

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Pääkkönen, Salli; Pölönen, Ilkka; Raita-Hakola, Anna-Maria; Carneiro, Mariana; Cardoso, Helena; Mauricio, Dinis; Rodrigues, Alexandre Miguel Cavaco; Salmi, Pauliina

**Title:** Non-invasive monitoring of microalgae cultivations using hyperspectral imager

**Year:** 2024

**Version:** Published version

**Copyright:** © The Author(s) 2024

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Pääkkönen, S., Pölönen, I., Raita-Hakola, A.-M., Carneiro, M., Cardoso, H., Mauricio, D., Rodrigues, A. M. C., & Salmi, P. (2024). Non-invasive monitoring of microalgae cultivations using hyperspectral imager. *Journal of Applied Phycology*, Early online.  
<https://doi.org/10.1007/s10811-024-03256-4>



# Non-invasive monitoring of microalgae cultivations using hyperspectral imager

Salli Pääkkönen<sup>1</sup> · Ilkka Pölönen<sup>1</sup> · Anna-Maria Raita-Hakola<sup>1</sup> · Mariana Carneiro<sup>3</sup> · Helena Cardoso<sup>2</sup> · Dinis Mauricio<sup>2</sup> · Alexandre Miguel Cavaco Rodrigues<sup>3</sup> · Pauliina Salmi<sup>1</sup>

Received: 23 February 2024 / Revised: 18 April 2024 / Accepted: 19 April 2024  
© The Author(s) 2024

## Abstract

High expectations are placed on microalgae as a sustainable source of valuable biomolecules. Robust methods to control microalgae cultivation processes are needed to enhance their efficiency and, thereafter, increase the profitability of microalgae-based products. To meet this need, a non-invasive monitoring method based on a hyperspectral imager was developed for laboratory scale and afterwards tested on industrial scale cultivations. In the laboratory experiments, reference data for microalgal biomass concentration was gathered to construct 1) a vegetation index-based linear regression model and 2) a one-dimensional convolutional neural network model to resolve microalgae biomass concentration from the spectral images. The two modelling approaches were compared. The mean absolute percentage error (MAPE) for the index-based model was 15–24%, with the standard deviation (SD) of 13–18 for the different species. MAPE for the convolutional neural network was 11–26% (SD = 10–22). Both models predicted the biomass well. The convolutional neural network could also classify the monocultures of green algae by species (accuracy of 97–99%). The index-based model was fast to construct and easy to interpret. The index-based monitoring was also tested in an industrial setup demonstrating a promising ability to retrieve microalgae-biomass-based signals in different cultivation systems.

**Keywords** Green microalgae · Hyperspectral imaging · Non-invasive monitoring · Vegetation indices · Convolutional neural network · Model comparison

## Introduction

Microalgae are important in biotechnology for their ability to accumulate useful biomolecules, such as healthy fatty acids and pigments (Hachicha et al. 2022). Furthermore, microalgae assimilate carbon dioxide and sequester nutrients from wastewater – features that support a circular economy (Lam & Lee 2012). However, improvements in microalgae production are still required to increase its economic viability. The development of comprehensive cultivation monitoring

methods has been recognized to be essential in resolving this challenge (Teng et al. 2020; Havlik et al. 2022).

Monitoring of microalgal cultivations is traditionally based on a combination of multiple-point sensors, sampling, and laboratory analyses (Havlik et al. 2022). Manual sampling-based assessments can bring several disadvantages such as unintentionally expose a cultivation to contaminants and introduce human error in the measurements, as well as requiring laboratory work, which reduces the temporal and spatial coverage of the assessments. Recent studies have suggested that spectral imagers could be a solution for efficient monitoring of microalgae cultivations (Murphy et al. 2014; Havlik et al. 2022; Solovchenko 2023). In addition to a typical RGB image, a hyperspectral imager produces a stack of images, also called a data cube, on hundreds of different wavebands (e.g., Raita-Hakola 2022). Previous studies indicated hyperspectral imagers' potential in monitoring microalgal growth (Murphy et al. 2014; Xu et al. 2020; Salmi et al. 2021, 2022) and in the assessment of lipid (Li et al. 2020) or pigment (Pyo et al. 2019) concentrations in the microalgal

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

<sup>1</sup> Faculty of Information Technology, University of Jyväskylä, Mattilanniemi 2, P.O. Box 35, FI-40014 Jyväskylä, Finland

<sup>2</sup> ALLMICROALGAE - Natural Products, SA, Pataias, Portugal

<sup>3</sup> Necton SA, Olhão, Portugal

biomass in samples and small volumes. However, further testing and comparison of different modelling approaches are needed to benchmark the most robust, non-invasive, monitoring pipeline.

To date, the high price of hyperspectral imagers and the complexity of data processing have been considered as main obstacles to a broader utilization of the technology (Dierssen et al. 2021; Liu et al. 2021). However, spectral cameras are becoming more common, commercially available and miniaturized, data processing algorithms more established, and the computing power required for data processing is no longer a limiting factor (Annala 2020; Dierssen et al. 2021). Spectral imaging could support traditional analyses due to its spatial coverage, low cost, and – when combined with expedient calibration models – the possibility for robust assessment of the desired parameters. For these reasons, the broader implementation of hyperspectral imagers is feasible and potentially very useful in microalgae cultivation.

The performance of a calibration model is affected by the complex matrix of a microalgal culture where the connection between optical properties and concentration of components is typically non-linear (Bricaud et al. 1988; Mehrubeoglu et al. 2014). The organization of microalgal pigments affect the propagation of light within the cell suspension. The pigment composition and materials, such as the cell wall composition of microalgae, affect the spectral absorption, scattering coefficients and refraction indices, and therefore their overall optical signature (Nair et al. 2008; Kirk 2011). Also, the cell size, arrangement and growth form (i.e., unicellular, filamentous, colonial) of microalgae affect the optical signature. In addition to the microalgal cell structures and morphology, microalgae are accompanied by other organisms, such as bacteria and protozoa, as well as organic and inorganic, dissolved and particulate material, that affect the propagation of light and, for that reason, the information captured by an imager.

The simplest way to retrieve a proxy of microalgae biomass concentration from hyperspectral data is to calculate vegetation indices, such as the ratio of a target's reflectance or absorbance on specific wavebands (Murphy et al. 2014; Salmi et al. 2021). Due to the volume of the data, machine learning algorithms that predict or generate an outcome, once they have been trained with previous data, have evolved as a popular way to extract information from hyperspectral data (Annala 2020; Havlik et al. 2022). Recently, convolutional neural networks (CNN) have been increasingly used in this context as they have potential of solving non-linear problems. The power of a CNN is based on the treatment of data through convolutional kernels before the data is fed to a neural network consisting of computation nodes organized in layers (Bengio et al. 2017). CNNs have previously been applied to automatically classify microalgae species from microscopy images (Pant et al. 2020; Yadav

et al. 2020) but also to classify species and predict their biomass concentrations from microalgae cultivation samples (Salmi et al. 2022). Machine learning models have been argued to be unnecessarily powerful algorithms in monitoring microalgae growth (Solovchenko 2023), so comparing their performance against simple index models in solving this regression problem is interesting and potentially reveals new insights into the choice of algorithms.

This study's aim was to construct a robust approach for non-invasive monitoring of microalgae growth and to compare two modelling approaches within the same imaging pipeline: vegetation index-based regression models and a CNN-based calibration model to resolve microalgae biomass concentration from liquid cultures. The pipeline from imaging to modelling was developed and tested in two separate experiments on a laboratory scale using three strains of green algae cultivated in cell culturing flasks. Finally, spectral imaging as a non-invasive monitoring method was demonstrated in an industrial microalgae cultivation facility of green algae and cyanobacteria.

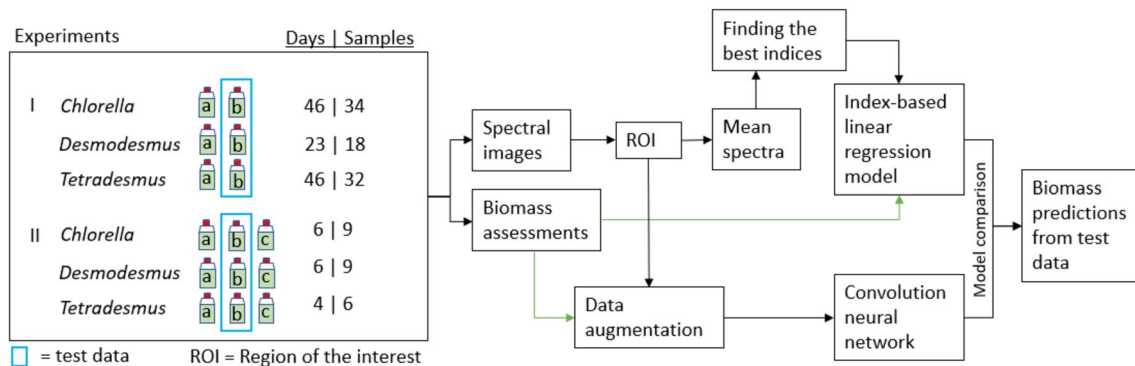
## Materials and methods

### Laboratory scale experiments

The growth of *Chlorella vulgaris* (CCAP 211/11B), *Desmodesmus maximus* (CCAC 3524B) and *Tetradesmus obliquus* (CCAP 276/3A) was monitored in two separate sets of experiments (I and II), in addition to the reference methods, by imaging the cultures through the cell culturing flask using a spectral imager. Hereafter, the algal strains are referred to by their genus names for readability. The aim was to test the ability of vegetation index- and machine learning-based calibration models to resolve biomass concentration from spectral images. Both Experiments I and II were batch cultivations in cell culturing flasks. The experiments lasted for 4–46 days resulting in a total of 84 and 24 spectral images and reference biomass concentration assessments in Experiments I and II respectively (Fig. 1). A CNN was trained using replicates *a* of Experiment I and *a* and *c* of Experiment II (Fig. 1). The models were tested using replicates *b* of Experiments I and II.

### Laboratory cultivations

In Experiment I each algal strain was cultivated in 250 mL cell culturing flasks with filter caps (VWR international) in 250 mL modified WC medium (Guillard and Lorenzen 1972) in duplicates (named *a* and *b*). The cultures were maintained at  $22 \pm 1$  °C under continuous illumination using fluorescent lamps of  $22\text{--}49 \mu\text{mol photons m}^{-2} \text{s}^{-1}$  measured with a quantum sensor (HiPoint, Taiwan). In Experiment



**Fig. 1** Left-to-right and up-to-down workflow of laboratory scale experiment. On the left are the experimental designs of the Experiments I and II, and on the right is the workflow for the Experiments. Replicate cultivations are illustrated by letters *a*, *b*, and *c*

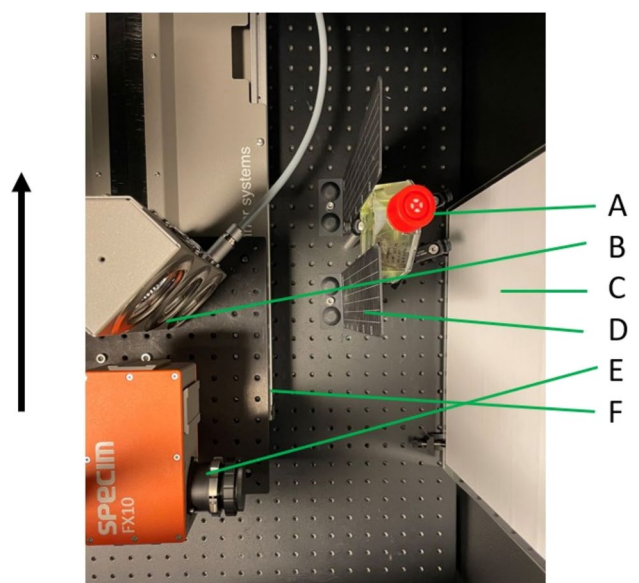
II, triplicates (named *a*, *b*, and *c*) of each algal strain were cultivated in the same scale and setup as in Experiment I.

In Experiment I, *Chlorella* and *Tetradesmus* cultivations lasted 46 days, while *Desmodesmus* cultivations lasted 23 days (see Supplementary Figure 1 for the growth of each species quantified with CASY). *Chlorella* cultures were refreshed once during the cultivation time by aseptically replacing 135 mL of the culture with WC medium to avoid nutrient limitation. *Tetradesmus* cultures were refreshed twice with 100 mL of WC medium. *Desmodesmus* cultures were not refreshed. All cultures were imaged 1–3 times per week immediately followed by aseptic sampling (1–2 mL) for biomass concentration assessment. The flasks were shaken before imaging and sampling. This experiment resulted in 34 spectral images and biomass concentration samples from *Chlorella*, 18 from *Desmodesmus* and 32 from *Tetradesmus* (Fig. 1).

In Experiment II, *Chlorella* and *Desmodesmus* cultivations lasted 6 days and *Tetradesmus* cultivations 4 days. During this time, the cultures were imaged and aseptically sampled in 2–3 days intervals leading to nine spectral images and biomass concentration samples taken from *Chlorella* and *Desmodesmus*, and six of *Tetradesmus* (Fig. 1).

### Laboratory spectral imaging

Both laboratory experiments deployed the same imaging setup (Fig. 2). Imaging was done in a darkroom to avoid any specular reflections or stray light using Specim FX10 hyperspectral imager (Specim, Finland). The imager has wavelength bands of 400–1000 nm with a spectral resolution of 5.5 nm FWHM (full width at half maximum). The imaging setup consisted of a broad-band halogen light source (3 bulbs of DECOSTAR 51 ALU 20W 12V 36deg GU5.3 halogen) and a white reference (PTFE diffuse reflector sheet PMR10P1, Thorlabs). Both white and black references were imaged for each image separately. The black reference was



**Fig. 2** Imaging setup in laboratory, top view. A – target, B – light source, C – white reference, D – laser shields, E – spectral imager, F – computer-guided scanner mounting the imager and the light source. The black arrow outside the panel marks the direction of movement of the scanner

taken automatically, with the mechanical shutter of the imager closing before each image. Reflectance (*R*) images were calculated using equation 1:

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

where *I* is the irradiance from the cultivation, *I<sub>d</sub>* the black reference and *I<sub>w</sub>* irradiance of the white reference both taken in conjunction with each image (Fig. 2).

The imager and the light source were placed on a motorized scanner (LabScanner 40 × 20, Specim, Finland). The

distance between the camera lens and the target was 25 cm. Laser shields with square engravings were placed to prevent reflections, adjust focus and scanning speed, and serve as spatial references for choosing the region of interest (ROI) in each image (Fig. 2). The selected scanning speed was 2.0 mm s<sup>-1</sup>, frame rate 20 fps and exposure time 15 s. The scanner and imager were operated through Lumo Scanner software (Specim, Finland).

### Assessment of biomass concentration in the laboratory experiments

Wet biomass concentration for both laboratory experiments was assessed with an electronic cell counter (CASY, Omni Life Sciences, Germany) immediately after the spectral imaging. In the CASY cell counter, a liquid sample flows through a capillary and a detector counts particle abundances and biovolumes based on pulse amplitude modulation caused by a bypassing particle. The capillary size of CASY was chosen according to algal cell size: 60 µm capillary for *Chlorella* and 150 µm capillary for *Desmodesmus* and *Tetrademus*. A different sample volume was added to 10 mL of CasyTon buffer based on a preliminary estimate of algae abundance: in Experiment I 40–200 µL of *Chlorella* culture, 300 µL for *Desmodesmus* and 40–700 µL for *Tetrademus* and in Experiment II 10 µL for *Chlorella*, 300 µL for *Desmodesmus* and 100–500 µL for *Tetrademus*. The results were processed with the CASY workX 1.26 macro (Omni Life Sciences) for Microsoft Excel. With *Desmodesmus*, a lower limit (left evaluation cursor) was set to 10 µm to delimit particles counted as algae cells (>10 µm). The electronic cell counter converts the pulse amplitude modulation to biovolume (fL mL<sup>-1</sup>), and biovolumes were converted to wet biomass concentrations (mg mL<sup>-1</sup>) by assuming the cells were isopycnic to water. Microscopic scrutinization at the beginning and at the end of all cultivations ensured the absence of contaminants.

The coefficient of variation (CV) of the CASY measurements was determined by measuring separate cultivations of each species grown in 600 mL cell culturing flasks in the same growth conditions as the experiments. The CV was assessed from five replicates, the error estimate including variation caused by sampling, sample preparation and measurement. For all species, the CV was low for cell abundances and for *Chlorella* and *Desmodesmus* only slightly higher for biomass concentrations (Table 1). However, for *Tetrademus*, the variation in biomass assessments was high due to the tendency of the cells to form aggregates (see Supplementary Figure 2 for example cell size distributions obtained with CASY).

**Table 1** Coefficients of variation (CV) for the reference assessments with an electronic cell counter CASY ( $n = 5$ )

CASY	CV <sub>Biomass concentration (%)</sub>	CV <sub>Abundance (%)</sub>
<i>Chlorella</i>	11	7.1
<i>Desmodesmus</i>	9.9	3.1
<i>Tetrademus</i>	41	3.7

### Index-based regression model

A 50×50-pixel ROI was extracted from each spectral image. This corresponds to a 5.2×4.9 mm area on the targets. For each captured image, the ROI was cropped from the same area of the cell culturing flasks related to the laser shields in Experiments I and II. For index calculation, mean reflectance spectra were calculated from the ROI. Images were cropped, and mean spectra and indices were calculated with Python 3.9.12 in Jupyter notebooks. Linear regression models fitting biomass concentration assessments over the best indices were derived in Microsoft Excel version 2209.

All possible indices formulated as A/B, where A and B are wavebands of the spectral camera, were calculated, and correlation coefficients determined using the biomass concentrations in replicate cultivations *a* of Experiment I and *a* and *c* of Experiments II ( $n = 58$ ). The ratio of wavebands with the highest significant ( $p < 0.001$ ) Pearson correlation form the index-based linear regression model to resolve biomass concentration. Separate models were constructed for each individual species and another describing all species together. The models were tested on data from replicate *b* of Experiments I and II ( $n = 50$ ). Mean absolute percentage error (MAPE) was calculated to evaluate the adequacy of the models. MAPE was calculated according to the Eq. 2:

$$MAPE(\%) = \frac{|B_{ex} - B_{pred}|}{B_{ex}} \times 100 \quad (2)$$

where  $B_{ex}$  is the biomass concentration assessed with CASY electronic cell counter, and  $B_{pred}$  the biomass concentration predicted by the model.

### One-dimensional convolutional neural network (1D CNN)

The one-dimensional convolutional neural network (1D CNN) was implemented with Python Version 3.9.12 in Jupyter notebooks, Keras library, and Tensorflow backend and computed on a Nvidia Tesla V100-SXM2 16 GB GPU unit. The CNN was constructed for two purposes: biomass concentration prediction and species classification.

The 50×50-pixel ROIs from replicate cultivations *a* and *c* of Experiments I and II were divided into 10×10-pixel data cubes shifting the window every 5 pixels. This

allowed data augmentation to train a 1D CNN. The augmented data ( $n = 1458$ ) were split into training and validation data by randomly dividing 80% as training and 20% as validation data. Training and validation input to the 1D CNN were the min-max-normalized mean spectra of the 10×10-pixel data cubes, with corresponding biomass concentration values of each species as data labels. Mean spectra of 50×50-pixel ROIs from Experiments I and II *b* replicate ( $n = 50$ ) were normalized by the minimum and maximum values of the training and validation data. These spectra served as test data.

The basic architecture of the 1D CNN was selected by first iterating the number of convolutional and dense layers by adding layers one at a time, with filter and node counts of 256, 128, 64, 32 and 16 in descending order, until validation losses and validation root mean squared error (RMSE) in the validation dataset were minimized (Table 1). Maxpooling layers with pool size 2 were added after each convolutional layer and a dropout layer with 0.2 drop was added before the output layer. A rectified linear unit (ReLU) was selected as an activation function for the convolutional and dense layers, excluding the output layer where linear activation was used. The model was optimized through a gradient-based stochastic optimizer (Adam) with a learning rate of 0.001,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and  $\epsilon = 1e^{-7}$ . Models were trained for 50 epochs in a sample batch size of 32.

The architecture's 5 and 9 (Table 2) hyperparameters were both tuned using Keras Tuner Random Search as their validation losses and validation RMSEs were the lowest compared to other architectures. The tuned hyperparameters were convolutional kernel sizes, number of nodes in dense layers, and learning rate. The tested node counts were between 32-256 with step size 32, filter sizes 3 and 5 and learning rates 0.01 and 0.001. Keras Tuner evaluated 100 different combinations of these hyperparameters while keeping patch size 32 for epochs constant,

with minimum validation loss as objective. The architecture 9 (Table 2) had lower validation loss after hyperparameter tuning than the architecture 5, and therefore it was selected as the most adequate model architecture (architecture 12 in Table 2). The architecture 12 had a learning rate of 0.001 and is shown in Fig. 3 (See Supplementary Figure 3 for training and validation loss, RMSE training and validation curves).

The test classification accuracy of the model was calculated with Eq. 3:

$$\text{Classification accuracy (\%)} = \frac{\text{Correctly assigned algal labels}}{\text{Total number of labels}} \times 100 \quad (3)$$

where algal labels were considered correctly assigned if the biomass concentration values exceeded the lowest biomass concentration ( $0.0069 \text{ mg mL}^{-1}$ ) of the training and validation data. This final model architecture was trained three separate times, and the means of the triplicate trainings gave the results (see Supplementary Table 1 for error metrics of the three trainings).

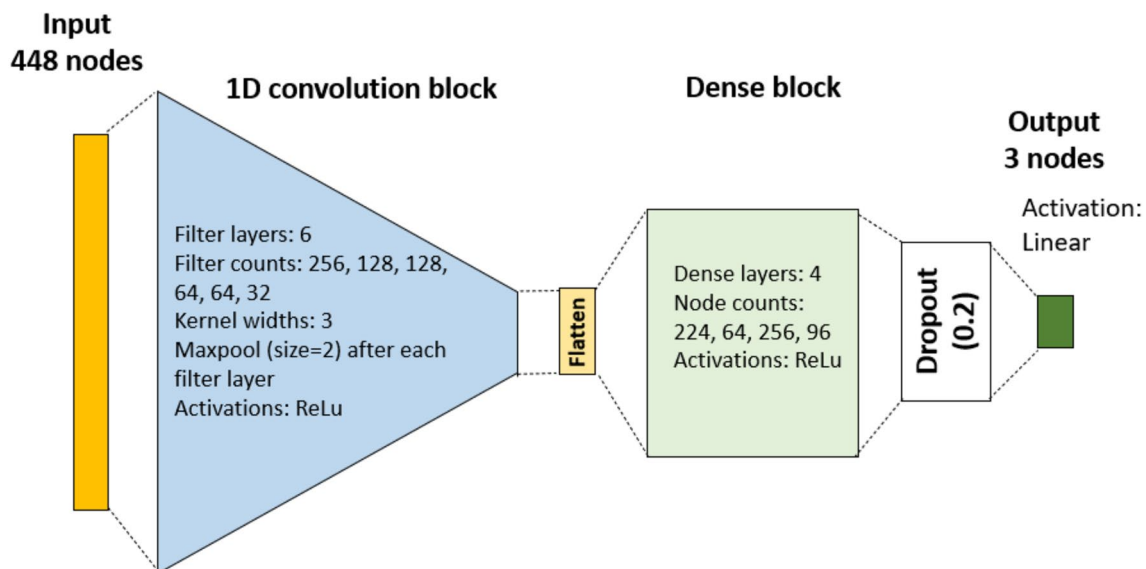
## Spectral imaging of industrial microalgae cultivations

Hyperspectral imaging was tested on-site in industrial scale photobioreactors (PBRs) and raceway ponds (Allmicroalgae, S.A., Portugal) by applying the index-based regression model approach tested previously in the laboratory cultivations. Nine flat panel PBRs with *Nannochloropsis* sp. (hereafter, referred to by its genus name) and four raceway ponds with *Arthrospira platensis* (hereafter, referred to by its common name Spirulina) were imaged (Fig. 4). The cultivation volumes of the flat panel PBRs were 0.1-1 m<sup>3</sup>, and the raceway ponds were 200-685 m<sup>3</sup>.

Specim IQ hyperspectral imager (Specim, Finland) was used in the test. The imager works in a range of 400-1000

**Table 2** Testing of the basic architectures of the 1D CNN in laboratory scale experiment. Model 12 is Model 9 architecture after hyperparameter tuning. Convolution filter counts in convolution blocks and node counts in dense blocks are separated with “/”

ID	Convolution block	Dense block	Validation loss	Validation RMSE
1	256	256	0.0071	0.084
2	256/128	256	0.0050	0.071
3	256/128	125/128	0.0030	0.054
4	256/128/64	125/128	0.0024	0.049
5	256/128/64	256/128/64	0.0013	0.036
6	256/128/64/32	256/128/64	0.0022	0.047
7	256/128/64/32	256/128/64/32	0.0057	0.076
8	256/128/64/64/32	256/128/64/32	0.0023	0.048
9	256/128/128/64/64/32	256/128/64/32	0.0015	0.039
10	256/256/128/128/64/64/32	256/128/64/32	0.0031	0.18
11	256/256/128/128/64/64/32	256/128/64/64/32	0.0064	0.080
12	256/128/128/64/64/32	224/64/256/96	0.0009	0.03



**Fig. 3** The architecture of the 1D CNN after hyperparameter tuning. This model was used to predict the algae species and biomass concentrations in Experiments I and II in laboratory scale experiment

**Fig. 4** Industrial scale imaging setups for *Nannochloropsis* sp. in flat panel photobioreactors and of *Spirulina* (*Arthrospira platensis*) in raceway ponds

*Nannochloropsis*, flatpanel



*Spirulina*, raceway



nm with a spectral resolution of 7 nm FWHM. Both spectral imagers (Specim FX10 in the laboratory experiments and IQ in the industrial tests) used in this study have identical operating principles as line scanners. The primary distinction between the two imagers lies in their adaptability to various contexts. The FX imagers are engineered for integration with conveyor belts or linear scanning systems, whereas the IQ model simulates the functionality of a snapshot imager by housing a line scanner within its framework. Consequently, while both imagers operate on the same fundamental principle, the IQ model offers enhanced portability for field measurements.

The industrial scale imaging happened inside greenhouses, in daylight. Because of the slightly opaque

coverings of the greenhouses, the lighting inside was diffuse, allowing images to be taken without notable reflections. The spectral imager was placed on a tripod, making imaging setups replicable for every imaged reactor. Variable illumination was considered by imaging a white reference Teflon tile (Specim, Finland), which was set at the reactor surface level, before each image to normalize the spectra between the white reference and a black reference. The black reference was taken automatically, with the mechanical shutter of the imager closing before each image. Depending on the lighting conditions, an automatically varying exposure time of the imager was used, which ranged between 17–93 ms with flat panels and 7–18 ms with raceway ponds. The imaging geometry was kept

similar for each cultivation system (44 cm height and 66–67 cm distance in direct angle for flat panels, 125–127 cm height, in 50° angle set on the edges of the raceway ponds).

The full spectral images (512×512 pixel ROI) were used to calculate mean spectra. The mean spectra were truncated between 426–691 nm (channels 10–100) to scrutinize the area with photosynthetic pigments and to exclude variation in the infrared area possibly caused by the greenhouse materials and varying cloudiness. The whole spatial area was used also with flat panels, even though they were surrounded by a metal frame, as the metal had little effect on the spectral mean. To calculate the best index to correlate with the flat panels' biomass, all the possible indices formulated as A/B, where A and B are wavebands of the spectral camera, were calculated and the best index was determined based on highest significant ( $p < 0.001$ ) Pearson correlation coefficient. In Allmicroalgae, the growth of *Nannochloropsis* cultures was assessed through optical density (OD) measurement on 540 nm waveband (Genesys 10s UV-Vis, Thermo Scientific) and a yearly updated linear regression model between OD and biomass dry weight (DW, in  $\text{g L}^{-1}$ ). Assessments of DW were done in triplicates by collecting 30 mL of culture on pre-weighed 0.7  $\mu\text{m}$  glass microfiber filters (698, VWR). The samples were washed with 10 mL of ammonium formate ( $35 \text{ g L}^{-1}$ ) and oven-dried at 60 °C until constant weight. For *Spirulina*, 600 nm waveband was used for OD measurement. The protocol for assessing dry weight was the same as for *Nannochloropsis*, but the samples were not washed with ammonium formate.

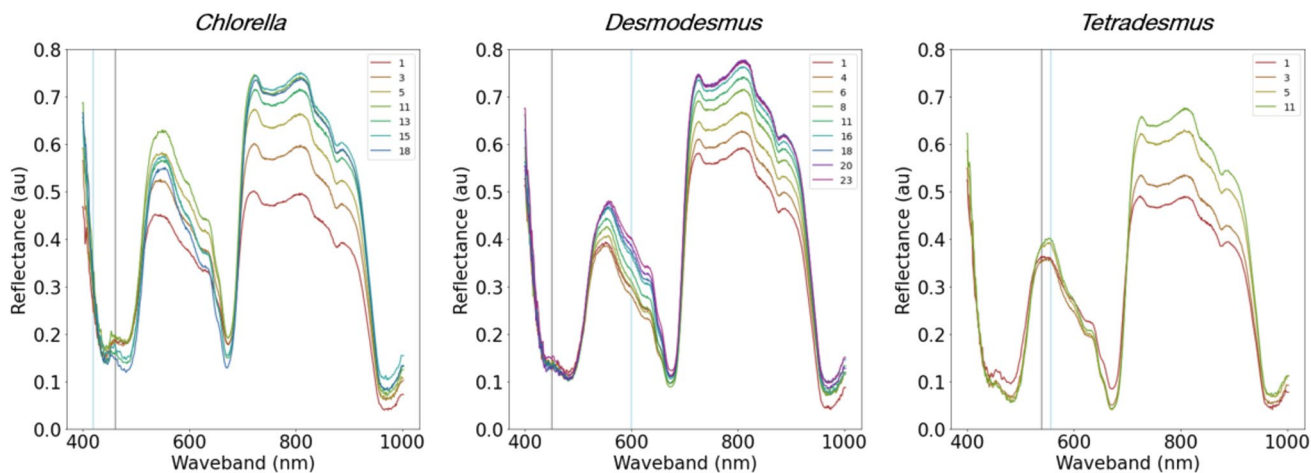
## Results

### Laboratory experiments

The differences in the spectral signatures between the studied microalgal strains were subtle, with all studied microalgae exhibiting distinct absorbance maxima on blue and red waveband areas (Fig. 5). Algal growth over time could be observed as an increase of reflectance on green wavebands and the near-infrared (NIR) area (Fig. 5) (see Supplementary Figure 1 for the growth of each species quantified with CASY). This study did not scrutinize the spatial distribution of microalgae in the cultures because the cell culturing flasks were thoroughly mixed before imaging and sampling. Therefore, notable aggregation of algae was not assumed, and only the spectral information was considered in the modelling approaches.

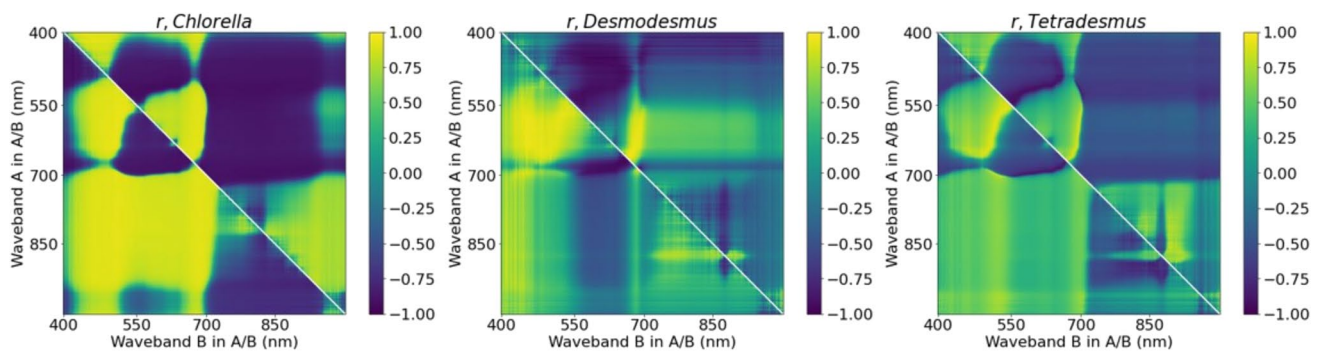
### Index-based regression model

Spectra of all studied algal strains showed strong Pearson correlation between biomass concentration measurements and A/B indices, where A and B are wavebands on green and blue waveband areas (Fig. 6). *Chlorella* biomass concentration correlated strongly also when A and B were on red-to-NIR and blue or green waveband areas (Fig. 6). The highest positive correlation between *Chlorella* biomass concentration and A/B index was with wavebands A = 419 nm and B = 461 nm ( $r = 0.95$ ,  $p < 0.001$ ,  $n = 23$ ), hereafter called the best index. For *Desmodesmus* the best index was with A = 600 nm and



**Fig. 5** Example of mean spectra of replicate *b* of each species in Experiment I in laboratory scale experiment. Vertical lines mark the location of the best indices. Numbers represent the day of the experiment. The spectral data was min-max normalized for the visualization





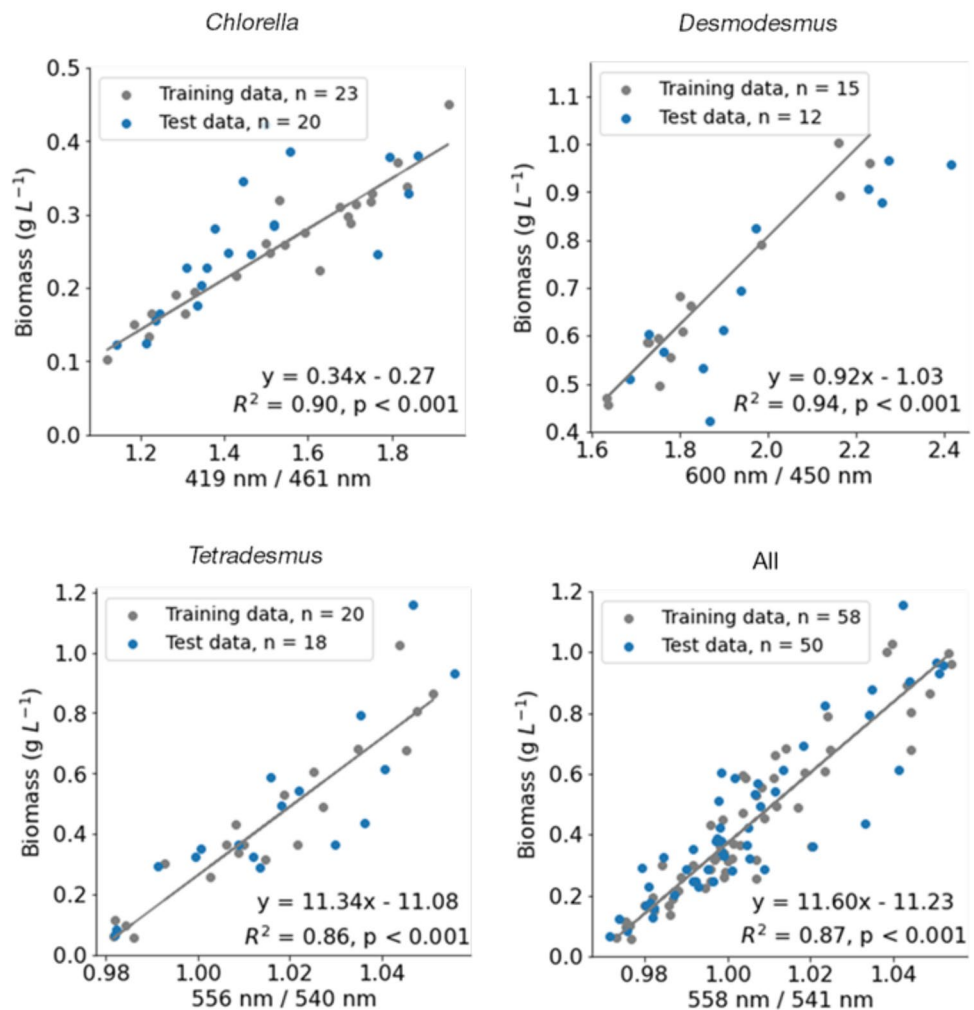
**Fig. 6** Pearson correlation matrices for biomass concentrations and indices A/B where A and B are wavebands in Experiments I and II in laboratory scale experiment

B = 450 nm ( $r = 0.97$ ,  $p < 0.001$ ,  $n = 15$ ) and for *Tetradesmus* with A = 556 nm and B = 540 nm ( $r = 0.93$ ,  $p < 0.001$ ,  $n = 20$ ) (Fig. 5).

A linear regression model fitted between the best index for *Chlorella* and biomass concentration reference data of replicates *a* and *c* (training data) predicted well the biomass concentration

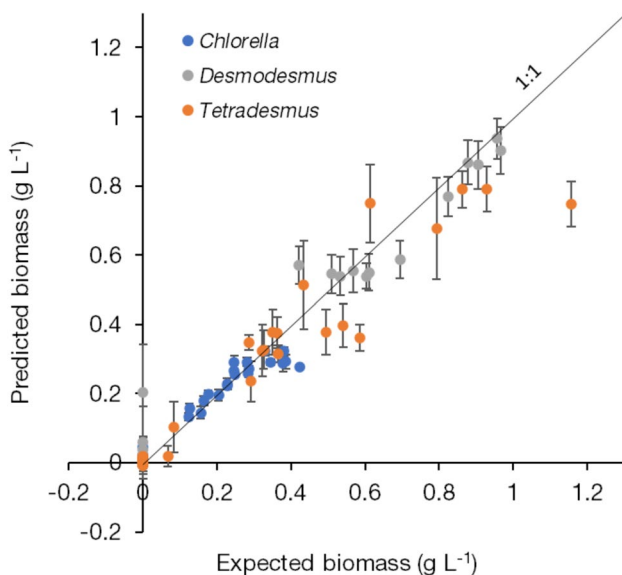
of replicates *b* (test data,  $n = 20$ , Fig. 7). The MAPE was 15%, with the standard deviation (SD) being 13. Similarly, the predictability of the *Desmodesmus* model was good (MAPE 16%, SD = 16,  $n = 12$ ); as well as the *Tetradesmus* model (MAPE 24%, SD = 18,  $n = 18$ ) and the model predicting all three species together (MAPE = 21%, SD = 17,  $n = 50$ ).

**Fig. 7** Linear regression model between the biomass concentration estimates and the best indices in Experiments I and II in laboratory scale experiment



**Table 3** Comparison of average and standard deviation (SD) (in brackets) of mean absolute percentage error (MAPE) of biomass predictions using index-model and the 1D CNN in Experiment I and II in laboratory scale experiment. The values are given for the test data (replicates *b* of Experiment I and II). Values for the 1D CNN are means and standard deviations of three replicate trainings of the chosen model architecture

Strain	Index-based model MAPE (%)	1D CNN MAPE (%)
<i>Chlorella</i> ( <i>n</i> = 20)	15 (13)	12 (10)
<i>Desmodesmus</i> ( <i>n</i> = 12)	16 (16)	11 (10)
<i>Tetradesmus</i> ( <i>n</i> = 18)	24 (18)	26 (22)
All ( <i>n</i> = 50)	21 (17)	17 (17)



**Fig. 8** Species classification and the ratio of the expected and 1D CNN predicted biomass concentrations in the test data (replicates *b* of Experiment I and II in laboratory scale experiment). Means values and standard deviations of the three replicate trainings of the chosen model architecture are shown

### One-dimensional convolutional neural network (1D CNN)

Like the index-based models, the 1D CNN performed well (Table 3, Fig. 8), considering the model performance cannot be expected to be better than the variation in the reference biomass concentration assessment (Table 1). MAPEs were between 11–26 % for the different species (Table 3) being highest for *Tetradesmus* which exhibited the highest variation also in the reference biomass assessments (Table 1).

The advantage of the 1D CNN over the index-based model is that it simultaneously predicted biomass concentration and accurately classified the algal monocultures of different species (Fig. 8). The species classification accuracies varied between 97–99 % (See Supplementary Table 1

for precision, recall, specificity and F1 score metrics). The species accuracies for *Chlorella*, *Desmodesmus* and *Tetradesmus* were 99 % (SD = 1, *n* = 20), 97 % (SD = 3, *n* = 12) and 97 % (SD = 1, *n* = 18), respectively. Occasionally *Desmodesmus* biomass concentrations were confused with others so a low biomass concentration of *Desmodesmus* was predicted in cultivations of *Chlorella* and *Tetradesmus* (Fig. 8, See supplementary Fig. 4 for a confusion matrix).

### Spectral imaging of industrial scale cultivations

The spectral imager was able to obtain distinctive algal spectra from both cultivation systems regardless of the plastic surface and metal frame of the flat panels (Fig. 9A). There were differences between the spectra of the two species, probably due to their different pigment compositions. For example, *Spirulina* shows a phycocyanin absorption peak around 620 nm, unlike *Nannochloropsis*, which does not possess phycobiliproteins. One of the distinctive spectra from raceway spectra plot (green spectra, Fig. 9A) was from a raceway that had not been properly mixed.

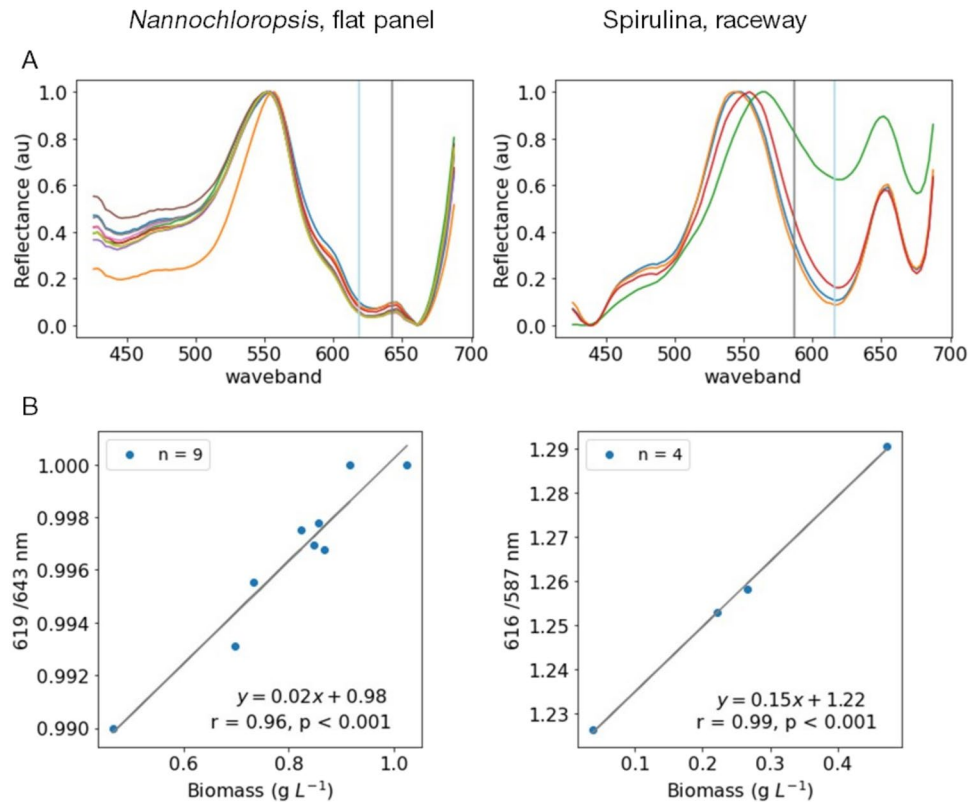
The best index determined for *Nannochloropsis* was 619/643 nm and for *Spirulina* 616/587 nm (Fig. 9B) (See Supplementary Figure 5 for correlation matrices). Both indices correlated strongly with the expected biomass. The best index visualization is an example of how the predicted biomass of *Spirulina* in one raceway (Fig. 10A) is distributed in a spectral image (Fig. 10B).

### Discussion

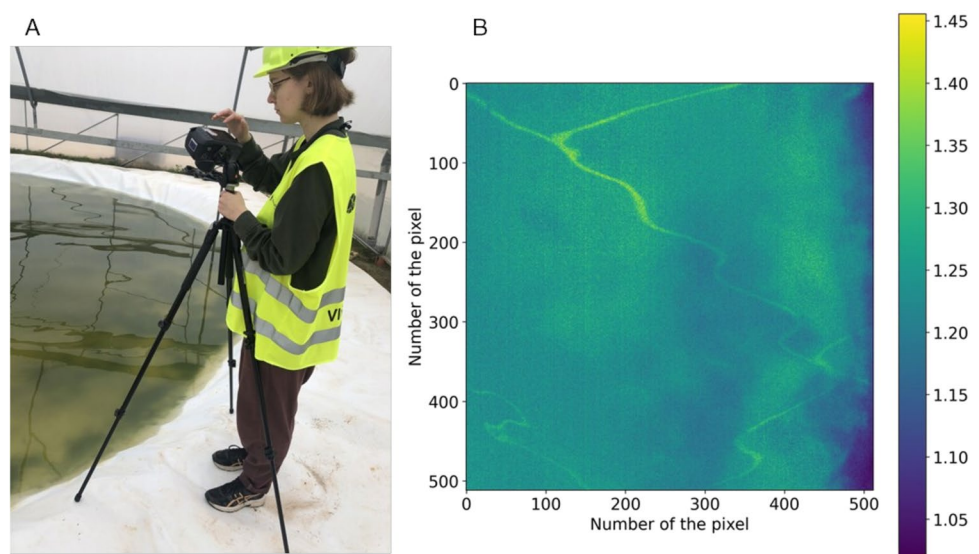
This study aimed to construct an effective, non-invasive, and robust method for monitoring microalgae growth. Both an index-based linear regression model and a CNN model were constructed to resolve the biomass concentration of three considered microalgae strains on a laboratory scale. The simpler, index-based approach was initially tested also for industrial scale. The imaging was fast, with a duration of less than a minute per sample. The non-invasive nature of the imaging arrangements posed no risk of contamination to the culture. Capturing the dark and white reference with each target image controlled the variation caused by possible warming of the sensor or the light source during the imaging sessions, and the variation by daylight conditions in the industrial setup. The results show that the reflectance imaging system combined with both the index-based approach and CNN work equally well and that the imaging setup can be implemented in different volumetric scales.

Murphy et al. (2013) achieved a prediction error of 15 % when predicting the biomass concentration of *Anabaena variabilis* from liquid samples with a multispectral imager. They imaged 400 mL samples of different concentrations

**Fig. 9** **A** Mean spectra of *Nannochloropsis* sp. in flat panel photobioreactors and *Spirulina* (*Arthrospira platensis*) in raceway ponds in industrial scale test. Vertical lines mark the location of the best indices. Each spectrum was min-max normalized for the visualization. **B** Pearson correlation between the biomass concentration estimates (dry weight) and the best indices



**Fig. 10.** **A** Spectral imaging of a raceway with *Spirulina* (*Arthrospira platensis*) in industrial scale test. **B** An index visualization calculated from a spectral image of a raceway pond using the best determined index (616/587 nm). This figure is an example of how the predicted biomass is distributed in a spectral image



and calculated the correlation between green waveband and areal biomass concentration. In this study, the predictabilities achieved both with the vegetation index-based linear regression model and 1D CNN were comparable to the results by Murphy et al. (2013). The poorer predictability for *Tetrademus* biomass concentration in this study can be explained by the tendency of the cells to form aggregates,

which affects the reference biomass concentration assessment by CASY.

Vegetation indices, such as the Normalized Difference Vegetation Index (NDVI) based on NIR and red wavebands are common methods for monitoring vegetation, including microalgal biomass, in spectral applications (Huang et al. 2021). Salmi et al. (2021) found a good correlation

between NIR/Red index calculated from transmittance spectral images of 2 mL samples on a well-plate and the total biomass concentration of microalgae when five different microalgae species were studied ( $r = 0.86$ ,  $p < 0.001$ ). In their study, changing the index architecture had no significant effect on the correlation between vegetation index and biomass, and therefore for this study, only the simplest index architecture (waveband A/ waveband B) was examined. In this study, only the best index is presented in the results which for all three species was on the green wavebands 558/541 nm ( $R^2 = 0.87$ ,  $p < 0.001$ , Figs. 5 and 6). This is in line with the observation that the laboratory cultures became visibly greener during growth (Fig. 5). The best index was also in the green wavebands for *Tetradesmus*. In the case of *Chlorella* and *Desmodesmus*, important wavelengths were also found between 420 and 450 nm, where the absorption peaks of Chlorophyll *a* and *b* are located (Chazaux et al. 2022). However, in this study, several wavebands including the NIR/Red areas correlated well (Fig. 6).

Salmi et al. (2021) studied the same *Desmodesmus* strain as in the results presented here. In their study, the best index to describe the biomass concentration of *Desmodesmus* was found to be 631/643 nm ( $r = 0.66$ ,  $p < 0.001$ ). In this study, the best index for *Desmodesmus* was 600/450 nm. Comparing with the best index found by Salmi et al. (2021), it becomes apparent, that the waveband reaching around 600 nm contains important information about the algal biomass concentration, even though the indices between the two studies varied. This was also observed in the industrial scale tests where the best indices occurred on the red wavelengths for the green microalgal cultivations. For the *Spirulina* cultivations, the most descriptive indices occurred in the area of phycocyanin absorbance (Fig. 8B).

The use of indices in biomass monitoring has several challenges. Due to the properties of light, the wavelengths in a green part of the spectrum will saturate with increasing biomass. This is particularly evident in the cultivation of dense biomass (Gitelson et al. 1996). The possible saturation of the index should therefore be considered if the wet biomasses are higher than in the laboratory experiments of this study ( $> 1.2 \text{ mg mL}^{-1}$ ). Since the visible light range contains a signal in the pigments of the microalgae, changes in the pigment properties of the microalgal culture can therefore also affect the index used. For example, pigments can be affected by the adaptation of the algae to changing environmental conditions, such as changes in nutrient levels (Goiris et al. 2015). In addition, the measured signal is affected by lighting, imaging setup and the photobioreactor (especially in non-invasive monitoring). This work highlights that spectral imaging allows the use of multiple wavebands for monitoring. Thus, using this approach, with a specifically defined index, can give better results than a pre-determined literature-based index.

In the laboratory experiments, the classification accuracies for microalgae monocultures of this study achieved with 1D CNN varied between 97–99%. The results were in line with both Yadav et al. (2020) and Pant et al. (2020) who reached classification accuracy of 99.97 % and 98.45 %, respectively. They classified algae with CNN, although from microscopic images, unlike in this study where the spectra were used. The prediction accuracies of this study were also in line with the study of Xu et al. (2020), who classified three species of algae (*Phaeocystis*, *Chlamydomonas*, and *Chaetoceros*) imaged with a transmission hyperspectral microscopy imager, comparing two different methods. The first method applied principal component analysis to normalized (min-max) transmission spectra followed by linear SVM (support vector machine) for classification. The second method calculated ratios 680/550 nm and 440/550 nm from the transmission data, again followed by linear SVM for classification. Both methods yielded an accuracy of 94.4%. In the same study, a random forest model predicted the growth stage of *Phaeocystis* from spectral images, achieving a prediction accuracy of 98.1 % ( $R^2 = 0.998$ ). It can be noted that in this study with 1D CNN, in addition to classification, biomass concentration could be predicted, while the accuracy remains at the same level compared to previous studies that focused on classification (Pant et al. 2020; Xu et al. 2020; Yadav et al. 2020).

The industrial scale test was promising as it showed that the method investigated in this study can also be applied to various industrial scale cultivation systems and therefore can be useful for the algae sector. The different photobioreactors gave a clear spectral signal with species-specific differences. As a contrast to sampling-based measurements, the spatial variation of biomass can be observed from a spectral image (Fig. 10). The same benefit can also be observed when comparing spectral instruments with optical fibers, which have been shown to yield good results in monitoring microalgae biomass (Morgado et al. 2024): optical fibers are fast to acquire data but have a relatively small observed area. Spectral cameras may be comparatively more expensive to deploy but provide information from a wider area of the cultivation. Therefore, the use of an imager could be beneficial, particularly when the algae are not well mixed in the PBR. Overall, the imaging setups used in this large-scale test were replicable and easy to use, which makes future research worthwhile.

In Allmicroalgae, biomass determination is based on OD measurements, which in general are proven to be an accurate basis for biomass monitoring models. (Griffiths et al. 2011). The wavelengths of the best indices determined in this study do not match the wavelengths used in Allmicroalgae's calibration model. In the case of *Spirulina*, the wavelength of the calibration model falls between the best indices (cf. 600 nm and 616/587 nm), while in the case of *Nannochloropsis*,

the wavelengths used in the index are higher (cf. 540 nm and 619/643 nm). However, several indices in the correlation matrices correlate well with biomass (Supplementary Figure 5) so a good correlation result could also have been achieved at other wavelengths. The correlation between the indices obtained in this study and the biomass results from the calibration model of the Allmicroalgae is promising and encourages further implementations.

Machine learning methods are knowingly data-intensive, i.e. CNN needs more data for training than the regression model. Training the CNN also requires more computing power and time to train the model. Therefore, for relatively simple problems, such as biomass prediction, a simpler regression model may be an adequate algorithm. For these reasons, the CNN was not applied for the initial tests on the industrial scale in this study. However, a more extensive study with a longer duration could reveal the benefits of CNN on a larger scale, too. Once the monitoring pipelines have been established, the practical applications of both CNN and regression models are similar, both allowing real-time monitoring.

The non-invasive imaging setup seemed to work well in laboratory conditions, as both the index-based regression models and the 1D-CNN model predicted well the biomass of each species. In addition to laboratory conditions, Specim imagers are adaptable to industrial use. Future research could concentrate on on-site and further online monitoring. One possible area of research could be the detection of contaminations in algal cultures. In this study, in addition to predicting biomass, 1D-CNN was able to distinguish three species of algae monocultures from each other at a laboratory scale, even though they were all green algae. This could indicate that contaminations by other algae species could be distinguished from spectral data using a machine learning-based approach. In this study, biomass concentration was the target parameter but simultaneous assessment of pigment or lipid composition and concentration in laboratory and industrial scale cultivations could also be an interesting target for further development.

## Conclusions

This study described non-invasive imaging arrangements for microalgae cultivations in cell culturing flasks and industrial scale cultivation systems. In this study, both index-based linear regression models and machine learning-based calibration models resolved the biomass concentration from spectral images with adequate prediction errors in laboratory scale cultivations. Index-based biomass concentration prediction was a simple and replicable way to monitor algal biomass concentration. The availability of multiple wavebands in a spectral imager enabled the fitting of different models for different microalgae species and volumetric scales.

Whilst the machine learning models might be unnecessarily complex for simply predicting biomass concentration, the advantage of the 1D CNN over the index-based model was that it simultaneously predicted biomass concentration and accurately classified the green algae monocultures by species. An industrial-scale test in this work showed that non-invasive spectral imaging could have the potential to be implemented in large-scale cultivation systems meeting the needs for more efficient and comprehensive monitoring.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1007/s10811-024-03256-4>.

**Authors' contributions** Salli Pääkkönen: Conceptualization, Data Curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing

Ilkka Pölönen: Conceptualization, Funding acquisition, Methodology, Writing - review & editing

Anna-Maria Raita-Hakola: Conceptualization, Methodology, Writing - review & editing

Mariana Carneiro: Conceptualization (Industrial scale test), Writing - review & editing

Helena Cardoso: Conceptualization (Industrial scale test), Methodology (Industrial scale test), Writing - review & editing

Dinis Mauricio: Conceptualization (Industrial scale test), Methodology (Industrial scale test), Writing - review & editing

Alexandre Miguel Cavaco Rodrigues: Conceptualization (Industrial scale test), Writing - review & editing

Pauliina Salmi: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing - review & editing

**Funding** Open Access funding provided by University of Jyväskylä (JYU). This research was funded by the European Union - NextGenerationEU via Business Finland, funding decision number 7134/31/2021.

**Data availability** The datasets generated during and/or analysed during the current study are available in the CSC IDA, [<https://doi.org/10.23729/1e576402-1ccc-4974-a392-e014bd6cec38>].

**Code availability** The code used in this study is available in the CSC IDA, [<https://doi.org/10.23729/1e576402-1ccc-4974-a392-e014bd6cec38>].

## Declarations

**Competing interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Annala L (2020) Convolutional neural networks and stochastic modelling in hyperspectral data analysis. Dissertation, University of Jyväskylä, Finland, pp 56
- Bengio Y, Goodfellow I, Courville A (2017) Deep learning. MIT Press, Massachusetts, USA
- Bricaud A, Bédhomme A, Morel A (1988) Optical properties of diverse phytoplanktonic species: experimental results and theoretical interpretation. *J Plankton Res* 10:851–873
- Chazaux M, Schiphorst C, Lazzari G, Caffarri S (2022) Precise estimation of chlorophyll *a*, *b* and carotenoid content by deconvolution of the absorption spectrum and new simultaneous equations for Chl determination. *Plant J* 109:1630–1648
- 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
- Gitelson A, Qiuang H, Richmond A (1996) Photic volume in photobioreactors supporting ultrahigh population densities of the photoautotroph *Spirulina platensis*. *Appl Environ Microbiol* 62:1570–1573
- Goiris K, Van Colen W, Wilches I, León-Tamariz F, De Cooman L, Muylaert K (2015) Impact of nutrient stress on antioxidant production in three species of microalgae. *Algal Res* 7:51–57
- Griffiths MJ, Garcin C, van Hille RP, Harrison STL (2011) Interference by pigment in the estimation of microalgal biomass concentration by optical density. *J Microbiol Meth* 85:119–123
- Guillard RR, Lorenzen CJ (1972) Yellow-green algae with chlorophyllide *c*. *J Phycol* 8:10–14
- Hachicha R, Elleuch F, Hlima HB, Dubessay P, de Baynast H, Delattre C, Pierre G, Hachicha R, Abdelkafi S, Michaud P, Imen F (2022) Biomolecules from microalgae and cyanobacteria: Applications and market survey. *Appl Sci* 12:1924
- Havlik I, Beutel S, Scheper T, Reardon KF (2022) On-line monitoring of biological parameters in microalgal bioprocesses using optical methods. *Energies* 15:875–902
- 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
- Kirk JT (2011) Light and photosynthesis in aquatic ecosystems. Cambridge University Press, Cambridge
- 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 J, Zeng L, Ren Z (2021) The application of spectroscopy technology in the monitoring of microalgae cells concentration. *Appl Spectrosc Rev* 56:171–192
- Mehrubeoglu M, Teng MY, Zimba PV (2014) Resolving mixed algal species in hyperspectral images. *Sensors* 14:1–21
- Morgado D, Fanesi A, Martin T, Tebbani S, Bernard O, Lopes F (2024) Non-destructive monitoring of microalgae biofilms. *Bioresour Technol* 398:130520
- Murphy TE, Macon K, Berberoglu H (2014) Rapid algal culture diagnostics for open ponds using multispectral image analysis. *Biotechnol Prog* 30:233–240
- Murphy TE, Macon K, Berberoglu H (2013) Multispectral image analysis for algal biomass quantification. *Biotechnol Prog* 29:808–816
- Nair A, Sathyendranath S, Platt T, Morales J, Stuart V, Forget M, Deyred E, Bouman H (2008) Remote sensing of phytoplankton functional types. *Remote Sens Environ* 112:3366–3375
- 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
- Pyo J, Duan H, Baek S, Kim MS, Jeon T, Kwon YS, Lee H, Chon KH (2019) A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. *Remote Sens Environ* 233:111350
- Raita-Hakola AM (2022) From sensors to machine vision systems: Exploring machine vision, computer vision and machine learning with hyperspectral imaging applications. Dissertation, University of Jyväskylä, Finland, pp 132
- Salmi P, Eskelinen MA, Leppänen MT, Pölönen I (2021) Rapid quantification of microalgae growth with hyperspectral camera and vegetation indices. *Plants* 10:341
- Salmi P, Calderini M, Pääkkönen S, Taipale S, Pölönen I (2022) Assessment of microalgae species, biomass, and distribution from spectral images using a convolution neural network. *J Appl Phycol* 34:1565–1575
- Solovchenko A (2023) Seeing good and bad: Optical sensing of microalgal culture condition. *Algal Res* 71:103071
- Teng SY, Yew GY, Sukačová K, Show PL, Máša V, Chang J (2020) Microalgae with artificial intelligence: A digitalized perspective on genetics, systems and products. *Biotech Adv* 44:107631
- Yadav DP, Jala AS, Garlapati D, Hossain K, Goyal A, Pant G (2020) Deep learning-based ResNeXt model in phycological studies for future. *Algal Res* 50:102018
- Xu Z, Jiang Y, Ji J, Forsberg E, Li Y, He S (2020) Classification, identification, and growth stage estimation of microalgae based on transmission hyperspectral microscopic imaging and machine learning. *Optics Express* 28:30686–30700

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.