcome of the model – in this case, total lipid prediction – and the analysis is then performed with these components. The number of principal components (PC) was determined with training data using cross-validation with ten folds. The number of PCs that produced the minimum mean squared error (MSE), maximum residual predictive deviation (RPD), and maximum coefficient of determination (R^2) was chosen for the model. Based on previous factors, the model with six PCs was chosen (Fig. 3). Separate test data were used to test the model.

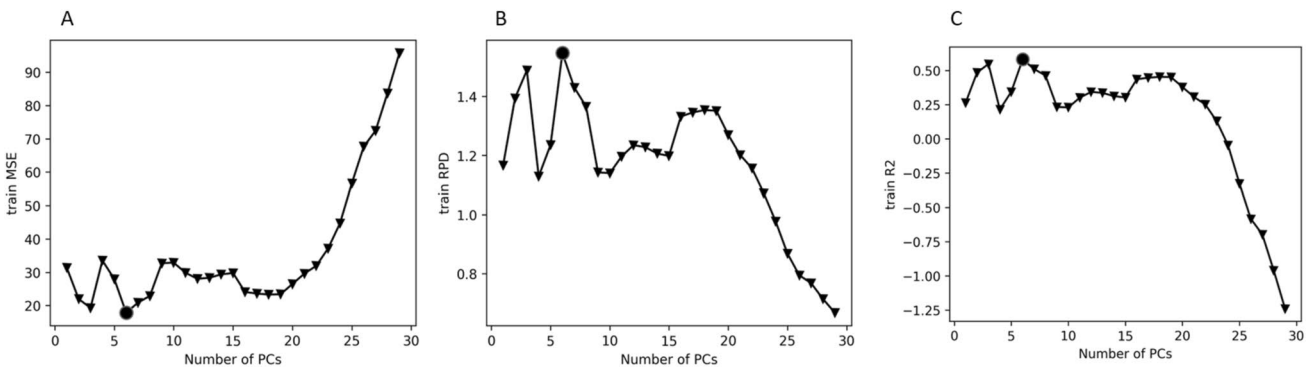

Fig. 3 Selection of the number of principal components (PC) of the partial least squares (PLS) model based on (**A**) the minimum mean squared error (MSE), (**B**) the maximum residual predictive deviation (RPD), and (**C**) the maximum coefficient of determination (R^2) of training data using cross-validation. The circle in the figure shows the best number of PCs

One-dimensional convolutional neural network

A one-dimensional convolutional neural network (1D CNN) was constructed to test microalgae lipid content monitoring. The model was implemented with a Keras library and Tensorflow backend, and computed on a Nvidia Tesla V100-SXM2 16 GB GPU unit. The input for the 1D CNN was min–max normalised mean spectra (Supplementary Fig. 2C) with corresponding lipid content (% per dry weight) rounded to an integer as labels for each spectrum.

The training data (200 × 200 pixel images) were divided into 100 × 100 pixel images for data augmentation. The total lipid content (% per dry weight) (i.e. ground truth) was assumed to remain the same in all smaller parts of the image as the flasks were shaken carefully before imaging. The training data was split into training and validation data by randomly dividing 80% as training and 20% as validation data (sklearn). To train the model efficiently, random oversampling was used to balance the differences between the lipid content and the corresponding mean spectra (Supplementary Fig. 4). This was done for training and validation data separately and resulted in a training data set of 648 and a validation data set of 140. Mean spectra from 200 × 200 pixel areas were used as test data.

To make a robust model, the 1D CNN was first built with the simplest possible architecture. After that, it was tested if the model prediction improves when layers are added to the architecture. Because validation loss and validation RMSE grew as the layers were added, the simplest architecture (ID 1 in Table 1, Fig. 4) was chosen as the best model. Max pooling layers with pool size 2 were added after each convolutional layer and a dropout layer with 0.2 drops was added after the first convolutional layer and before the output layer to control overfitting. A rectified linear unit (ReLu) was selected as an activation function for the convolutional and dense layers. The model was optimised through a gradient-based stochastic optimiser (Adam) with a learning rate of 0.001. Models were trained for 50 epochs in a sample batch size of 32. The best model was trained three times, and the results are means of the triplicate training (See learning curves in Supplementary Fig. 5).

Comparison of models

The coefficient of determination (R^2) and the test root mean squared error (RMSE) were used to evaluate the performance of the index-based LR, PLS, and 1D CNN models. The factors were defined in Eqs. 2 and 3:

$$R^2 = \frac{\sum_{i=1}^{n}\left(exp_i - pred_i\right)^2}{\sum_{i=1}^{n}\left(exp_i - \bar{y}\right)^2} \quad (2)$$

Table 1 Testing the architectures of the one-dimensional convolutional neural network (1D CNN) model for lipid monitoring. Convolution filter counts in convolution blocks and node counts in dense blocks are separated with "/" and "do" = dropout layer of 0.2

ID	Convolution block	Dense block	Train loss	Train RMSE	Validation loss	Validation RMSE	Validation Cor
1	256 + do	256	12.13	3.48	12.90	3.59	0.84
2	256 + do/128	256	17.73	4.20	22.33	4.73	0.84

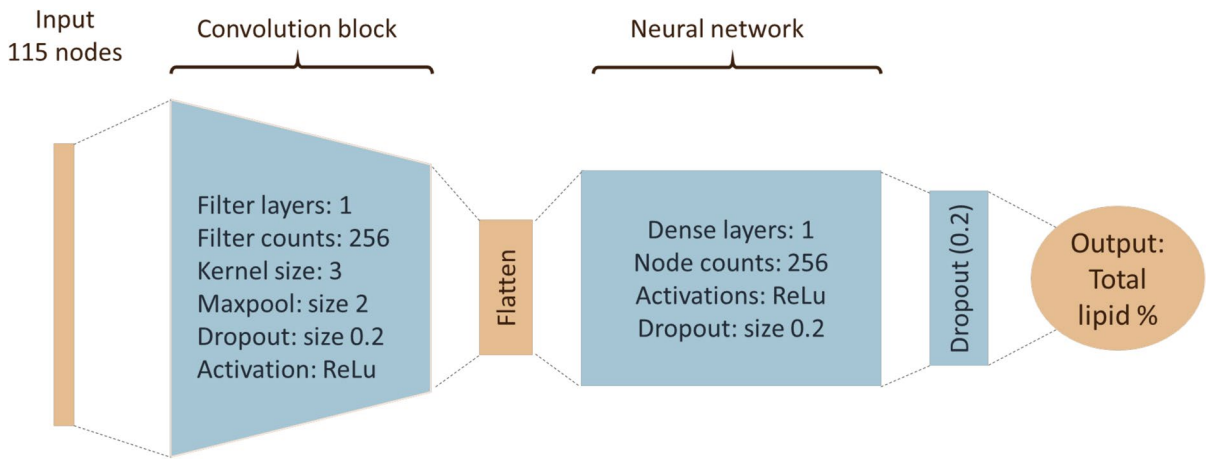

Fig. 4 The architecture of the 1D CNN used for total lipid content (% per dry weight) prediction

$$RMSE, \%pt = \sqrt{\frac{1}{n}\sum_{i=1}^{n-1}(exp_i - pred_i)^2} \quad (3)$$

where n = number of samples, exp = expected value, $pred$ = predicted value, and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} exp_i$.

Results

Near-infrared spectra of the cultivations

The mean spectra of 200 × 200 pixel ROIs of *Chlorella* cultivations in three different experiments are shown in Fig. 5. The spectra of all experiments show two clear absorption valleys: a smaller valley around 980 nm and a clearer valley around 1150 nm. The clearest reflection peaks are between 1000 and 1130 nm. All spectra become noisier after 1150 nm.

Assessed total lipid and biomass content

The wet biomass of the *Chlorella* cultures assessed with the electronic cell counter ranged between 0.05 and 3.50 mg L^{-1} and the gravimetrically assessed lipid content between 2 and 22% per dry weight. The biomass and the lipid content did not increase linearly (Fig. 6). The S-control cultivations had the highest average biomass as well as the greatest variation in the biomass compared to other experiments. The S-depleted cultivations, on the other hand, had the highest total lipid content on average. In the nitrogen experiment, a similar phenomenon can be observed as in the sulphur experiment: the N-depleted cultivations produced on average

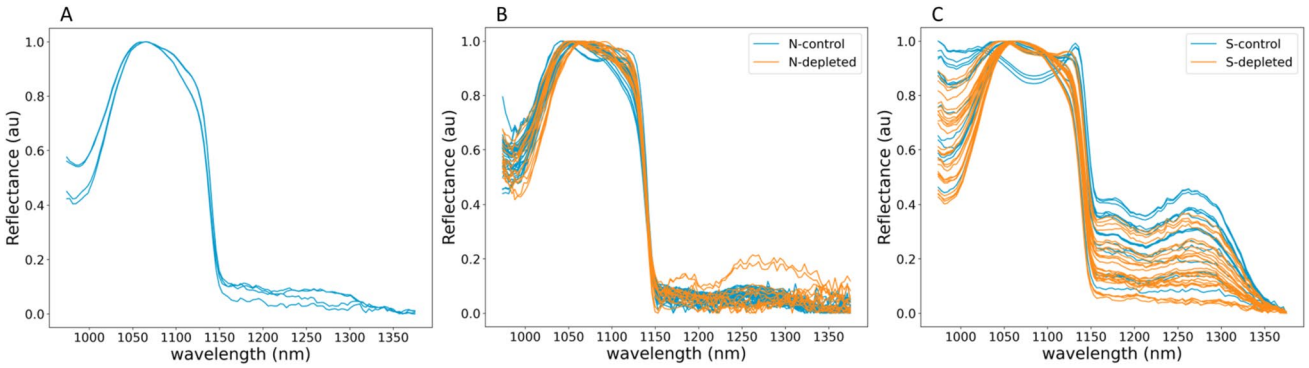

Fig. 5 Min–max normalised mean spectra of *Chlorella vulgaris* from three different experiments. (**A**) Standard, (**B**) Nitrogen, and (**C**) Sulphur, from 200 × 200 pixel areas

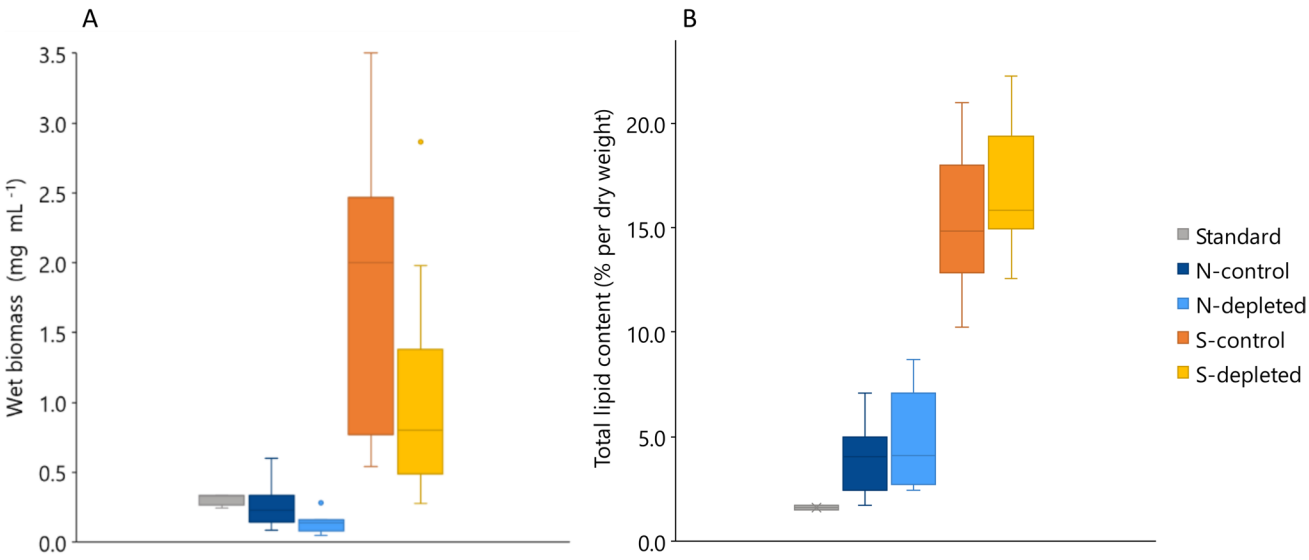

Fig. 6 (**A**) Wet biomass (mg mL^{-1}) and (**B**) total lipid content (% per dry weight), of *Chlorella vulgaris* from three different experiments

more lipids than the N-control cultivations, although their average biomass was lower.

Visualisation with confocal fluorescence microscope

The Nile Red stained lipids produced by *Chlorella* were visualised using fluorescence microscopy (Fig. 7). The fluorescence signal highlights storage lipids detected in each treatment (Fig. 7A, C–F). Fluorescence was not observed in *Chlorella* without the Nile Red (Fig. 7B). As expected, the fluorescence intensity varied between the experiments. The highest fluorescence was detected in the S-depleted treatments and the lowest in the control treatments of the sulphur experiment.

Lipid class analysis

Variation in lipid classes was observed between the different experiments in the lipid class analysis (Fig. 8). Concerning all experiments, the biggest variation was in triacylglycerols (TAG), meaning the most significant variation was in lipid

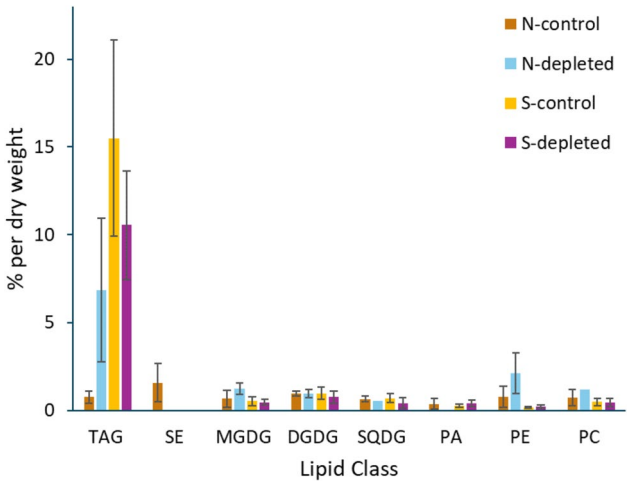

Fig. 8 Lipid class analysis for different experiments (% per dry weight). TAG = triacylglycerols, SE = sterol esters, MGDG = mono galactosyl diacylglycerol, DGDG = di galactosyl diacylglycerol, SQDG = sulfoquinovosyl diacylglycerols, PA = phosphatidic acid, PE = phosphatidylethanolamine, PC = phosphatidylcholines

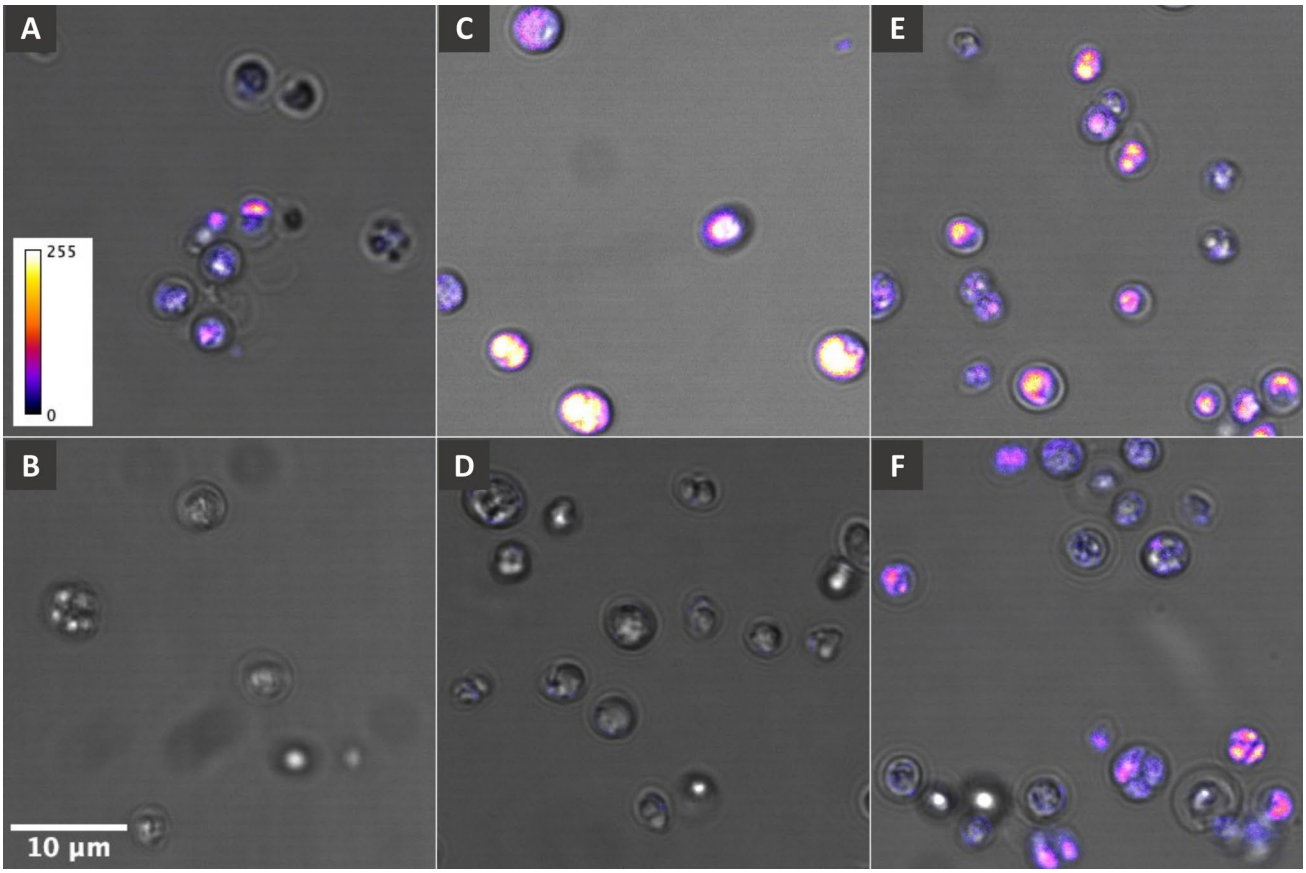

Fig. 7 Nile Red staining of lipids. (**A**) Standard experiment, (**B**) negative control without Nile Red, (**C**) S-depleted, (**D**) S-control (**E**) N-depleted, and (**F**) N-control. Fluorescence intensity is presented in a fire lookup table (calibration bar in panel A) overlayed with a differential interference contrast (DIC) signal describing the cell morphology. The images were selected to represent the highest observed fluorescent intensities

storage inside the cells. The mean concentration of TAGs was, in % per dry weight, 0.8 ± 0.4, 6.9 ± 4.1, 15.5 ± 5.6, and 10.6 ± 3.1 in the N-control, N-depleted, S-control, and S-depleted samples, respectively. Other lipid classes, such as membrane lipids of plastids, did not add much variation to the data. The only exception was the N-control cultivations where the main source of variation was sterol esters (SE, mean 1.6 ± 1.1% per dry weight). For a more detailed analysis, see Supplementary Table 1.

Table 2 Comparison of index-based linear regression (LR), partial least squares (PLS), and one-dimensional convolutional neural network (1D CNN) models. The table shows both test and validation data from the 1D CNN, which reflects the learning of the model. The values to be compared are the coefficient of determination (R^2) and the test root mean squared error (RMSE)

Model	test R^2	test RMSE (%pt)	p value
LR ($n = 34$)	0.31	6	< 0.001
PLS ($n = 34$)	0.64	4	< 0.001
CNN val ($n = 140$)	0.70 ± 0.02	4 ± 0.1	< 0.001
CNN test ($n = 34$)	0.65 ± 0.05	4 ± 0.3	< 0.001

Modelling

A Pearson correlation matrix determined from training data of the index model showed that several indices (a division of wavebands) correlated well with the *Chlorella* total lipid content (% per dry weight), that is, the ground truth (Fig. 9A). The highest positive correlation between total lipid content and A/B index was with wavebands A = 1171 nm and B = 1223 nm ($r = 0.75$, $p < 0.001$, $n = 70$), hereafter called the best index (Fig. 9B).

The index-based LR model was the weakest model to predict the total lipid content (% per dry weight) with a moderate correlation between predicted and expected lipid content and the highest test RMSE (Table 2, Fig. 10A). There were no major differences in prediction between the PLS and 1D CNN models as both had strong correlations and same level test RMSE between predicted and expected lipid content (Table 2, Fig. 10B–C). Based on the validation data, the learning of the 1D CNN model was at a higher level than the performance shown by the test data (Table 2).

The SHAP algorithm was used to find the most effective wavebands for lipid prediction in the PLS and CNN models, that is, to find the wavelengths with the highest SHAP value. Based on the modelling results, there were two separate lipid groups, one with lower and one with higher total lipid content (% per dry weight) (Fig. 10). Therefore, the three wavelengths that were most significant in predicting low lipid content (< 10% of dry weight) and higher lipid content (> 10% of dry weight) were investigated separately. For the PLS model, the most important wavelengths were at the beginning of the spectrum at 974 nm and at the end of the spectrum at 1336 nm and 1347 nm for low lipid content (grey dashed lines, Fig. 11A). For high lipid content, the corresponding wavelengths were near the largest absorption valley at 1139 nm, 1143 nm, and 1171 nm (black lines, Fig. 11A). For the 1D CNN model, the most important wavelengths are located close to the absorption valley and to the smaller reflection peak: For low lipid content at 1195 nm,

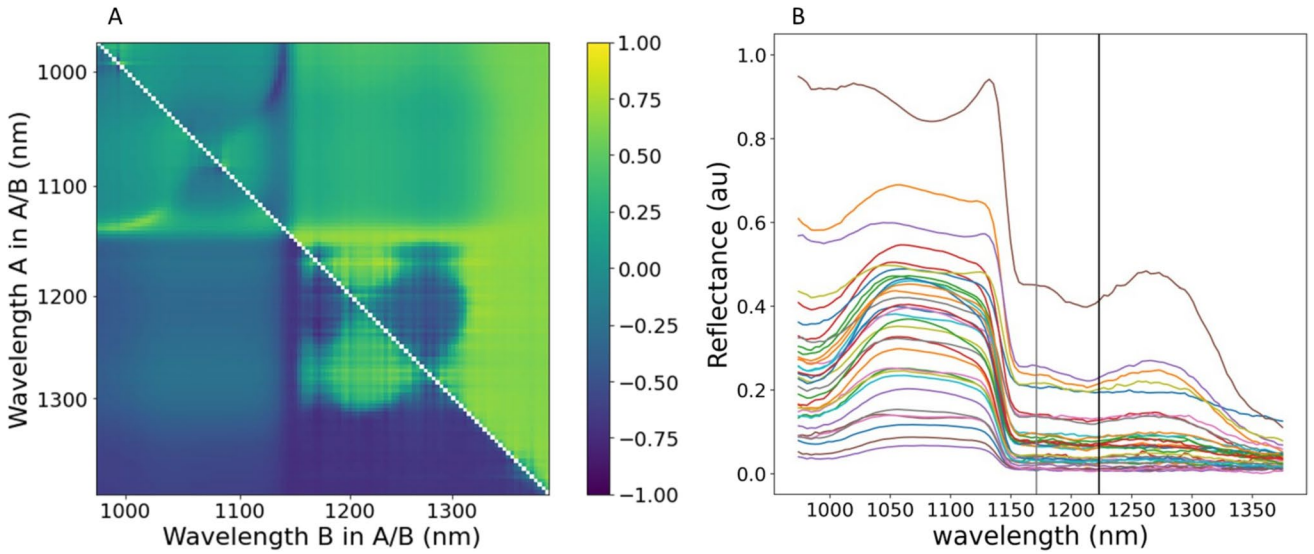

Fig. 9 (**A**) Pearson correlation matrix, determined from training data, for total lipid concentration and indices A/B where A and B are wavebands and (**B**) Mean spectra of test data from a 200 × 200 pixel area used in the index-based linear regression model. Vertical lines mark the location of the best index. The spectral data was min–max normalised for the visualisation

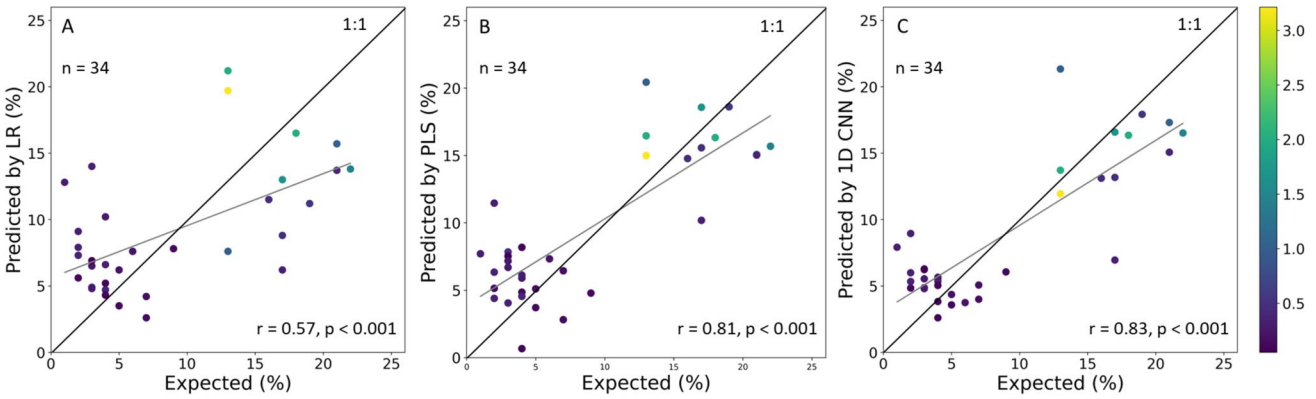

Fig. 10 The comparison of the expected and predicted total lipid content (% per dry weight) of (**A**) the index-based linear regression model (LR), (**B**) the partial least squares (PLS) and (**C**) the first of three trainings of the one-dimensional convolutional neural network (1D CNN). The results are on test data. The colour bar represents the measured wet biomass concentration (mg mL^{-1}) and the regression line represents the Pearson's correlation between expected and predicted lipid content

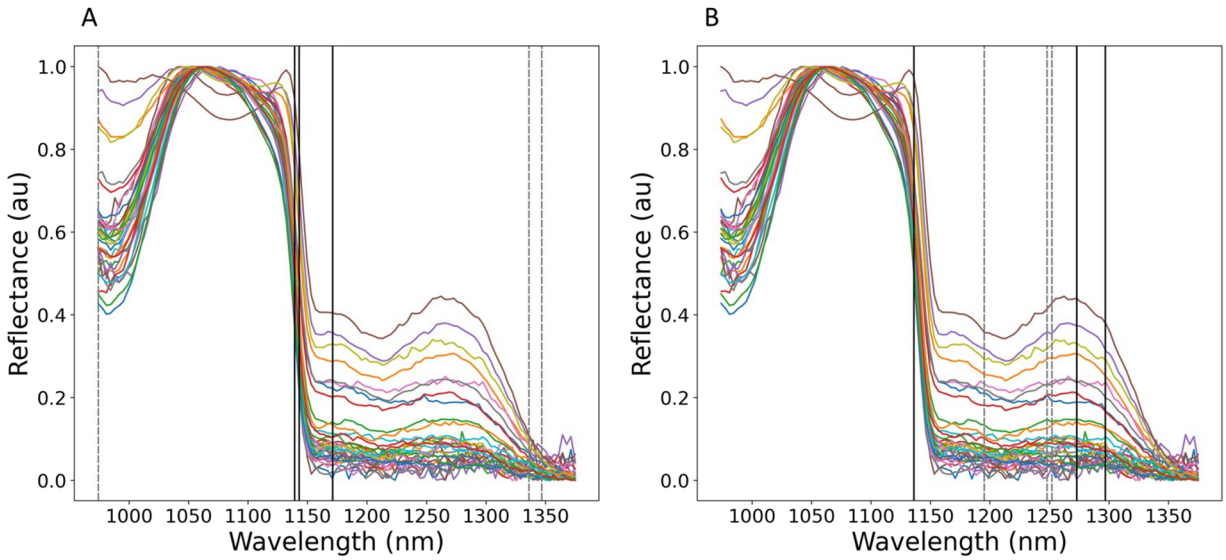

Fig. 11 A SHAP algorithm was used to determine the three most effective wavebands in predicting low total lipid content (< 10% of dry weight, grey dashed lines) and higher lipid content (> 10% of dry weight, black lines) for (**A**) partial least squares (PLS) and (**B**) one-dimensional neural network (1D CNN) models. Spectra are mean spectra of test data from a 200 × 200 pixel area

1248 nm, and 1252 nm (grey dashed lines, Fig. 11B) and, for high lipid content, at 1136 nm, 1273 nm, and 1297 nm (black lines, Fig. 11B). For more details on the relationship between SHAP values and lipid content, see Supplementary Figs. 6–7.

Discussion

This study described a non-invasive approach for lipid monitoring of *Chlorella*. The cultures were imaged with an NIR range hyperspectral imager in a reflectance setup. The imaging process was fast, as it took less than a minute to take one image. The method also posed no risk of contamination to the cultures. Total lipid content (% per dry weight) assessed with gravimetric methods served as the ground truth for three models that were compared: index-based LR, PLS, and 1D CNN model. The highest prediction performances were achieved with the PLS and 1D CNN models, which also seemed to learn the nonlinearity between biomass and total lipid concentration, as the low lipids in high biomasses could be detected and vice versa.

Brown et al. (2014) and Li et al. (2020) achieved promising results when predicting the lipids of microalgae samples using NIR spectroscopy methods. Brown et al. (2014) aimed to investigate the use of a NIR range portable reflectance

spectrometer as a fast tool to monitor microalgal lipids, more specifically fatty acid methyl esters (FAME), as a percentage of dry weight in filtered samples of two species, *Kirchneriella* sp. and *Nannochloropsis* sp. They used a PLS model achieving R^2 values of 0.97 and 0.94 for *Kirchneriella,* and *Nannochloropsis* respectively, indicating excellent model performance. Compared to Brown et al. (2014), who analysed filtered samples, Li et al. (2020) imaged liquid samples of *Scenedesmus obliquus* with a hyperspectral imager in a non-destructive way, that is, the sampling or imaging did not harm the cultivations. They compared different models in their study and achieved the best results with a multiple linear regression (MLR) model ($r = 0.94$). They also tested a PLS model and achieved excellent results ($r = 0.92$). Compared to previous studies, this study investigated a completely non-invasive way to monitor microalgae lipids in a reflectance setup, that is, cultures were imaged directly from the cultivation flasks. The closest to the level of the previous studies was the performance of the PLS ($r = 0.81$, $R^2 = 0.64$, $p < 0.001$) and the 1D CNN ($r = 0.83$, $R^2 = 0.65$, $p < 0.001$) models. The performance of the index-based LR was weak ($r = 0.55$, $R^2 = 0.31$, $p < 0.001$). The results of this study thus show that non-invasive microalgae monitoring is also a potential approach for lipid monitoring when using a suitable algorithm. It is reasonable, however, to examine why the models did not perform as well as in previous studies.

CNN models can often solve complex problems. In this study, the original training data set was rather small for the CNN model (70 images), which generally requires a lot of data (Alzubaidi et al. 2021). Due to the small dataset, the training images of the CNN model were augmented by dividing them into four smaller images. In this way, in addition to more data, variation was also added to the training data. Following the image division, the shape of a few of the mean spectra from smaller ROI became very different from the average shape of the spectra, possibly due to the illumination conditions of the specific ROI (Supplementary Fig. 2C). This may have affected the predictive ability of the 1D-CNN model. It is also possible that the uneven distribution of the data may have affected the performance of the CNN model. The model learning was enhanced by smoothing out the differences using a random oversampling algorithm (Supplementary Fig. 4). However, no lipid contents from the middle parts of the data (around 10% per dry weight) ended up in the validation data, which may have increased the prediction error. Despite this, the learning of the 1D CNN was good.

In this study, the PLS model performed at the same level as the 1D CNN in predicting lipids and required less data processing. However, the behaviour of RMSE and R^2 as a function of PC shows variation based on training data (Fig. 3). Thus, in the case of PLS, it should be noted that for the selected number of components, the predictions were at the same level as 1D CNN, but the choice of the number of PCs adds some uncertainty to the PLS. The index-based LR model performed at a lower level in predicting lipid contents than did the PLS and 1D CNN models (Table 2). Models based on vegetation indices are a common way to obtain information on, for example, microalgal biomass based on spectral data (Huang et al. 2021). However, the results show that it may not be a suitable approach for predicting lipids, probably due to the non-linearity of the data, at least in this study.

The relatively low lipid content of *Chlorella* cultures in this study may explain the magnitude of the prediction error. Li et al. (2020) achieved a prediction error of 2.4%pt (RMSE) when predicting the total lipids of *S, obliquus.* In their study, the lipid content ranged between 35 and 62% per dry weight. In this study, the error (RMSE) between expected and predicted total lipid content (% per dry weight) was higher: 6%pt with the index-based LR model and 4%pt with the PLS and 1D CNN models. The lipid contents in this study (2–22% per dry weight) were not as high as in the previous studies, but on the other hand typical for the strain (Sakarika and Kornaros 2016, 2017). In general, the error of gravimetric lipid determinations is around 3% (Lu et al. 2008). At low lipid concentrations, however, the analytical error may be higher than at higher concentrations. However, since the model error cannot be better than the error of the background data, it is reasonable to think that the prediction error of the PLS and 1D-CNN models in this study could not have been much better, or else the models would have been overfitted.

It is not known exactly what causes the spectral shape in the NIR region, and hence what correlates with the total lipids used as ground truth in this study. The experiment was designed so that the relationship between biomass and total lipid content was non-linear (Fig. 6). This is relevant because NIR spectra can correlate with biomass (Sandnes et al. 2006; Brown et al. 2014), and lipid content does not always have a linear relationship with biomass (Hu et al. 2008), as lipid accumulation in microalgae is usually highest when growth resources are limited. Both the PLS and the 1D-CNN models learned to predict total lipid content relatively well despite the non-linear relationship between lipid content and wet biomass as Fig. 10 shows that predictions of high lipid content also include low biomass cultures and vice versa. It could therefore be concluded that the differences in spectra were due to lipids, or at least some factor that correlates with lipids.

Based on the lipid class analysis, storage lipids were the main source of variation in total lipid content in *Chlorella* cultures (Fig. 8). The variation in storage lipid levels can also be observed in Nile Red images, which visualise the difference in lipid levels between treatments (Fig. 7). Therefore, the results suggest that storage lipids caused the largest

variation in the total lipid content of different treatments and had the largest impact on the model's output. According to the microscopy observations, there were many cell-to-cell Nile red intensity variations within each sample. However, since the models used average spectra from regions containing numerous cells, differences in individual cells are unlikely to affect significantly the results. Surprisingly, the images of the S-control samples were dim, although the lipid content in these samples was the second highest compared to the other treatments. There was no obvious explanation for this, however, the motive here was to visualise the main features rather than quantitative analysis. In the future, it would be interesting to investigate the monitoring of lipids in microalgal cultures based on spectral data using TAGs as ground truth, as they are of interest in different algal applications (Khan and Fu 2020).

The advantage of the CNN and PLS models used in this study and previous studies (Brown et al. 2014; Li et al. 2020) is that they self-select the features they consider important for the prediction, making them convenient to implement. Increasingly popular explanatory AI, in this study the SHAP algorithm, was used to discover the factors that the PLS and CNN model emphasised. Based on the SHAP algorithm (Fig. 11), both the PLS and 1D CNN models had important wavelengths for lipid monitoring between 1130 and 1200 nm. The result is partly aligned with Li et al. (2020) who used the same brand of hyperspectral imager and NIR wavelengths for lipid content prediction. They used competitive adaptive reweighted sampling (CARS) to find the most effective wavebands for lipid prediction, after which they used those wavebands for modelling with PLS and MLR. They found the most important wavelengths between 1050 and 1130 nm and near 1200 nm. Lipids have absorption signals between 1100 and 1200 nm (Westad et al. 2008), to which the waveband area identified as important for lipid prediction in this study corresponds. The PLS and 1D CNN models used in this study also had important wavelengths in other areas, such as between 1200 and 1300 nm, but they are more difficult to interpret. Lipids also have absorbance areas between 1350 and 1450 nm as well as 1650 and 1850 nm (Westad et al. 2008), but in this study spectra were truncated because the end of the signal was flat and noisy.

Today, lipid monitoring is based on point sampling and laboratory analysis. Bouillaud et al. (2020) used nuclear magnetic resonance (NMR) for automated lipid monitoring in a non-invasive way. Larger-scale monitoring could benefit from the spatial data provided by a spectral imager (Pääkkönen et al. 2024), which could provide more detailed information about the algal culture over a wider area compared to point sampling. For example, in such large-scale monitoring, a PLS or 1D CNN model, applied pixel-wise, could be used. In this study, the experiment was carried out on a laboratory scale. The flasks were shaken before imaging, making the microalgae culture rather homogenous, so spatial information would not necessarily have added much information about the culture. The challenge for industrial-scale monitoring is that cultivations are often outdoors, where the varying cloudiness could introduce variation into the NIR signal. The application of the approach to outdoor conditions therefore needs further research. Nevertheless, different species and conditions would probably require a new model, as many factors, including species, illumination, imaging setup, and photobioreactor (especially in non-invasive monitoring), influence the measured signal. On the other hand, if the measurement geometry remained similar, and only the species changed, the 1D CNN model could potentially be retrained, thereby eliminating the need to start the model training from scratch. Eventually, although the models used in this study were promising, they need further research before they can be used in practice for non-invasive monitoring.

Conclusion

This study presented a non-invasive spectral data-based approach for lipid monitoring of microalgae cultivations. The study monitored biomass and validated lipid accumulation using two methods to ensure that the model learned total lipid variation in different treatments. The approaches were promising, as PLS and 1D CNN could predict the total lipid content of a microalgal culture with an appropriate prediction error and learn the difference between lipid content and biomass. The results were corroborated by the SHAP result, which showed a similarity between the models in predicting lipid content when the lipid content was high. The models could benefit from further development with a larger dataset of a broader lipid content range, as their prediction errors were still larger than, for example, the error of gravimetric methods traditionally used for total lipid determination. However, the approaches used in this study could potentially provide indicative information on the lipid content of microalgal cultures during cultivation. The information could further be used to make necessary adjustments to the culturing conditions if required. The method also has the potential for online monitoring to provide up-to-date information on cultivation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10811-024-03397-6.

Authors' contributions S.P. Conceptualisation, Data Curation, Formal analysis, Investigation, Methodology, Visualisation, Writing—original draft, Writing—review & editing. I.P. Conceptualisation, Funding acquisition, Methodology, Writing—review & editing. M.C. Conceptualisation, Data Curation, Investigation, Methodology, Writing—review & editing. A.Y. Data Curation, Investigation, Writing—review & editing. V.R. Conceptualisation, Data Curation, Investigation,

Writing—review & editing. M.V. Data Curation, Investigation, Writing—review & editing. P.S. Conceptualisation, Funding acquisition, Investigation, Methodology, Project administration, Visualisation, Writing—review & editing.

Ilkka Pölönen: Conceptualisation, Funding acquisition, Methodology, Writing—review & editing.

Marco Calderini: Conceptualisation, Data Curation, Investigation, Methodology, Writing—review & editing.

Aliisa Yli-Tuomola: Data Curation, Investigation, Writing—review & editing.

Visa Ruokolainen: Conceptualisation, Data Curation, Investigation, Writing—review & editing.

Maija Vihinen-Ranta: Data Curation, Investigation, Writing—review & editing.

Pauliina Salmi: Conceptualisation, Funding acquisition, Investigation, Methodology, Project administration, Visualisation, Writing—review & editing.

Funding Open Access funding provided by University of Jyväskylä (JYU). This project received funding from the European Union – NextGenerationEU instrument and is funded by the Academy of Finland under grant number 352764 and via Business Finland, funding decision number 7134/31/2021.

Availability of data and material The datasets generated during and/or analysed during the current study are available in the CSC IDA, [https://doi.org/https://doi.org/10.23729/96494a42-bc7f-4e0f-9310-8ac8babae9b4].

Code availability The code used in this study is available in the CSC IDA, [https://doi.org/https://doi.org/10.23729/96494a42-bc7f-4e0f-9310-8ac8babae9b4].

Declarations

Competing interests The authors declare no competing interests.

References

Ahmad MT, Shariff M, Md. Yusoff F, Goh YM, Banerjee S (2020) Applications of microalga *Chlorella vulgaris* in aquaculture. Rev Aquac 12:328–346

Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8:1–74

Ansari FA, Guldhe A, Gupta SK, Rawat I, Bux F (2021) Improving the feasibility of aquaculture feed by using microalgae. Environ Sci Pollut Res 28:43234–43257

Beć KB, Grabska J, Huck CW (2021) Principles and applications of miniaturized near-infrared (NIR) spectrometers. Chem - Eur J 27:1514–1532

Bellou S, Baeshen MN, Elazzazy AM, Aggeli D, Sayegh F, Aggelis G (2014) Microalgal lipids biochemistry and biotechnological perspectives. Biotechnol Adv 32:1476–1493

Bouillaud D, Drouin D, Charrier B, Jacquemmoz C, Farjon J, Giraudeau P, Gonçalves O (2020) Using benchtop NMR spectroscopy as an online non-invasive in vivo lipid sensor for microalgae cultivated in photobioreactors. Process Biochem 93:63–60

Bricaud A, Bédhomme A, Morel A (1988) Optical properties of diverse phytoplanktonic species: experimental results and theoretical interpretation. J Plankton Res 10:851–873

Brown MR, Frampton DM, Dunstan GA, Blackburn SI (2014) Assessing near-infrared reflectance spectroscopy for the rapid detection of lipid and biomass in microalgae cultures. J Appl Phycol 26:191–198

Challagulla V, Nayar S, Walsh K, Fabbro L (2017) Advances in techniques for assessment of microalgal lipids. Crit Rev Biotechnol 37:566–578

Chen W, Zhang C, Song L, Sommerfeld M, Hu Q (2009) A high throughput Nile Red method for quantitative measurement of neutral lipids in microalgae. J Microbiol Meth 77:41–47

Coronado-Reyes JA, Salazar-Torres JA, Juárez-Campos B, Gonzalez-Hernandez JC (2020) *Chlorella vulgaris*, a microalgae important to be used in Biotechnology: a review. Food Sci Technol 42:1678–2457

De Bhowmick G, Guieysse B, Everett DW, Reis MG, Thum C (2023) Novel source of microalgal lipids for infant formula. Trends Food Sci 135:1–13

Dierssen HM, Ackleson SG, Joyce KE, Hestir EL, Castagna A, Lavender S, McManus MA (2021) Living up to the hype of hyperspectral aquatic remote sensing: science, resources and outlook. Front Environ Sci 9:649528

Dos Santos CAT, Lopo M, Páscoa RN, Lopes JA (2013) A review on the applications of portable near-infrared spectrometers in the agro-food industry. Appl Spectrosc 67:1215–1233

Fernández FGA, Reis A, Wijffels RH, Barbosa M, Verdelho V, Llamas B (2021) The role of microalgae in the bioeconomy. N Biotechnol 61:99–107

Guillard RR, Lorenzen CJ (1972) Yellow-green algae with chlorophyllide *c*. J Phycol 8:10–14

Havlik I, Beutel S, Scheper T, Reardon KF (2022) On-line monitoring of biological parameters in microalgal bioprocesses using optical methods. Energies 15:875

Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant J 54:621–639

Huang S, Tang L, Hupy JP, Wang Y, Shao G (2021) A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J for Res 32:1–6

Khan S, Fu P (2020) Biotechnological perspectives on algae: a viable option for next generation biofuels. Curr Opin Biotechnol 62:146–152

Kotapati HK, Bates PD (2020) Normal phase HPLC method for combined separation of both polar and neutral lipid classes with application to lipid metabolic flux. J Chromatogr B 1145:122099

Lam MK, Lee KT (2012) Microalgae biofuels: a critical review of issues, problems and the way forward. Biotechnol Adv 30:673–690

Li X, Chen K, He Y (2020) In situ and non-destructive detection of the lipid concentration of *Scenedesmus obliquus* using hyperspectral imaging technique. Algal Res 45:101680

Liu B, Liu J, Chen T, Yang B, Jiang Y, Wei D, Chen F (2015) Rapid characterization of fatty acids in oleaginous microalgae by near-infrared spectroscopy. Int J Mol Sci 16:7045–7056

Liu JY, Zeng LH, Ren ZH (2020) Recent application of spectroscopy for the detection of microalgae life information: a review. Appl Spectrosc Rev 55:26–59

Lu Y, Ludsin SA, Fanslow DL, Pothoven SA (2008) Comparison of three microquantity techniques for measuring total lipids in fish. Can J Fish Aquat Sci 65:2233–2241

Martínez-Guijarro R, Pachés M, Ferrer J, Seco A (2018) Model performance of partial least squares in utilizing the visible spectroscopy data for estimation of algal biomass in a photobioreactor. Environ Technol Innov 10:122–131

Mehrubeoglu M, Teng MY, Zimba PV (2013) Resolving mixed algal species in hyperspectral images. Sensors 14:1–21

Müller-Navarra DC, Brett MT, Liston AM, Goldman CR (2000) A highly unsaturated fatty acid predicts carbon transfer between primary producers and consumers. Nature 403:74–77

Pääkkönen S, Pölönen I, Raita-Hakola A-M, Carneiro M, Cardoso H, Mauricio D, Cavaco Rodrigues AM, Salmi P (2024) Non-invasive monitoring of microalgae cultivations using hyperspectral imager. J Appl Phycol 36:1653–1665

Pant G, Yadav DP, Gaur A (2020) ResNeXt convolution neural network topology-based deep learning model for identification and classification of *Pediastrum*. Algal Res 48:101932

Patel AK, Albarico FPJB, Perumal PK, Vadrale AP, Nian CT, Chau HTB, Anwar C, Wani HMUD, Pal A, Saini R, Ha LH, Senthilkumar B, Tsang YS, Chen CW, Dong CD, Singhania RR (2022) Algae as an emerging source of bioactive pigments. Bioresour Technol 351:126910

Sakarika M, Kornaros M (2016) Effect of pH on growth and lipid accumulation kinetics of the microalga *Chlorella vulgaris* grown heterotrophically under sulfur limitation. Bioresour Technol 219:694–701

Sakarika M, Kornaros M (2017) Kinetics of growth and lipids accumulation in *Chlorella vulgaris* during batch heterotrophic cultivation: effect of different nutrient limitation strategies. Bioresour Technol 243:356–365

Salmi P, Eskelinen MA, Leppänen MT, Pölönen I (2021) Rapid quantification of microalgae growth with hyperspectral camera and vegetation indices. Plants 10:341

Sandnes JM, Ringstad T, Wenner D, Heyerdahl PH, Källqvist T, Gislerød HR (2006) Real-time monitoring and automatic density control of large-scale microalgal cultures using near infrared (NIR) optical density sensors. J Biotech 122:209–215

Senatore V, Buonerba A, Zarra T, Oliva G, Belgiorno V, Boguniewicz-Zablocka J, Naddeo V (2021) Innovative membrane photobioreactor for sustainable CO_2 capture and utilization. Chemosphere 273:129682

Sheu YH (2020) Illuminating the black box: interpreting deep neural network models for psychiatric research. Front Psychiatry 11:551299

Solovchenko A (2023) Seeing good and bad: Optical sensing of microalgal culture condition. Algal Res 71:103071

Storms ZJ, Cameron E, de la Hoz Siegler H, McCaffrey WC (2014) A simple and rapid protocol for measuring neutral lipids in algal cells using fluorescence. J Vis Exp 87:e51441

Westad F, Schmidt A, Kermit M (2008) Incorporating chemical band-assignment in near infrared spectroscopy regression models. J Near Infrared Spectrosc 16:265–273

Wong Y, Ho YH, Ho KC, Leung HM, Yung KKL (2017) Growth medium screening for *Chlorella vulgaris* growth and lipid production. J Aquac Mar Biol 6:00143

Yadav DP, Jalal AS, Garlapati D, Hossain K, Goyal A, Pant G (2020) Deep learning-based ResNeXt model in phycological studies for future. Algal Res 50:102018

Zheng X, Yin L, Qiang S, Li S, Chen Y (2022) Rapid method for lipid determination in *Chlorella* sp. based on Nile Red fluorescence. Bioresour Technol 18:101077