

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

Author(s): Calderini, Marco L.; Pääkkönen, Salli; Yli-Tuomola, Aliisa; Timilsina, Hemanta; Pulkkinen, Katja; Pölönen, Ilkka; Salmi, Pauliina

Title: Accurate non-invasive quantification of astaxanthin content using hyperspectral images and machine learning

Year: 2025

Version: Published version

Copyright: © 2025 The Authors. Published by Elsevier B.V.

Rights: CC BY 4.0

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

Please cite the original version:

Calderini, M. L., Pääkkönen, S., Yli-Tuomola, A., Timilsina, H., Pulkkinen, K., Pölönen, I., & Salmi, P. (2025). Accurate non-invasive quantification of astaxanthin content using hyperspectral images and machine learning. *Algal Research*, 87, Article 103979.
<https://doi.org/10.1016/j.algal.2025.103979>

The University of Jyväskylä's Rights Retention Strategy (RRS) enables accepted versions (final drafts) to be made openly available through self-archiving. For more information, see: <https://www.jyu.fi/en/rrs>



Accurate non-invasive quantification of astaxanthin content using hyperspectral images and machine learning

Marco L. Calderini^{a,*}, Salli Pääkkönen^a, Aliisa Yli-Tuomola^b, Hemanta Timilsina^b,
Katja Pulkkinen^b, Ilkka Pölönen^a, Pauliina Salmi^a

^a Faculty of Information Technology, University of Jyväskylä, P.O. Box 35, FI-40014 Jyväskylä, Finland

^b Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35, FI-40014 Jyväskylä, Finland

ARTICLE INFO

Keywords:

Astaxanthin
Haematococcus pluvialis
Hyperspectral imaging
Machine learning
Monitoring

ABSTRACT

Commercial cultivation of *Haematococcus pluvialis* to produce natural astaxanthin has gained significant traction due to its application in feeds, cosmetics and nutraceuticals. However, monitoring of astaxanthin content in cultures remains challenging and requires invasive, time consuming and expensive approaches. Here, we employed reflectance hyperspectral imaging (HSI) of *H. pluvialis* suspensions within the visible spectrum, combined with a 1-dimensional convolutional neural network (CNN) to predict the astaxanthin content ($\mu\text{g mg}^{-1}$) quantified by high-performance liquid chromatography (HPLC). This approach had low average prediction error (5.9 %) for samples cultured under the same conditions and was only unreliable at very low astaxanthin contents ($<0.6 \mu\text{g mg}^{-1}$). In addition, our CNN model outperformed single or dual wavelength linear regression models even when the spectral data was obtained with a spectrophotometer coupled with an integrating sphere. Different cultivation conditions to the ones used for the training of the CNN model affected the reflectance spectra of *H. pluvialis* cultures, leading to higher prediction errors (32.9 %). Transfer learning, a partial re-training of the model, fine-tuned the CNN to the variability observed in the new samples leading to a lower average prediction error (8.2 %). Overall, this study proposes the use of HSI in combination with a CNN for precise non-invasive quantification of astaxanthin in cell suspensions. Optimal performance is achieved when the training dataset accurately reflects the inherent variability of the target samples to be quantified.

1. Introduction

Microalgae are a diverse group of single-cell photosynthetic organisms that can serve as a biotechnological platform to produce high-value biomolecules and quality biomass. Their high photosynthetic efficiency and rapid growth, their ability to thrive in water that is unsuitable for human consumption and their high content of pigments, polyunsaturated fatty acids and other health promoting compounds have increased commercial interest in microalgae production over the past decades [1]. Currently, commercial production of microalgae in Europe focuses on high-value products for the supplement and nutraceutical industry (24 %) as well as the cosmetics industry (24 %) [1]. Common to these industries is the use of astaxanthin, a red-orange secondary ketocarotenoid that in microalgal cells serves as a scavenger of free radicals to protect against photooxidative stress [2]. The high antioxidant capacity of astaxanthin has been observed in vivo and in vitro [3,4], driving interest in its potential therapeutic applications for

cardiovascular, neurodegenerative and immune diseases [5]. Additionally, astaxanthin is used as a colorant and is responsible for the natural coloration observed in salmonids, shrimp, lobsters and crayfish [6]. Microalgal production of astaxanthin is centred in the cultivation of the green algae *Haematococcus pluvialis* [6]. Under stress conditions such as nutrient limitation and high light, this species can accumulate up to 7 % of its dry weight as astaxanthin [7]. Astaxanthin is not a single molecule since the presence of multiple conjugated double bonds allows for several possible geometric isomers [8] with the most abundant being all-E, followed by 9Z-, 13Z- and 15Z-forms in *H. pluvialis* [8,9]. In microalgal cells, only small amounts of astaxanthin remain as free molecules (0.3–4 %), while most of this ketocarotenoid is esterified with fatty acids to form monoesters (76–94 %) and diesters (2–23 %) [9–11].

Monitoring of cultures is a critical aspect of microalgal production systems, since parameters such as biomass concentration and biochemical composition significantly influence the decision-making process of business operations [12]. Traditionally, microalgae culture

* Corresponding author.

E-mail address: marco.92.calderini@jyu.fi (M.L. Calderini).

<https://doi.org/10.1016/j.algal.2025.103979>

Received 4 October 2024; Received in revised form 27 January 2025; Accepted 23 February 2025

Available online 25 February 2025

2211-9264/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

monitoring depended on invasive methods in which a sample of biomass is taken, and different analytical laboratory methods are used to determine the parameter of interest. Currently however, efforts have been placed in developing non-invasive monitoring strategies to minimize the time lag between the measurement and availability of results as well as lowering the labour and operational costs required for the measurement [13]. Spectroscopic methods based on the absorption of light by microalgae remain the most widely used approach for routine monitoring of cultures [12,14]. The Bouguer-Beer-Lambert law defines the linear correlation between the absorbance of a non-scattering medium and the sum of the products of each dissolved analyte's molar absorptivity and its concentration, for a specified optical path length. This law is primarily applied to pigment extracts, but it can be an adequate approximation to study the absorbance of microalgal suspensions. Given that microalgal pigments absorb light in the visible range, the absorbance spectra between 350 and 700 nm can be used for the quantification of pigments while optical density (OD) in the near infrared (e.g. OD₇₅₀) can be used for whole biomass estimation [15]. However, microalgal cell morphology and pigment packaging can affect light absorption and scattering by cell suspensions [16], leading to non-linearities that deviate from the Bouguer-Beer-Lambert law. To minimize the impact of light scattering on a measured spectrum, an integrating sphere (IS) can be utilized, where most forward-scattered light from the sample is integrated into the measurement [17]. Nevertheless, a high-end IS can be expensive [12].

Estimation of total astaxanthin content in *H. pluvialis* cultures is one of the crucial steps during commercial production. Commonly, cells are harvested, pigments extracted using physical disruption methods in combination with organic solvents and astaxanthin is subsequently quantified based on the light absorption properties of the total pigment extract [18]. Pure all-E astaxanthin has a maximum absorption between 470 and 506 nm depending on the solvent [19], but there is significant overlap with the absorption peaks of chlorophylls and other carotenoids. This, added to the fact that stress conditions lead to the degradation of chlorophyll and the synthesis of secondary carotenoids other than astaxanthin [20], has prompted researchers to suggest using 530 nm as a wavelength to quantify astaxanthin while minimizing interference with the absorption peaks of other pigments [21]. More sophisticated approaches for astaxanthin quantification require pigment extraction and de-esterification of astaxanthin followed by chromatographic separation and quantification of each astaxanthin isomer independently [8,9]. Absorbance-based quantification is relatively fast, but due to the overlap of the astaxanthin absorption peak with other carotenoids, deviations of up to 30 % in astaxanthin content can occur when astaxanthin content is lower than 2 % of dry weight [9]. This large error makes this estimation method very unreliable at low astaxanthin contents. In contrast, chromatography-based methods present high accuracy, but they are expensive, time consuming and require highly trained personal in addition to expensive equipment.

Reflectance hyperspectral imaging (HSI) has been proposed as a powerful and cost-effective alternative spectroscopic method for non-invasive monitoring of scattering microalgal cultures [13,22,23]. A hyperspectral imager produces a set of images (>20) or data cube, where each image corresponds to a different wavelength band and each pixel contains spatial information of the imaged target. One or two HSI wavelength bands can be used directly to, for example, quantify microalgal biomass [22,24,25]. However, this approach misses the opportunity of utilizing the information contained across the whole visible spectrum. Machine learning (ML) algorithms represent an attractive approach to work with microalgal reflectance spectra since they can deal with the difficulties inherent to this kind of data such as the collinearity between close wavelengths. Additionally, ML can deal with non-linear relationships which might be more suitable for modelling reflectance data from scattering suspensions [26]. ML algorithms that have shown promising results when dealing with spectroscopic data are convolutional neural networks (CNN) [23,25,27,28]. CNN were originally

developed for computer vision tasks and consists of one or several convolution layers followed by a deep-forward neural network (dense layers). The convolution kernels work as filters that select the most important features in the input while the dense layers perform the classification or regression task. CNN applied to microalgae reflectance HSI data have been successfully used for species classification [23,25] as well as for quantification of biomass and pigments [23,25,29,30]. Transfer learning is a ML technique in which a pre-existing model is adapted (partially re-trained) to perform in different domains or distributions [31]. The aim of transfer learning is to conserve the valuable information acquired during the original training and only optimize the model for a new dataset. In the case of a CNN, transfer learning can involve retaining the convolution layers intact and retrain the model's dense layers. The logic behind such approach is to utilize the capacity of the CNN for feature recognition and only adapt the output of the model to such features. This approach requires a smaller number of training samples than training a complete model, making the process efficient and not so computationally expensive.

In this study, we tested the use of a CNN model to predict the astaxanthin content per dry weight ($\mu\text{g mg}^{-1}$) of *H. pluvialis* suspensions based on visual reflectance HSI data. As a reference, we tested the prediction error of linear models constructed using single or dual wavelengths from reflectance HSI or absorbance data from a high-end IS. We hypothesize that the variation in carotenoids and chlorophylls during stress conditions is coordinated in *H. pluvialis* and that such spectral information can be used to quantify astaxanthin accurately. In addition, we hypothesize that although a costly IS can produce high-quality absorbance spectra, reflectance HSI data in combination with the non-linear capabilities of a CNN can outperform linear regression models constructed using IS data.

2. Materials and methods

2.1. Laboratory cultivations for base model training and testing

The *H. pluvialis* strain used in this study was obtained from the Norwegian Culture Collection of Algae (NORCCA) and maintained as a stock monoculture in Modified Wright's Cryptophyte (MWC) algae medium [32] with the addition of filter-sterilized (0.22 μm) vitamins (0.5 $\mu\text{g L}^{-1}$ biotin [B7], 0.5 $\mu\text{g L}^{-1}$ cyanocobalamin [B12], 0.5 $\mu\text{g L}^{-1}$ pyridoxine [B6], 0.1 mg L^{-1} thiamine HCL [B1]). Cultivation was divided into two stages, with stage one prioritizing biomass accumulation while stress conditions were applied in stage two to maximize astaxanthin accumulation. In stage one, 8 parallel cultivations took place in 300 mL glass funnels aerated from the bottom with ambient room air as described in [33] utilizing MWC media with higher concentrations of nitrogen and phosphorus (0.17 g L^{-1} of NaNO_3 and 0.023 g L^{-1} of $\text{K}_2\text{HPO}_4 \cdot 3\text{H}_2\text{O}$). Light was supplied at 38–42 $\mu\text{mol m}^{-2} \text{s}^{-1}$ with LED growth lights (18W, L-series T8 tubes, Valoya Oy) under a 12:12 photoperiod and room temperature was set at 20 °C. Cells were maintained in stage one until cultivations presented a green-brown colour. At that point, 200 mL of culture were centrifuged for 5 min (2500g, 18 °C) and cells were transferred to 500 mL Erlenmeyer flasks containing 200 mL of nitrogen deficient MWC media. Erlenmeyer flasks were placed in a shaker under constant high light (1400–1550 $\mu\text{mol m}^{-2} \text{s}^{-1}$) supplied by LED lights (R150 LED Roof Lighting, Valoya Oy) and aerated with filtered room air. Fresh MWC media was added to the culture funnels (stage one) to replace the subtracted volume and such funnel was considered a new culture. Stage one lasted no more than two weeks from the addition of fresh media while stage two did not last more than three weeks. In total, 35 and 21 cultures were sampled from stage one and two, respectively. Sampling for spectroscopy was conducted using a pseudo-randomized approach, meaning that samples were not collected at consistent, predefined time intervals. This approach helped to capture the variability in astaxanthin contents withing cultivation stages one and two. Additionally, there was a focus on early stages of astaxanthin

accumulation to obtain a dataset with spectral data of cell suspensions with low astaxanthin contents per dry weight. In practice, 100 mL of microalgae culture were collected at different timepoints from cultivation stages one and two to obtain its absorbance and reflectance spectrum. To increase the variability in astaxanthin concentration of the dataset, mixtures of samples from stages one and two were carried out by mixing different volumes of each sample (up to 1 and 99 % of each culture). In total, 98 spectra were recorded of which 42 corresponded to sample mixtures, while the rest (56) corresponded to non-mixed samples (35 and 21 spectra from cultures of stage one and two, respectively).

2.2. Microalgal suspension spectroscopy

Absorbance spectra were recorded with a spectrophotometer (Lambda 850 UV/VIS, Perkin Elmer) equipped with an integrating sphere by placing a glass cuvette (outer dimensions 50 × 50 × 14 mm, OP38, eCuvettes) containing the microalgal suspension in front of the integrating sphere. The cuvette had an optical path length of 10 mm (approximately 30 mL of microalgae). Spectra were measured across 350–750 nm with a wavelength resolution of 1 nm.

Reflectance spectra were recorded with a hyperspectral imager (Specim FX10, Specim, Finland) as described in [25]. Briefly, microalgal suspensions (40 mL) were placed in 50 mL cell culture flasks (EasYFlask, Nunc) at 25 cm from the camera lens. The camera and the light source (3 bulbs of DECOSTAR 51 ALU 20 W 12 V 36 deg. GU5.3 halogen) were mounted in a motorized scanner (LabScanner 40 × 20, Specim, Finland) and scanning speed was set to 2.0 mm s⁻¹. White (PTFE diffuse reflector sheet PMR10P1, Thorlabs) and black references were imaged for each image separately. Spectra were measured across 400–1000 nm with a spectral resolution of 5.5 nm.

2.3. Biomass collection and astaxanthin quantification

After spectroscopical measurements, a known volume of microalgal suspension was filtered through a pre-weighed fiber filter (GF/A, Whatman) and subsequently re-weighed after 12 h at 105 °C to determine its dry weight (mg L⁻¹). The rest of the microalgal suspensions were transferred to 50 mL Falcon tubes and centrifuged for 10 min (5000g, 18 °C). Cell pellets were transferred to 1.5 mL Eppendorf tubes and stored at -80 °C. Microalgal biomass was freeze-dried overnight before pigment extraction. Freeze-dried biomass (1 ± 0.1 mg) was transferred to ZR BashingBead Lysis tubes (0.1 & 0.5 mm, Zymo research) and 3.1 µg of internal standard (Trans-β-Apo-8'-carotenal, Sigma Aldrich) were added together with 1 mL of ice-cold acetone. Tubes were homogenized for 20 s at 4 m s⁻¹ for 4 rounds (Bead Ruptor Elite, OMNI) and subsequently centrifuged for 15 min (15,000 RCF at 4 °C). The pigment-rich supernatant was placed in a 10 mL KIMAX tube and stored at -20 °C until further use. A second extraction round with ice-cold acetone, exactly as described, was carried out and supernatants were combined. To the final combined supernatants (approx. 2 mL), 1 mL of acetone, 2 mL of Tris solution 50 mM (pH = 7) and 0.6 mL of cholesterol esterase enzyme (3.3 U mL⁻¹) isolated from *Pseudomonas* sp. (MP Biomedicals, Germany) were added. Freeze-dried enzyme was initially resuspended in Tris 50 mM to achieve the desired concentration (3.3 U mL⁻¹). Final enzyme concentration in KIMAX tube was 0.36 U mL⁻¹ as suggested by [9]. De-esterification of astaxanthin was carried out at 37 °C in a water bath (in darkness) for 45 min. After de-esterification, samples were allowed to cool down to room temperature, 2 mL of petroleum ether were added and tubes were centrifuged for 3 min (2000 g) to facilitate phase separation. The pigment-rich upper phase was transferred to a new KIMAX tube and a second extraction with 2 mL of petroleum ether was carried out (as before). Upper phases were combined and evaporated under a nitrogen stream in total darkness. To dry tubes, 300 µL of acetone were added and pigment extracts were transferred to brown vials. During the whole extraction process, samples were protected from direct light exposure and only dim was used when

necessary.

Pigment extracts were separated and analysed via high-performance liquid chromatography (HPLC; Nexera, Shimadzu) coupled with a SPD-M20A diode array detector (Shimadzu) using an Acclaim C30 column (3 µm 150 × 3.0 mm, Thermo Fisher Scientific). In total, three astaxanthin peaks were identified by their retention time and characteristic spectra [8] as all-E, 13Z- and 9Z-form (in decreasing order of abundance, respectively). It is likely that other less abundant astaxanthin isomers co-eluted with the three identified forms. Quantification of astaxanthin (µg) was done by interpolation in a 3-point calibration curve generated with synthetic

(3S,3'S)-Astaxanthin standard (DHI). It was assumed that the area-to-concentration relationship of the astaxanthin standard was the same for all identified astaxanthin isomers. Astaxanthin (µg) was then normalized by the weighed dry weight to obtain astaxanthin content per dry weight (µg mg⁻¹). To correct for the astaxanthin lost during the extraction process, the recovery percentage of the internal standard was used as a correction factor for astaxanthin. Sample recovery percentage was 77 % on average with two of the analysed samples been outliers (Fig. S1). These two outliers were not included in further analysis. Finally, astaxanthin concentration per litre was calculated by multiplying astaxanthin content in mg g⁻¹ by the dry weight (g L⁻¹) of the culture.

2.4. Linear regression

HSI data was first cropped to the region of interest (120 × 150 pixels; ROI) where the *H. pluvialis* suspension was located in each image (Fig. S2). Then, the dataset without outliers ($n = 96$) was randomly divided into training and testing data ($n = 76$ and $n = 20$, respectively). Given that each ROI had 18,000 pixels, each with its own reflectance spectrum, the average reflectance was calculated for each training and test sample. Principal component analysis (PCA) of training samples was then carried out using the library sklearn and test samples were projected into the calculated principal components. The IS data train-test partition was the same as with HSI data, meaning that the same samples were training and testing for both HSI and IS to ensure that the models created with both instruments were comparable. Selection of wavelengths for modelling was done using training data only. The absolute Pearson correlation was calculated between wavelengths from absorption or reflectance data with astaxanthin concentration per litre (mg L⁻¹) and content per dry weight (µg mg⁻¹). This was done by iterating over every single wavelength directly or a wavelength divided by 750 nm and selecting the highest absolute correlation. A linear model was then built between the selected wavelength and astaxanthin in the corresponding unit using the Python library SciPy (module scipy.stats). The presence or absence of an intercept was decided based on which model presented the highest coefficient of determination (R^2) with the training data. Model fitting on test data was evaluated with the root mean squared error (RMSE) and mean absolute percentage error (MAPE), calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (C_i - \hat{C}_i)^2}{N}}$$

and

$$MAPE = \left(\frac{1}{N} \sum_{i=1}^N \frac{|C_i - \hat{C}_i|}{C_i} \right) \times 100$$

where C_i is the expected astaxanthin (µg L⁻¹ or µg mg⁻¹), $\hat{C}_i$ is the predicted astaxanthin and N is the number of samples. To study if MAPE was affected by cultures with low astaxanthin, the MAPE of samples with quantified astaxanthin was >2 µg L⁻¹ or µg mg⁻¹ was calculated. This statistic is referred in the text as MAPE_{>2}. Additionally, Pearson correlation and mean squared error (MSE) of test data was calculated.

2.5. Data augmentation and 1D convolutional neural network

HSI training data (not averaged) was augmented by an iterative process where a random training sample was selected and the reflectance spectrum of 100 randomly selected pixels were averaged. This process was repeated 15,000 times, to obtain 15,000 augmented samples. To visualize the effect of data augmentation on the distribution on the training data, augmented data was projected into the principal components previously calculated (see “Linear regression” section). A portion of the augmented training data (20 %) was randomly selected as validation data for model training. Only the wavelengths in the visible range were used for modelling (398.91–757.38 nm) since the prediction error was lower than using the full HSI spectra (data not shown). Before model training, all spectral data (training, validation and test) were min-max-normalized.

A One-dimensional Convolutional Neural Network (1D CNN) was developed to predict astaxanthin content per dry weight ($\mu\text{g mg}^{-1}$). The model was implemented with Python (v 3.10.9) using Keras library (v 3.1.1) and Tensorflow (2.16.1) backend on Jupyter lab (3.5.3) editing environment. A simplified custom CNN architecture was employed, which omitted the use of max-pooling layers [34] and connected each neuron in the last convolution layer to only one neuron in the first dense layer (Table S1, Fig. 1). By using this architecture, the number of trainable parameters was reduced from 208,881 to 31,051. Convolution layers had no activation function while dense layers, except for the output layer, used Leaky Rectified Linear Unit (Leaky ReLU). The output layer had a linear activation function. To avoid overfitting, each neuron in the dense layers had a dropout rate of 10 %. The model was trained with a gradient-based stochastic optimizer (Adam) using a learning rate of 0.003 for a maximum of 200 epochs (with early stopping). Batch size was 30 samples and mean squared error (MSE) was used as a loss function. The number of convolution and dense layers were manually adjusted until high performance was achieved. Once the best model was obtained (based on lowest MSE of validation data), the SHAP algorithm (SHapley Additive exPlanations) was used to analyse which areas of the spectra were important for the model prediction.

2.6. Laboratory cultivations for further model testing

To test the generalization capabilities of the CNN model, five new

cultivation conditions were applied. Firstly, *H. pluvialis* was cultivated in filtered (27 μm) UV-sterilized (3 h) recirculating aquaculture system wastewater ($\text{NO}_3\text{-N}$: 32.11 mg L^{-1} , $\text{PO}_4\text{-P}$: 2.52 mg L^{-1}) obtained from the Natural Resources Institute Finland (LUKE) Laukaa fish farm. In total, 12 photobioreactors (1 L) were set under constant light (55–67 $\mu\text{mol m}^{-2} \text{s}^{-1}$), of which 6 were aerated with ambient room air (CO_2 : 400 ppm) and 6 with elevated CO_2 air (CO_2 : 1900–2500 ppm). Once the cultures reached early stationary phase, astaxanthin production was stimulated with high light (200–400 $\mu\text{mol m}^{-2} \text{s}^{-1}$; 18 W L-series T8 tubes, Valoya Oy). Three random photobioreactors from each CO_2 condition were sampled for HSI after 7 days of the start of the cultivations and then all photobioreactors were sampled for HSI at days 1, 3 and 10 after the start of the high light conditions (total HSI: 42). In addition to these samples, 3 different concentrations of NaCl (1, 2 and 3 g L^{-1} ; each concentration: $n = 3$) were used to promote astaxanthin accumulation in *H. pluvialis* [35]. Before the addition of NaCl, biomass was grown in 300 mL glass funnels containing high nutrient MWC (NaNO_3 : 0.17 g L^{-1} , $\text{K}_2\text{HPO}_4 \cdot 3\text{H}_2\text{O}$: 0.023 g L^{-1}). Funnels were aerated from the bottom ambient room air under a 18:6 photoperiod (38–42 $\mu\text{mol m}^{-2} \text{s}^{-1}$; 18 W, L-series T8 tubes, Valoya Oy). All cultivations were then transferred to high light (1000–1200 $\mu\text{mol m}^{-2} \text{s}^{-1}$, R150 LED Roof Lighting, Valoya Oy) and NaCl was added to achieve the appropriate concentration. The lowest NaCl treatment (1 g L^{-1}) did not receive aeration, while the rest of the treatments were aerated with room air. Cultures were sampled for HSI on day 6 after the start of high light and NaCl treatment. All 51 cultures described were imaged with a hyperspectral imager as described in Section 2.2, images were cropped as described in Section 2.4, and astaxanthin quantified as described in Section 2.3.

2.7. CNN transfer learning

For further testing of the CNN base model, the ROIs of samples ($n = 51$) from the five new cultivation conditions were first averaged individually to obtain their reflectance spectra. The astaxanthin contents of these samples were predicted using our CNN base model and the predictions were compared to their HPLC values. For transfer learning, ROIs were divided into training ($n = 20$) and testing ($n = 31$) datasets, from now on called transfer learning (TL) - datasets. The TL - training dataset was augmented as described in Section 2.5 to obtain 4000 training

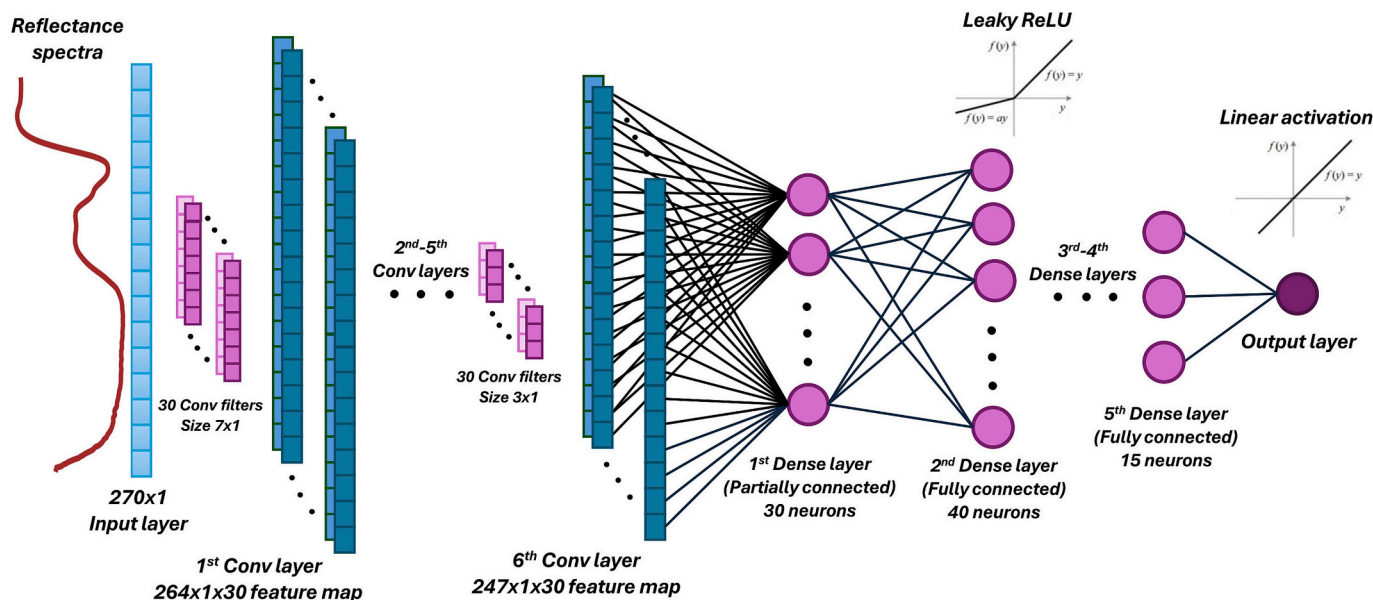


Fig. 1. 1-Dimensional (1D) convolutional neural network (CNN) architecture to predict astaxanthin content per dry weight ($\mu\text{g mg}^{-1}$) utilizing reflectance hyperspectral imaging (HSI) data of *H. pluvialis* suspensions.

spectra, while the average reflectance was calculated for each TL - test sample. The TL - training dataset was further divided into TL - training ($n = 3200$) and TL - validation ($n = 800$). Transfer learning was carried out by freezing the weights of all convolution layers and further training the weights of the dense layers using the augmented TL - training data. Adam optimizer using a learning rate of 0.003 for a maximum of 200 epochs (with early stopping) for training. Batch size was 30 samples and mean squared error (MSE) was used as a loss function.

3. Results

3.1. Dataset characteristics and recorded spectra

Dry weight of *H. pluvialis* cultures ranged between 0.13 and 1.77 g L⁻¹ with most samples having a dry weight < 0.8 g L⁻¹ (Fig. S3a). Total astaxanthin content ranged from 0.01 to 1 % of dry biomass, with roughly half of the samples in the dataset presenting contents in the range of 0.01–0.2 % (Fig. S3b). Of the three identified and quantified astaxanthin isomers, all-E astaxanthin constituted on average 86 % of the total astaxanthin content, with approximately 10 % attributed to 13Z-form and ~4 % of 9Z-form (Fig. S3c). Both IS and HSI spectra presented clear differences between *H. pluvialis* cultures with increasing stress conditions (Fig. 2). Although the use of an IS allowed to obtain high spectral resolution and an overall high quality absorbance spectrum, scattering of light markedly increased in stressed cells (Fig. 2a). This led to shifts in peak maxima which was not observed in HSI spectra (Fig. S4).

3.2. Single and dual wavelength regression

Astaxanthin concentration per litre (mg L⁻¹) reflects the total volumetric amount of astaxanthin and should therefore best correlate with a single wavelength. The highest absolute Pearson correlations were observed with wavelengths close to 565 nm for both IS (568 nm, $r = 0.91$; Table S2) and HSI (564 nm, $r = 0.90$; Table S2). Linear models using these highly correlating wavelengths did not accurately predict test samples and had MAPE of 432 % and 214.3 % for IS and HSI, respectively (Fig. 3A,B; Table S3; Table S4). Since these linear models showed poor fitting at low astaxanthin (<2 mg L⁻¹), MAPE was calculated for test samples with higher concentrations (>2 mg L⁻¹; $n = 8$). MAPE_{>2} was notably lower for IS (27.4 %) and HSI (35.3 %), reflecting lower prediction error at higher astaxanthin concentrations (Table S4). However, astaxanthin concentration per litre might not be an ideal unit for culture monitoring given the effect of cell density. Therefore, an ideal model would predict the astaxanthin content per dry weight (μg mg⁻¹)

directly. Absolute Pearson correlation between a single wavelength and astaxanthin content per dry weight was highest with 560 nm for both IS and HSI ($r = 0.67$ and 0.88 , respectively; Table S2). Linear models of 560 nm and astaxanthin (μg mg⁻¹) showed a lower MAPE than previous models (145.4 % and 76 % for IS and HSI, respectively; Fig. 3C,D; Table S3; Table S4), although their performances at >2 μg mg⁻¹ were similar (MAPE_{>2} = 41.1 % and 36.6 % for IS and HSI, respectively). Lastly, the use of two wavelengths was tested using optical density at 750 nm (OD 750) as a normalizing value given its high correlation with dry weight ($r = 0.99$ for both IS and HSI; Table S3, Fig. S5). The highest Pearson correlations using this approach were observed at 497/750 for IS and 560/750 for HSI ($r = 0.84$ and 0.83 for IS and HSI, respectively; Table S2). Linear models using these dual wavelengths gave >100 % MAPE (116.3 % for IS and 120.8 % for HSI; Table S4) and similar results as the previous tested models at higher astaxanthin contents (MAPE_{>2} = 24.7 % and 35 % for IS and HSI, respectively; Fig. 3; Table S3; Table S4).

3.3. 1D convolutional neural network regression

The use of a CNN for the prediction of astaxanthin content per dry weight (μg mg⁻¹) using HSI data strongly decreased the prediction error compared to single and dual wavelength linear model (Fig. 4). Pearson correlation between expected and predicted astaxanthin content was $r = 0.99$ (Table S3). The CNN had the lowest MAPE of all studied models in this work (26.8 %) and presented a very low prediction error for astaxanthin content >2 μg mg⁻¹ (MAPE_{>2} = 5.9 %; Table S4). On average, the test samples with <0.6 μg mg⁻¹ astaxanthin (or 0.06 % of biomass) had >10 times prediction error than samples with >0.6 μg mg⁻¹ astaxanthin (Fig. 4B), suggesting that at such low astaxanthin contents accurate predictions might not be possible using HSI. PCA summarized 95 % of the variance observed between sample's spectra in the first two principal components (Fig. 5A). Position of train and test samples in the two principal component axes shows that train samples represent well the variability observed in test samples. Plotting of the augmented spectra used for model training shows a larger portion of the two-dimensional space covered (Fig. 5B), validating the use of random pixel selection to increase the variability in training data during the model training process. Shapley-based explanation of the CNN model shows that the most important wavelengths used for model prediction are in the range of 550–700 nm (Fig. 5C), with 578.28, 584.99 and 630.79 nm presenting the highest contribution to the prediction value. A closer inspection at the effect of the 20 most important features for value prediction suggest that when wavelengths >580 nm have low values they have a negative effect on the astaxanthin content while at high values they have a positive effect on the prediction value (Fig. 5D). Since

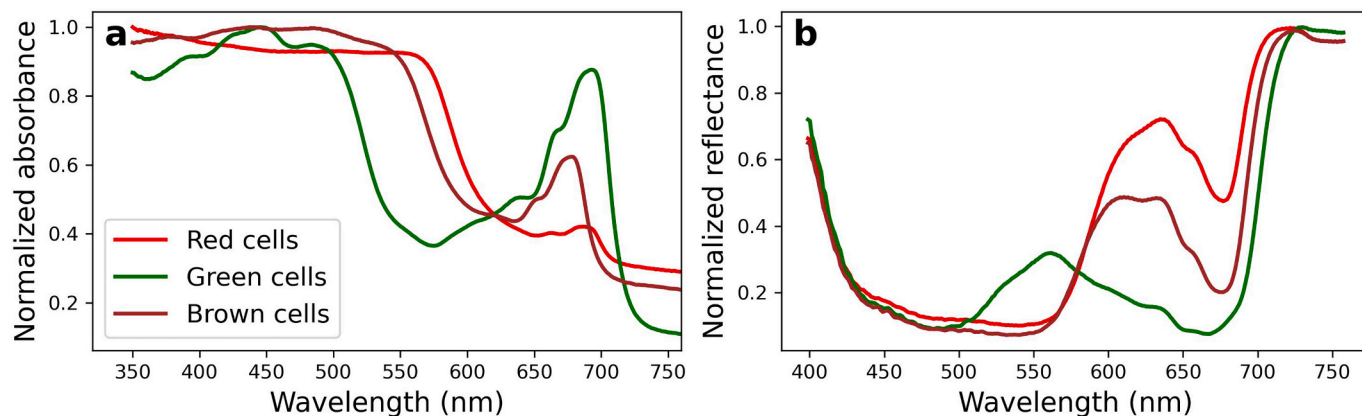


Fig. 2. Examples of normalized absorption (a) and reflectance (b) spectrum of *H. pluvialis* suspensions at different cultivation times. With longer cultivation time and stress conditions, cultivations transition from green to brown and then to red colour. Reflectance data was acquired with a hyperspectral imager while absorbance data was obtained with a spectrophotometer attached to an integrating sphere. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

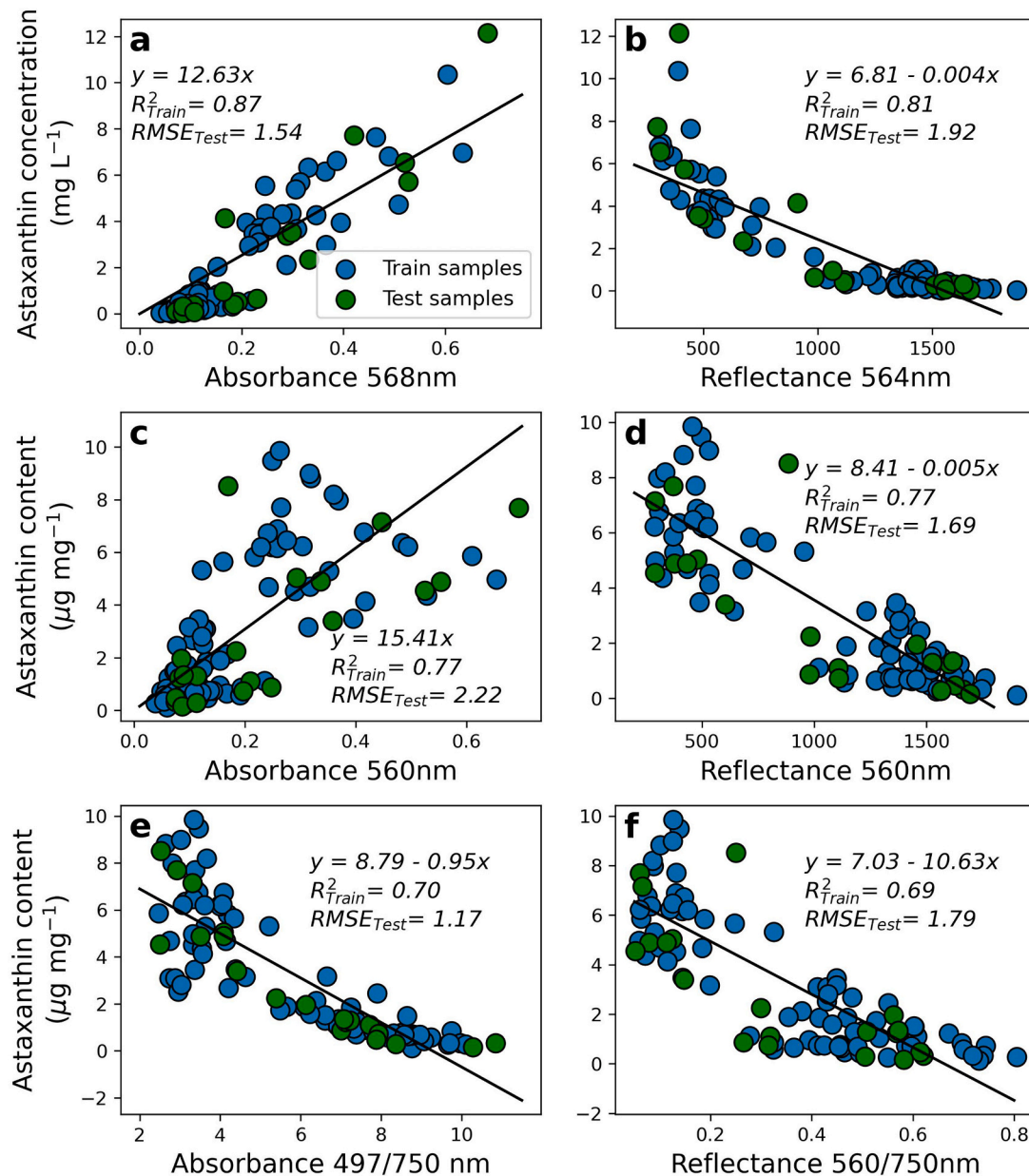


Fig. 3. Linear models describing the relationship between single (a-d) or dual (e,f) wavelengths and astaxanthin concentration per litre (a,b) or content per dry weight (d-f). Reflectance data was acquired with a hyperspectral imager while absorbance data was obtained with a spectrophotometer attached to an integrating sphere. Linear equation, train data coefficient of determination (R^2) and test data root mean squared error (RMSE) are presented in each plot.

the second absorption peak of chlorophylls are in this area of the spectrum, it can be interpreted that chlorophyll concentration is negatively related to the predicted content of astaxanthin. The opposite trend is seen for wavelengths <580 nm, suggesting that lower values in the carotenoid area of the spectra increase the predicted astaxanthin content by the CNN while higher values decrease the prediction estimate.

3.4. Varying cultivation conditions and transfer learning

To test if the CNN model generalizes to samples of *H. pluvialis* cultured under different conditions, we collected 51 new samples that include cultivations in wastewater at different CO₂ levels ($n = 42$) and three different NaCl concentrations ($n = 9$) (see [Materials and Methods, Section 2.6](#)). Model performance was lower than previously observed (MAPE = 36.6 %, MAPE_{>2} = 32.9 %; [Fig. 6A,B](#)) and a large discrepancy between prediction and HPLC values was seen for the new test samples

with high astaxanthin contents ($>8.5 \mu\text{g mg}^{-1}$). Although samples with high astaxanthin contents ($>8.5 \mu\text{g mg}^{-1}$) in the new dataset presented a broad low-reflectance signal on the carotenoid area of the spectra, they also had a lower chlorophyll red-absorbing reflectance compared to the samples used to train the CNN ([Fig. S6](#)). This suggests less chlorophyll degradation at similar and higher astaxanthin contents. Given that the CNN partially uses the chlorophyll peak to quantify astaxanthin ([Fig. 5D](#)), we tested if transfer learning can fine-tune the model for the characteristics of the new dataset. Improved prediction accuracy was observed after transfer learning (MAPE = 24.9 %, MAPE_{>2} = 8.2 %; [Fig. 6C,D](#)), suggesting that the CNN model can be adapted to deal with the variability observed between different cultivation conditions.

4. Discussion

Astaxanthin quantification from *H. pluvialis* cultures requires

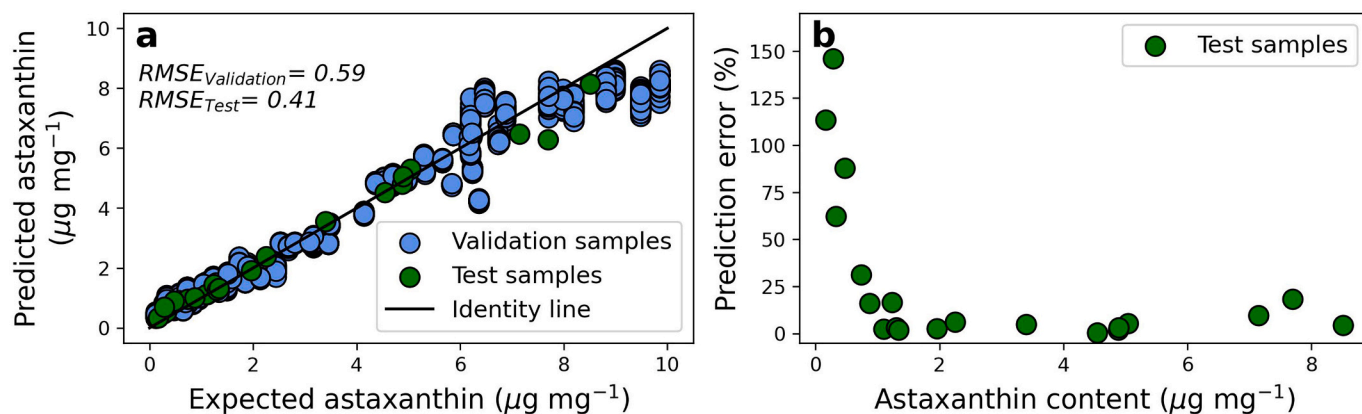


Fig. 4. Relationship between the predicted astaxanthin content ($\mu\text{g mg}^{-1}$) using a convolutional neural network (CNN) and the real values of validation and test datasets (a) together with the dependency of absolute prediction error with the astaxanthin concentration of test samples (b).

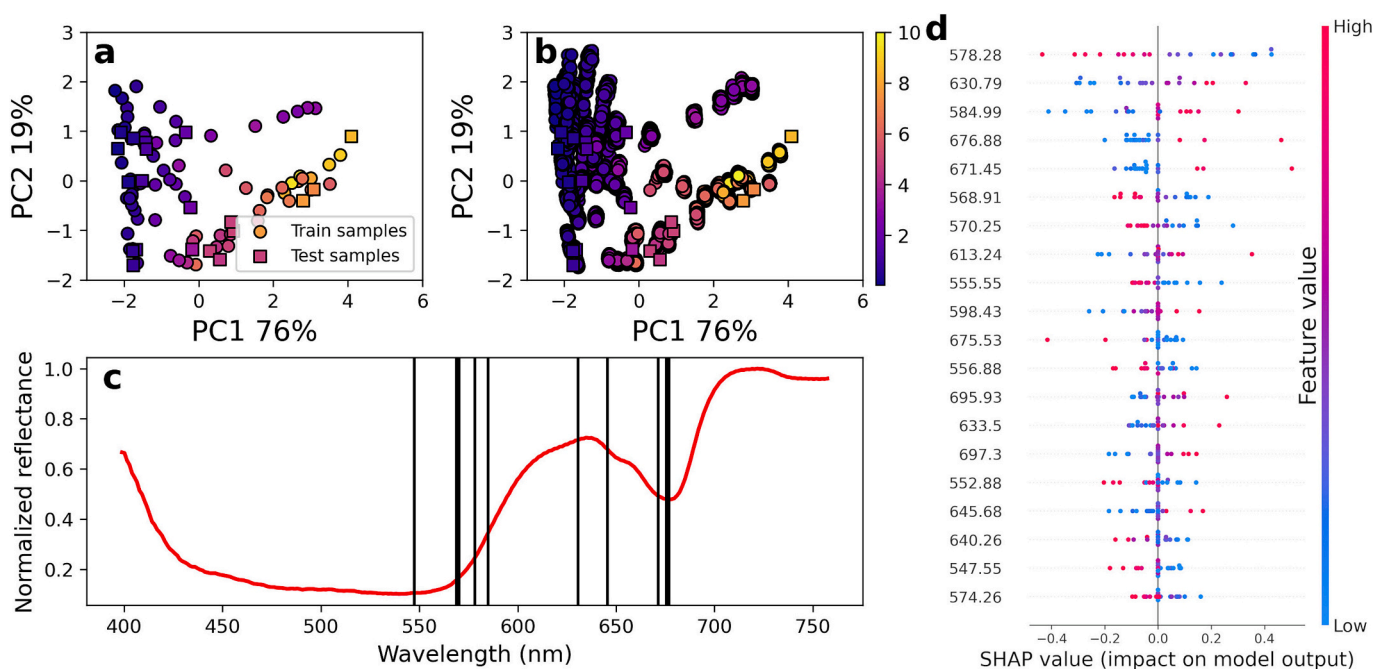


Fig. 5. Principal component analysis (PCA) of train and test samples without (a) or with augmented training data (b). Variance explained by each principal component is given as percentage in plot axes and colour coding reflects the astaxanthin content ($\mu\text{g mg}^{-1}$) of each sample as indicated in the colour bar (a, b). Most influential wavelengths (10) for prediction of astaxanthin identified by SHAP algorithm are plotted as vertical black lines on top of a normalized reflectance spectrum (red) of *H. pluvialis* suspension (c). Impact of wavelength value on model prediction value (d) of the 20 most influential wavelengths (ordered from higher to lower influence). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

laborious sample processing while accuracy at low astaxanthin contents can only be achieved using HPLC [9]. This study shows that applying machine learning algorithms to predict astaxanthin content using reflectance HSI of *H. pluvialis* suspensions is an effective approach for fast and accurate non-invasive quantification. Traditionally, spectroscopic analysis of microalgal suspensions has focussed on biomass monitoring using absorbance data (i.e., OD750) in combination with appropriate calibration curves [15]. Nevertheless, pigment-specific quantification is more challenging due to partial or total overlap of pigment absorption peaks, scattering of light by microalgal cells and shifts in pigment absorption maximum due to intermolecular interactions [16]. Our reflectance HSI data showed no major optical effects due to changes in cell density or culture conditions, although the chlorophyll absorption peak in the red region exhibited a slight rightward shift in stressed cells. In contrast, IS spectra had clear effects of light scattering in stressed cells. Stress conditions such as nitrogen

deficiency and high light, common conditions used to stimulate astaxanthin production, promote the encystment of *H. pluvialis* cells [20,36] leading to changes in cell size, chlorophyll degradation and the accumulation of secondary carotenoids in lipid droplets [20]. Such morphological changes have been shown to increase the scattering coefficient of *H. pluvialis* suspensions [37]. Despite an IS technically captures all forward-scattered light from a sample, our results show that the contribution of side- and back-scattering of red *H. pluvialis* suspensions can affect the quality of the obtained spectra leading to the overestimation of light absorption and shifts in peak absorption maxima. A way to circumvent the scattering of *H. pluvialis* in suspension is to measure the absorbance of cells upon deposition on glass fiber filters [16]. By doing so, the relationship between OD500/OD678 and total carotenoids/chlorophyll remains linear across a range of pigment concentrations. The downside of this method is that astaxanthin cannot be directly quantified and it only becomes the major carotenoid at high

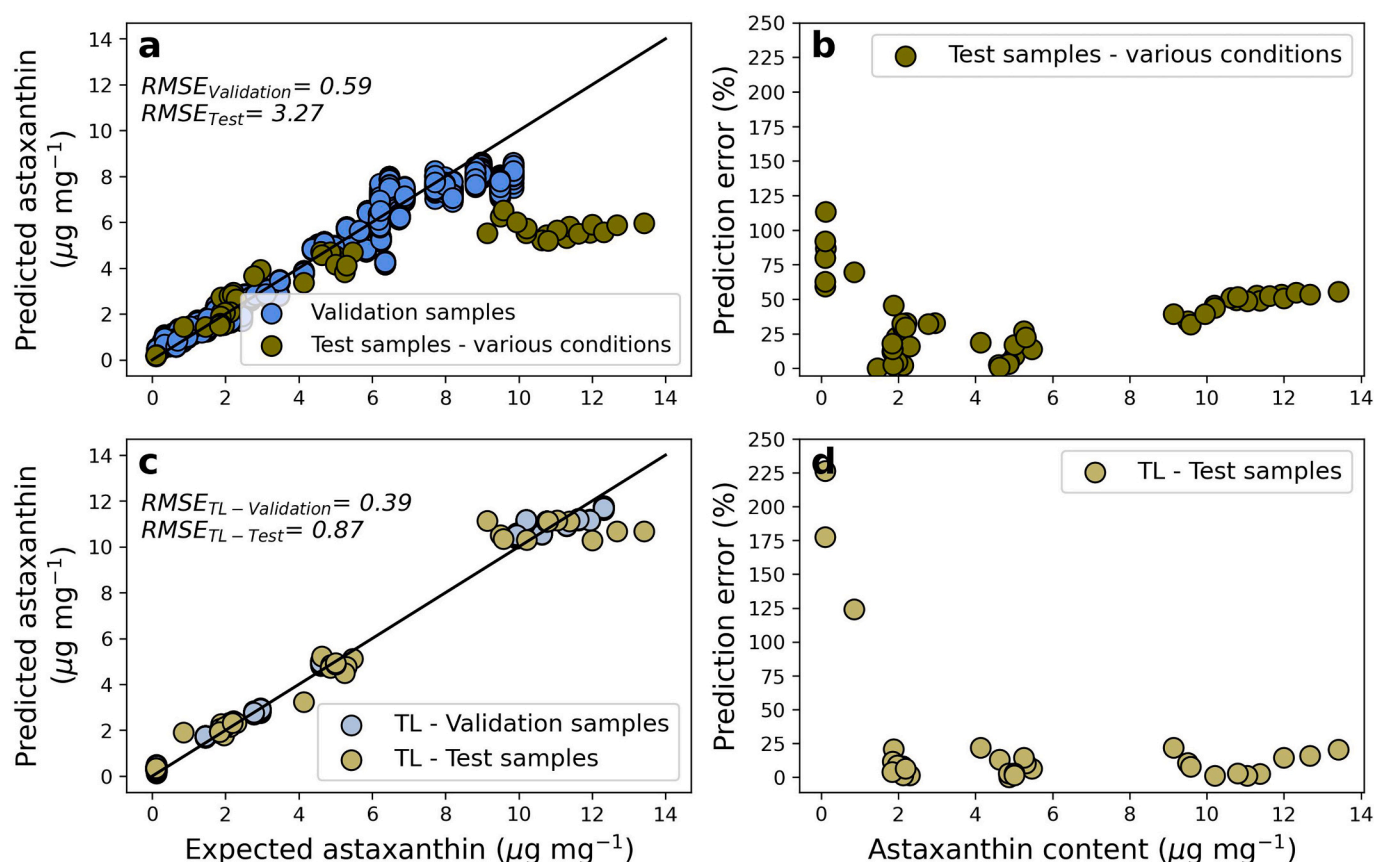


Fig. 6. Relationship between the predicted astaxanthin content ($\mu\text{g mg}^{-1}$) and the real values of validation and test datasets (a, c) using a convolutional neural network (CNN) before (a) and after (c) transfer learning (TL). Additionally, the dependency of absolute prediction error with the astaxanthin concentration of test samples (b, d) is presented for samples before (b) and after transfer learning (d). The test samples belonging to various cultivation conditions ($n = 51$; a, b) were divided into two groups to ultimately produce the augmented TL train and TL validation datasets together with TL test samples (see [Materials and Methods, Section 2.7](#)). Plots a and c, as well as plots b and d share their x axis title.

total carotenoid contents. Other non-destructive methods include total astaxanthin quantification via dry biomass near-infrared spectroscopy [38] and per cell estimation via flow cytometry [39] and Raman microspectroscopy [40,41]. These methods require more sample preparation than our approach, and it remains to be tested if they can accurately quantify *H. pluvialis* with low ($<10 \mu\text{g mg}^{-1}$) astaxanthin content. Quantification of cultures with low astaxanthin content is relevant in industrial settings because unintentional increases in astaxanthin can indicate undesired cell stress. Future research concerning techniques for non-invasive astaxanthin quantification should therefore include cultures containing low ($<10 \mu\text{g mg}^{-1}$) to high astaxanthin contents. The astaxanthin absorption peak overlaps to a high degree with other carotenoids present in *H. pluvialis* such as lutein and β -carotene [21]. Therefore, the use of higher wavelengths (i.e. 530 nm) than the absorbance maximum of astaxanthin have been suggested for the quantification of pigment extracts from *H. pluvialis* [21]. Our Pearson correlation analysis between astaxanthin concentration per litre (mg L^{-1}) or content per dry weight ($\mu\text{g mg}^{-1}$) with a single wavelength favoured the use of $\geq 560 \text{ nm}$. These results suggest that the range between 550 and 600 nm contains wavelengths where astaxanthin absorption has less overlap with other pigments in vivo. Nevertheless, precise determination of the astaxanthin absorption peak in vivo is difficult due to the presence of different astaxanthin esters [11] and the possibility of spectral changes due to strong pigment aggregation in stressed cells [16]. Prediction errors of linear models using single or dual wavelengths were high although large improvements were seen at higher astaxanthin amounts ($>2 \text{ mg L}^{-1}$ or $>2 \mu\text{g mg}^{-1}$). This difference in prediction error was expected due to the very low astaxanthin

amounts observed in some of the samples analysed. Considering that reflectance and absorption of a highly scattering suspension are not linearly related to changes in the concentration of the light absorbing analyte [26], it is likely that non-linear models would increase the prediction accuracy of the tested single and dual wavelengths. Nevertheless, since *H. pluvialis* presents different scattering coefficients at different life stages [37] and the overlap between the astaxanthin absorption peak and other carotenoids is so pronounced, it is possible that no single or dual wavelength model could produce very accurate astaxanthin estimations even when using a high-end IS.

Although CNNs were originally developed as an image classification algorithm [42], their suitability to deal with 1-dimensional spectroscopic data has also been demonstrated [23,25,28]. Our results provide further evidence that CNNs can handle the collinearity of spectroscopic data and produce accurate regression results. Estimation of astaxanthin content per dry weight had an average prediction error 3-fold lower than the best tested linear model using either HSI or IS data. Moreover, our CNN model outperformed the accuracy of spectrometric quantification of astaxanthin using pigment extracts at low astaxanthin contents [9]. This improvement in prediction error is likely explained by the capacity of the CNN to extract the important features from the HSI data and model a non-linear relationship between such features and the output. By doing so, the CNN model decreases its reliance on a single wavelength such as an absorption maximum which, as previously mentioned, can be subjected to spectral shifts due to changes in the spectral properties of the algal culture. The non-linearity of the HSI data is evidenced by our PCA results, where astaxanthin content presents a non-linear distribution in principal component axes.

As observed in the SHAP values, the CNN not only uses the wavelengths associated with astaxanthin but also the red absorption of chlorophyll to make predictions. When testing samples of *H. pluvialis* cultivated under different conditions than the one used to train the model (high light and nitrogen deficiency), this translated into lower prediction accuracy than expected. A commonality between the 51 extra tested samples was that none of those cultures were transferred into nitrogen deficient media to stimulate astaxanthin production. Nitrogen deficiency in combination with high light leads to acute chlorophyll degradation in *H. pluvialis* [43], even when compared with each of these treatments alone [43]. The higher chlorophyll reflectance observed in the samples with high astaxanthin content ($>8.5 \mu\text{g mg}^{-1}$) used to train the base model and the lower prediction accuracy observed in differently cultured samples suggest that the CNN model leverages the function that relate the astaxanthin and chlorophyll reflectance signals to improve prediction. Therefore, when the relationship between these signals is altered, the prediction accuracy can be affected. The application of transfer learning, a ML technique used to fine-tune a model to new (unseen) data, greatly improved the prediction error of the CNN within the new test samples. Importantly, no major differences in prediction accuracy were observed between the different treatments of the extra tested samples, indicating that the CNN model may possess generalization capabilities under comparable stress conditions. Overall, our results suggest that the best course of action is to train a CNN model with samples that are cultured under similar conditions to the samples that need to be quantified. By doing so, the prediction variability associated with changes in the relationship between astaxanthin and chlorophyll (given by differences in treatment conditions) can be avoided. Furthermore, since there are several strains of *H. pluvialis* with potential for astaxanthin production, they likely present variability in the relationship between astaxanthin and chlorophyll compared to the strain used in this study. Therefore, to obtain the samples for model training not only culture conditions should be considered but also the *H. pluvialis* strain being cultivated.

Common downsides to the use of CNNs are the large amounts of data and computing power required for model training. In this study, we overcame these limitations by augmenting our training datasets and utilizing a custom CNN architecture that significantly decreases the number of trainable parameters. Overall, our data augmentation strategy combined with the use of random dropout of neurons during training and transfer learning resulted in good performance of the CNN model despite the low number of independent training samples. Limitations to our study include the ranges of cell density and astaxanthin content of our samples. Unfortunately, higher biomasses or astaxanthin content per dry weight could not be obtained under our laboratory conditions. More research on the ability of HSI and CNN to predict astaxanthin content under relevant commercial conditions is needed.

5. Conclusions

Reflectance HSI in combination with a CNN offers a fast and cost-effective non-invasive monitoring approach to astaxanthin content quantification of *H. pluvialis* suspensions. The low prediction error observed in this study when the CNN predicted samples in similar cultivation conditions to the training data was unmatched by simple linear models utilizing one or two wavelength bands of the spectral data, even when using a high-end integrating sphere for absorbance spectrum acquisition. Data augmentation and a simplified CNN architecture allowed to obtain accurate predictions without the need for large amounts of training data or computing power. Fine-tuning of the model with samples from the same distribution as the test samples is required to obtain the highest possible performance. Future research on the detection limits of HSI of *H. pluvialis* commercial cultures will be key to define if the approach used in this study can translate to the industrial scale production of astaxanthin.

CRedit authorship contribution statement

Marco L. Calderini: Writing – review & editing, Writing – original draft, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Salli Pääkkönen:** Writing – review & editing, Methodology. **Aliisa Yli-Tuomola:** Writing – review & editing, Methodology. **Hemanta Timilsina:** Writing – review & editing, Methodology. **Katja Pulkkinen:** Writing – review & editing, Resources, Funding acquisition. **Ilkka Pölönen:** Writing – review & editing, Resources. **Pauliina Salmi:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Funding sources

This work was funded by the European Union – NextGenerationEU and the Research Council of Finland research grants awarded to P.S. (grant no. 352764) and K.P. (grant no. 352765) and by the Department of Biological and Environmental Science, University of Jyväskylä (support given to K.P.).

Declaration of competing interest

The authors declare that they do not have competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank laboratory technicians Mervi Koistinen and Emma Pajunen and laboratory engineer Hannu Pakkanen for their help during the experimental work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.algal.2025.103979>.

Data availability

All data and data analysis files necessary to replicate the results presented in this study are available at <https://doi.org/10.23729/be85c733-2555-45b9-8bca-5fe441beb48b>. A comprehensive description of the content of each file can be found in the same DOI under the name “Metadata file.”

References

- [1] A. Kuech, M. Breuer, I. Popescu, Research for PECH committee – the future of the EU algae sector, in: European Parliament, Policy Department for Structural and Cohesion Policies, Brussels, 2023.
- [2] Q. Li, L. Li, Y. Zhang, H. Gao, Y. Zhao, X. Yu, Chemical inducers regulate ROS signalling to stimulate astaxanthin production in *Haematococcus pluvialis* under environmental stresses: a review, *Trends Food Sci. Technol.* 136 (2023) 181–193.
- [3] X. Liu, T. Shibata, S. Hisaka, T. Osawa, Astaxanthin inhibits reactive oxygen species-mediated cellular toxicity in dopaminergic SH-SY5Y cells via mitochondria-targeted protective mechanism, *Brain Res.* 13 (1254) (2009) 18–27.
- [4] Y. Xue, Z. Qu, J. Fu, J. Zhen, W. Wang, Y. Cai, W. Wang, The protective effect of astaxanthin on learning and memory deficits and oxidative stress in a mouse model of repeated cerebral ischemia/reperfusion, *Brain Res. Bull.* 131 (2017) 221–228.
- [5] X. Liu, J. Xie, L. Zhou, J. Zhang, Z. Chen, J. Xiao, H. Xiao, Recent advances in health benefits and bioavailability of dietary astaxanthin and its isomers, *Food Chem.* 404 (B) (2023) 134605.
- [6] R.T. Lorenz, G.R. Cysewski, Commercial potential for *Haematococcus* microalgae as a natural source of astaxanthin, *Trends Biotechnol.* 18 (4) (2000) 160–167.
- [7] V.C. Liyanaarachchi, G.K. Nishshanka, R.G. Premaratne, T.U. Ariyadasa, P. H. Nimarshana, A. Malik, Astaxanthin accumulation in the green microalga *Haematococcus pluvialis*: effect of initial phosphate concentration and stepwise/continuous light stress, *Biotechnol. Rep.* 28 (2020) e00538.
- [8] B. Subramanian, N. Tchoukanova, Y. Djaoued, C. Pelletier, M. Ferron, J. Robichaud, Investigations on the geometrical isomers of astaxanthin: Raman spectroscopy of conjugated polyene chain with electronic and mechanical

- confinement: investigations on the geometrical isomers of astaxanthin, *J. Raman Spectrosc.* 45 (2014) 299–304.
- [9] I.K. Koopmann, A. Kramer, A. Labes, Development and validation of reliable astaxanthin quantification from natural sources, *PLoS One* 17 (12) (2022) e0278504.
 - [10] F. Miao, D. Lu, Y. Li, M. Zeng, Characterization of astaxanthin esters in *Haematococcus pluvialis* by liquid chromatography–atmospheric pressure chemical ionization mass spectrometry, *Anal. Biochem.* 352 (2006) 176–181.
 - [11] C. Grewe, C. Griehl, Time- and media-dependent secondary carotenoid accumulation in *Haematococcus pluvialis*, *Biotechnol. J.* 3 (2008) 1232–1244.
 - [12] A. Solovchenko, Seeing good and bad: optical sensing of microalgal culture condition, *Algal Res.* 71 (2023) 103071.
 - [13] I. Havlik, S. Beutel, T. Scheper, K.F. Reardon, On-line monitoring of biological parameters in microalgal bioprocesses using optical methods, *Energies* 15 (2022) 875.
 - [14] J.Y. Liu, L.H. Zeng, Z.H. Ren, The application of spectroscopy technology in the monitoring of microalgae cells concentration, *Appl. Spectrosc. Rev.* 56 (3) (2020) 171–192.
 - [15] M.J. Griffiths, C. Garcin, R.P. van Hille, S.T. Harrison, Interference by pigment in the estimation of microalgal biomass concentration by optical density, *J. Microbiol. Methods* 85 (2) (2011) 119–123.
 - [16] A. Solovchenko, C. Aflalo, A. Lukyanov, S. Boussiba, Nondestructive monitoring of carotenogenesis in *Haematococcus pluvialis* via whole-cell optical density spectra, *Appl. Microbiol. Biotechnol.* 97 (2013) 4533–4541.
 - [17] R.J. Ritchie, S. Sma-Air, Solvent-free chlorophyll spectrometry in unicellular algal research, *J. Appl. Phycol.* 32 (2020) 2711–2723.
 - [18] S. Boussiba, A. Vonshak, Astaxanthin accumulation in the green alga *Haematococcus pluvialis*, *Plant Cell Physiol.* 32 (1991) 1077–1082.
 - [19] M. Buchwald, W.P. Jencks, Optical properties of astaxanthin solutions and aggregates, *Biochemistry* 7 (1986) 834–843.
 - [20] S. Ota, A. Morita, S. Ohnuki, A. Hirata, S. Sekida, K. Okuda, Y. Ohya, S. Kawano, Carotenoid dynamics and lipid droplet containing astaxanthin in response to light in the green alga *Haematococcus pluvialis*, *Sci. Rep.* 8 (2018) 5617.
 - [21] P. Casella, A. Iovine, S. Mehariya, T. Marino, D. Musmarra, A. Molino, Smart method for carotenoids. Characterization in *Haematococcus pluvialis* red phase and evaluation of astaxanthin thermal stability, *Antioxidants* 9 (2020) 1–17.
 - [22] T.E. Murphy, K. Macon, H. Berberoglu, Multispectral image analysis for algal biomass quantification, *Biotechnol. Prog.* 29 (2013) 808–816.
 - [23] P. Salmi, M. Calderini, S. Pääkkönen, S. Taipale, I. Pölönen, Assessment of microalgae species, biomass, and distribution from spectral images using a convolution neural network, *J. Appl. Phycol.* 34 (2022) 1565–1575.
 - [24] P. Salmi, M.A. Eskelinen, M.T. Leppänen, I. Pölönen, Rapid quantification of microalgae growth with hyperspectral camera and vegetation indices, *Plants* 10 (2021) 341.
 - [25] S. Pääkkönen, I. Pölönen, A.M. Raita-Hakola, M. Carneiro, H. Cardoso, D. Mauricio, A.M. Cavaco-Rodrigues, P. Salmi, Non-invasive monitoring of microalgae cultivations using hyperspectral imager, *J. Appl. Phycol.* 36 (2024) 1653–1665.
 - [26] C.M. Gardner, Transmission versus reflectance spectroscopy for quantitation, *J. Biomed. Opt.* 23 (1) (2018) 018001.
 - [27] P. Ghamisi, Y. Chen, X.X. Zhu, A self-improving convolution neural network for the classification of hyperspectral data, *IEEE Geosci. Remote Sens. Lett.* 13 (10) (2016) 1537–1541.
 - [28] S. Malek, F. Melgani, Y. Bazi, One-dimensional convolutional neural networks for spectroscopic signal regression, *J. Chemometr.* 32 (2018) e2977.
 - [29] J. Pyo, H. Duan, S. Baek, M.S. Kim, T. Jeon, Y.S. Kwon, H. Lee, K.H. Cho, A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery, *Remote Sens. Environ.* 233 (2019) 111350.
 - [30] P. Salmi, I. Pölönen, D.A. Beckmann, M.L. Calderini, L. May, J. Olszewska, L. Perozzi, S. Pääkkönen, S. Taipale, P. Hunter, Resolving phytoplankton pigments from spectral images using convolutional neural networks, *Limnol. Oceanogr. Methods* 22 (2024) 1–13.
 - [31] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, M. Azim, Transfer learning: a friendly introduction, *J. Big Data* 9 (2022) 102.
 - [32] R.R.L. Guillard, C.J. Lorenzen, Yellow-green algae with chlorophyllide, *Phycol.* 8 (1972) 10–14.
 - [33] Č. Stević, K. Pulkkinen, J. Pirhonen, Screening of microalgae and LED grow light spectra for effective removal of dissolved nutrients from cold-water recirculating aquaculture system (RAS) wastewater, *Algal Res.* 44 (2019) 101681.
 - [34] J.T. Springenberg, A. Dosovitskiy, T. Brox, M.A. Riedmiller, Striving for Simplicity: The All Convolutional Net, *CoRR* (2014) abs/1412.6806.
 - [35] Q. Li, J. You, T. Qiao, D. Zhong, X. Yu, Sodium chloride stimulates the biomass and astaxanthin production by *Haematococcus pluvialis* via a two-stage cultivation strategy, *Bioresour. Technol.* 344 (2022) 126214.
 - [36] M. Kobayashi, Y. Kurimura, T. Kakizono, N. Nishio, Y. Tsuji, Morphological changes in the life cycle of the green alga *Haematococcus pluvialis*, *J. Ferment. Bioeng.* 84 (1997) 94–97.
 - [37] F. Sheng, J. Fan, W. Huang, J. Bai, S. Wang, W. Li, M. Li, Y. Wan, Investigation on models for light distribution of *Haematococcus pluvialis* during astaxanthin accumulation stage with an application case, *Algal Res.* 33 (2018) 182–189.
 - [38] J.H. Liu, L. Song, Q. Huang, Rapid screening astaxanthin-hyperproducing *Haematococcus pluvialis* mutants through near-infrared spectroscopy, *Lett. Appl. Microbiol.* 62 (1998) 185–191.
 - [39] J. Chen, D. Wei, G. Pohnert, Rapid estimation of astaxanthin and the carotenoid-to-chlorophyll ratio in the green microalga *Chromochloris zofingiensis* using flow cytometry, *Mar. Drugs* 15 (2017) 231.
 - [40] Y. Shao, W. Gu, L. Jiang, Y. Zhu, A. Gong, Study on the visualization of pigment in *Haematococcus pluvialis* by Raman spectroscopy technique, *Sci. Rep.* 9 (2019) 12097.
 - [41] K. Li, J. Cheng, Q. Ye, Y. He, J. Zhou, K. Cen, In vivo kinetics of lipids and astaxanthin evolution in *Haematococcus pluvialis* mutant under 15% CO₂ using Raman microspectroscopy, *Bioresour. Technol.* 244 (2017) 1439–1444.
 - [42] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
 - [43] L. Scibilia, L. Girolomoni, S. Berteotti, A. Alboresi, M. Ballottari, Photosynthetic response to nitrogen starvation and high light in *Haematococcus pluvialis*, *Algal Res.* 12 (2015) 170–181.