*CASE REPORT*

**Annals of Agricultural and Environmental Medicine 2025, Vol 32, No 1, 157–162**
www.aaem.pl



# Application of automatic image analysis using a Deep Learning Neural Network for assessing the growth of green algae containing carotenoids – importance for environment, health and aquaculture

**Monika M. Zdeb[1,B-D,F] ✉, Mateusz Walo[2,B-D], Grzegorz Łagód[3,A,D,F]**

[1] *Department of Water Purification and Protection, Rzeszów University of Technology, Rzeszów, Poland*
[2] *Department of Applied Mathematics, Faculty of Mathematics and Information Technology, Lublin University of Technology, Lublin, Poland*
[3] *Department of Water Supply and Sewage Disposal, Lublin University of Technology, Lublin, Poland*
*A – Research concept and design, B – Collection and/or assembly of data, C – Data analysis and interpretation, D – Writing the article, E – Critical revision of the article, F – Final approval of the article*

Zdeb MM, Walo M, Łagód G. Application of automatic image analysis using a Deep Learning Neural Network for assessing the growth of green algae containing carotenoids – importance for environment, health and aquaculture. Ann Agric Environ Med. 2025; 32(1): 157–162. doi: 10.26444/aaem/202673

## Abstract

Using deep learning and neural networks enables us to greatly speed-up quantitative studies and provide a useful tool for analyzing microscopic images. Studies conducted on selected algae *Haematococcus* and *Coelastrum* sp. confirm the feasibility of using the deep learning neural network. The confusion matrix demonstrated the numbers of errors generated by the YOLO v8 network in relation to the validation dataset. It indicated a higher number of errors in the detection of *Haematococcus* than *Coleastrum*. The F1 score, as the harmonic mean of precision and recall, is significantly higher for the class *Coelastrum* sp. than for *Haematococcus* sp. Machine learning can be applied not only to the detection of individual cells, but also to the detection of colonies over a wide range of sizes. This article discussed the technical and practical aspects of implementing these advanced methods and highlighted their importance in the aquaculture, food, medical, sustainable energy, and environmental sectors.

## Key words

carotenoids, green algae, automatic image analysis, *Haematococcus*, *Coelastrum*, Deep Learning Neural Network, bioindication, aquaculture

## INTRODUCTION

Various forms of freshwater and marine algae are used on a large scale worldwide. They are applied not only for consumption and industrial purposes, but also for bioindication. Algae are used in wastewater treatment technologies and in systems for removing nutrients [1, 2]. They have also been used to neutralize difficult-to-decompose substances, such as those found in landfill leachates [3]. Their greatest advantage, however, is $CO_2$ binding (about 2 kg of $CO_2$ is needed to produce 1 kg of dry matter of algae) and hence the great interest in using them to reduce gaseous pollutant emissions from coal-fired boilers [4]. They are also tested as a filling element in biofilters for flue gas purification and reduction of $CO_2$ emissions into the atmosphere by industrial plants [5]. At the same time, algal biomass can be used to produce biofuels, such as biohydrogen, biomethane, biodiesel or biogas [6, 7]. Moreover, algae – as useful bioindicators – are widely used in assessing changes in the ecological status of surface waters. They are an element of the saprobe system, the Phytoplankton Index for Polish Lakes, or form the basis of the diatom index. As a lichen component, they are indicators of the degree of atmospheric air pollution [8]. Balanced nutrition and pharmacological support in the fight against lifestyle diseases are also extremely important.

Traditionally and culturally in Asian countries, and currently also in highly developed countries, these plants play an important role as a food product. The latest dietary trends indicate health benefits resulting from the use of species, such as spirulina, nori, chlorella, kombu and wakame in the diet. Dried *Euglena gracilis* algae have been included in the EU list of novel foods [9]. Specially targeted cultures are intended for the cosmetics and pharmaceutical industries. Algae cultivation allows the use of not only the biomass itself, but also the products of algal cell metabolism, an example of which is the extraction of astaxanthin from green algae [10, 11]. Astaxanthin is a component of many dietary supplements, as it has been shown to have strong antioxidant, cardioprotective, antithrombotic properties, improving the elasticity of blood vessel walls, reducing the level of pro-inflammatory cytokines in neurophiles, preventing, e.g., atherosclerosis, and stimulating the immune system, which is important in inhibiting the growth of cancers [12]. It is estimated that up to 95% of the product currently on the market comes from chemical synthesis. Although it dissolves better in water, it is less stable; therefore, natural astaxanthin is more valuable and desirable. Most of the natural compound

✉ Address for correspondence: Monika M. Zdeb, Department of Water Purification and Protection, Rzeszów University of Technology, al. Powstańców Warszawy 6, 35-959 Rzeszów, Poland
E-mail: mzdeb@prz.edu.pl

Received: 15.01.2025; accepted: 06.03.2025; first published: 24.03.2025

is obtained from *Hematococcus pluvialis*, which synthesizes it in the amount of up to 4–7% of dry mass [13].

Control of the efficiency and purity of algae production for industrial or food purposes or analysis of species composition, as well as the number of individual taxa for bioindication purposes, is still based on microscopic observations. This requires time-consuming work of qualified employees, often with a biological education and at least several years of experience in observing and identifying organisms. To enhance and automate this process, the implementation of automatic image analysis techniques utilizing deep neural networks is recommended. Deep learning, which employs neural networks, represents a more sophisticated and comprehensive evolution of traditional machine learning. The latter typically revolves around classification and regression tasks, employing simpler models, including support vector machines, random forests, decision trees, etc. The core principle of deep learning, which falls under the umbrella of artificial intelligence (AI), is to train a neural network using available data, enabling it to perform tasks in a timely manner and with the highest level of efficiency [14].

This study was undertaken to improve the process of assessing the development of algal biomass, and to be the first step towards complete automation of the aquaculture microalgae cultivation process for use in food, pharma, energy and fuel industries, as well as in bioindication studies on the degree of degradation of surface waters or during monitoring of self-purification processes. For this purpose, research was undertaken on two selected subjects – species of freshwater algae used widely in the pharmaceutical and food industry and in bioindication, using deep learning techniques – to train the YOLO v8 network for automatic object detection.

## MATERIALS AND METHOD

Two types of freshwater algae were selected as the study subjects: *Hematococcus pluvialis* and *Coelastrum sp.*, obtained from periphyton samples collected in 2024 in the Bystrzyca River, near the city of Lublin in eastern Poland. Images were taken using an Olympus CX31 microscope (40x magnification) and an image acquisition kit: uEye Cockpit software and a camera. Algal cells were identified using keys for identifying plants and plankton [15].

*Hematococcus pluvialis* is a unicellular alga with spherical, free-swimming, biflagellate cells with a diameter of about 30 µm. It occurs in the form of a vegetative green cell or as red cysts, strongly accumulating astaxanthin, under stress conditions [15, 13] (Fig. 1).

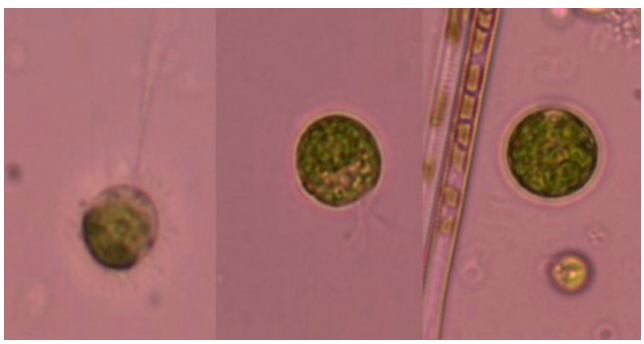

**Figure 1.** Instances of digital images featuring *Hemoatococcus pluvialis* utilized for the training, validation, and testing of the YOLO v8 network

*Coelastrum sp.* is a genus of green algae (Fig. 2). Spherical cells form spherical, less frequently angular cenobia with the number of cells 4, 8, 16, 32 or 64 (exceptionally 128), and a size of up to approx. 100 µm. In some situations, cells may also occur singly. Cells in cenobia are packed quite tightly, but with free spaces between cell walls. This genus occurs commonly in the phytoplankton of mesotrophic and eutrophic lakes in all climatic zones [16]. For this genus of green algae, a genetic mutation of a new strain of *Coelastrum* called *Coelastrum sp. TISTR 9501RE* was identified, which also has the ability to produce carotenoids, mainly astaxanthin, canthaxanthin and lutein, under environmental stress conditions [17].

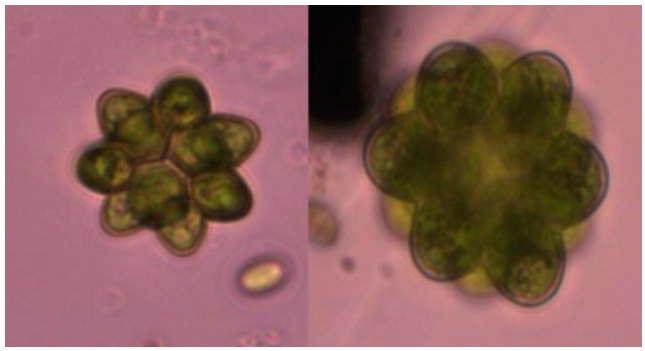

**Figure 2.** Instances of digital images featuring *Coelastrum sp* utilized for training, validation, and testing of the YOLO v8

**Software and procedures used for deep learning of neural networks.** The dataset annotation process was conducted using the Roboflow tool, fully compatible with Ultralytics, the distributor of all YOLO networks. The process utilized a functional API (Application Progranning Interface) requiring the user's private key. After accurately annotating *Haematococcus* and *Coleastrum* objects with bounding boxes, the entire dataset was split into 3 subsets: training, validation, and testing, in proportions of 7:2:1. This division ensured an adequate amount of data for both training and model evaluation. Given the limited number of samples (about 1,000 for each species), this split was deemed more reasonable than a 6:2:2 proportion, which is more applicable to datasets exceeding 10,000 microscopic images. The next step involved training the YOLO v8 network to address the object detection task. The training process lasted for 100 epochs, with validation performed after each epoch, enabling real-time monitoring of model performance, including metrics and loss functions. Google Colab, a free cloud-based computing tool, was utilized for training, allowing the avoidance of local hardware resource strain. The computations were accelerated using the Tesla T4 GPU provided by Google Colab. The training was conducted in a Python 3.8 environment using PyTorch version 1.7, which serves as the underlying framework for YOLO v8. This framework handles image processing and executes complex operations on large matrices. The training process took approximately one hour. YOLO v8 learning hyperparameters: batch size 32, learning rate 0,001. After training, the model was switched to validation mode, enabling it to fine-tune its detection capabilities by correcting previous errors and oversights based on the validation set. Finally, the model underwent evaluation, where final predictions were made, and the values of the loss function and selected metrics were assessed. Metrics, such as Accuracy, Recall, and Precision, were used to evaluate the model's performance.

## RESULTS AND DISCUSSION

After training YOLO v8 for 100 epochs, the YOLO v8 network was evaluated on the validation part of datasets. Results of the experiment are presented in Figs. 3–4.

The confusion matrix presents the number of errors made by the YOLO v8 network on the validation dataset. In this case, the neural network made more errors when detecting *Haematococcus* compared to *Coleastrum* (Fig. 3). Specifically, the total number of misclassifications for *Haematococcus* was 46, while for *Coleastrum* it was 7. Considering the proportion of errors relative to the total number of objects in the validation dataset, the network produced incorrect predictions for only 0.03 of all *Coleastrum* instances. In contrast, for *Haematococcus*, the error rate was higher, with 0.14 of all *Haematococcus* objects being misclassified.

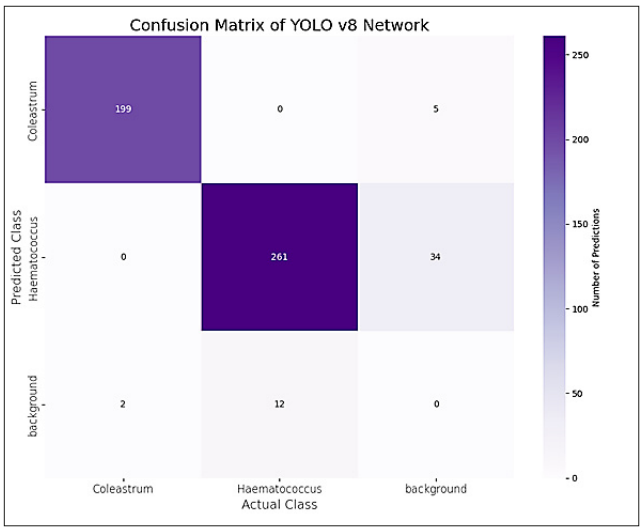


**Figure 3.** Confusion matrix for the YOLO v8

Using this confusion matrix, one can calculate performance metrics such as Accuracy, Precision, and Recall. The formulas for the mentioned metrics are as follows:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$Precision = TP/(TP + FP) \quad (2)$$

$$Recall = TP/(TP + FN) \quad (3)$$

$$F1\ score\ F_1 = \left(\frac{2}{recall^{-1} + precision^{-1}}\right) = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

(TP – true positive (result), TN – true negative (result), FP – false positive, FN – false negative).

The F1 score as the harmonic mean of precision and recall is significantly higher for the class *Coelastrum sp.* (0.9827) than for *Haematococcus sp.* (0.9190) (Tab. 1). The F1 score for the objects studied was >0.9. This is promising and indicates that the detection and classification of these objects can be supported.

**Table 1.** Performance metrics of YOLO v8 regarding localized and identified *Coleastrum* and *Haematococcus* species

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| *Coleastrum* | 0.9660 | 0.9751 | 0.9900 | 0.9827 |
| *Haematococcus* | 0.8501 | 0.8847 | 0.9560 | 0.9190 |

The provided plots illustrate the performance of the YOLO v8 model during training and validation over 100 epochs (Fig. 4).

The loss metrics consistently decrease, indicating effective learning. The box loss, which measures the accuracy of bounding box localization, steadily declines in both training and validation datasets, showing that the model becomes better at object localization, during subsequent epochs. The small gap between training and validation classification losses suggests good generalization. The distribution focal loss, which assesses the confidence of predictions, also converges smoothly, with validation closely tracking training, indicating stable performance without over-fitting. In terms of metrics, Precision stabilizes at a high value of approximately 0.95, demonstrating the model's ability to minimize false positives. Recall, which measures the model's ability to detect true positives, improves steadily and converges near 0.93, reflecting strong detection capabilities. The mean Average

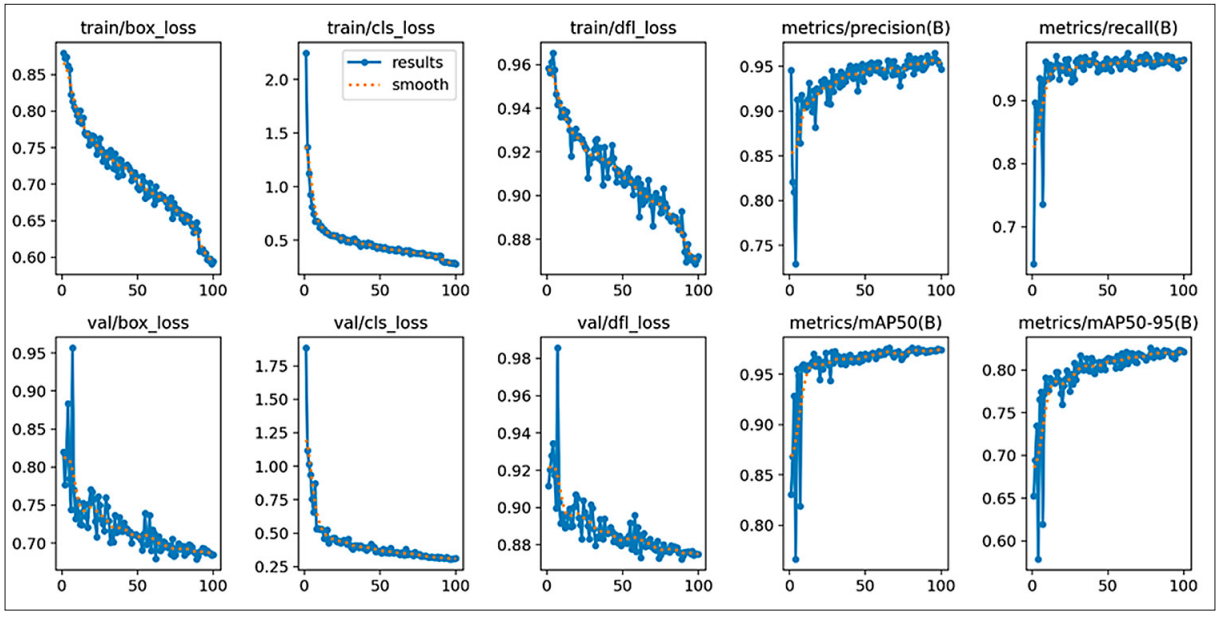


**Figure 4.** Performance of the YOLO v8 model during training and validation over 100 epochs

Precision (mAP) metrics further confirm this performance. The mAP@50 metric, which evaluates the model's accuracy at an Intersection over Union (IoU) threshold of 50%, stabilizes at about 0.95, indicating very good detection accuracy under lenient conditions. Meanwhile, the stricter mAP@50–95 metric, which averages performance across multiple IoU thresholds, converges at around 0.81, showcasing the model's robustness across various levels of overlap between predicted and ground truth bounding boxes.

Due to the fact that the sizes of the bounding boxes in the predicted objects are close to circular, the width-to-height ratio follows a distribution that is approximately normal, which simplifies its prediction (Fig. 5).

In the other 2 cases, a slight skew in the distributions can be observed, resulting from the asymmetry of certain objects. The marginal distributions show that the *x* and *y* coordinates follow an approximately normal distribution, indicating that object centres tend to cluster near the cenre of the space. The width and height distributions are more concentrated, suggesting less variability in object dimensions, compared to their positions. The scatter plot in-between reveals a clear linear correlation, implying proportionality in object dimensions. There is no noticeable correlation between the location and the size of the objects, indicating that object dimensions are independent of their spatial position. Both the width and height distributions show slight skewness, likely due to some objects being asymmetrical (Fig. 6).

Attempts are also being undertaken to use YOLO networks in the field of marking and counting microscopic organisms.

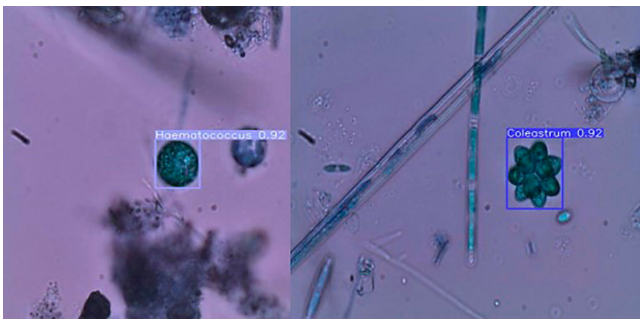


**Figure 6.** Correctly detected objects: A) *Hematococcus pluvialis*, B) *Coelastrum sp.*

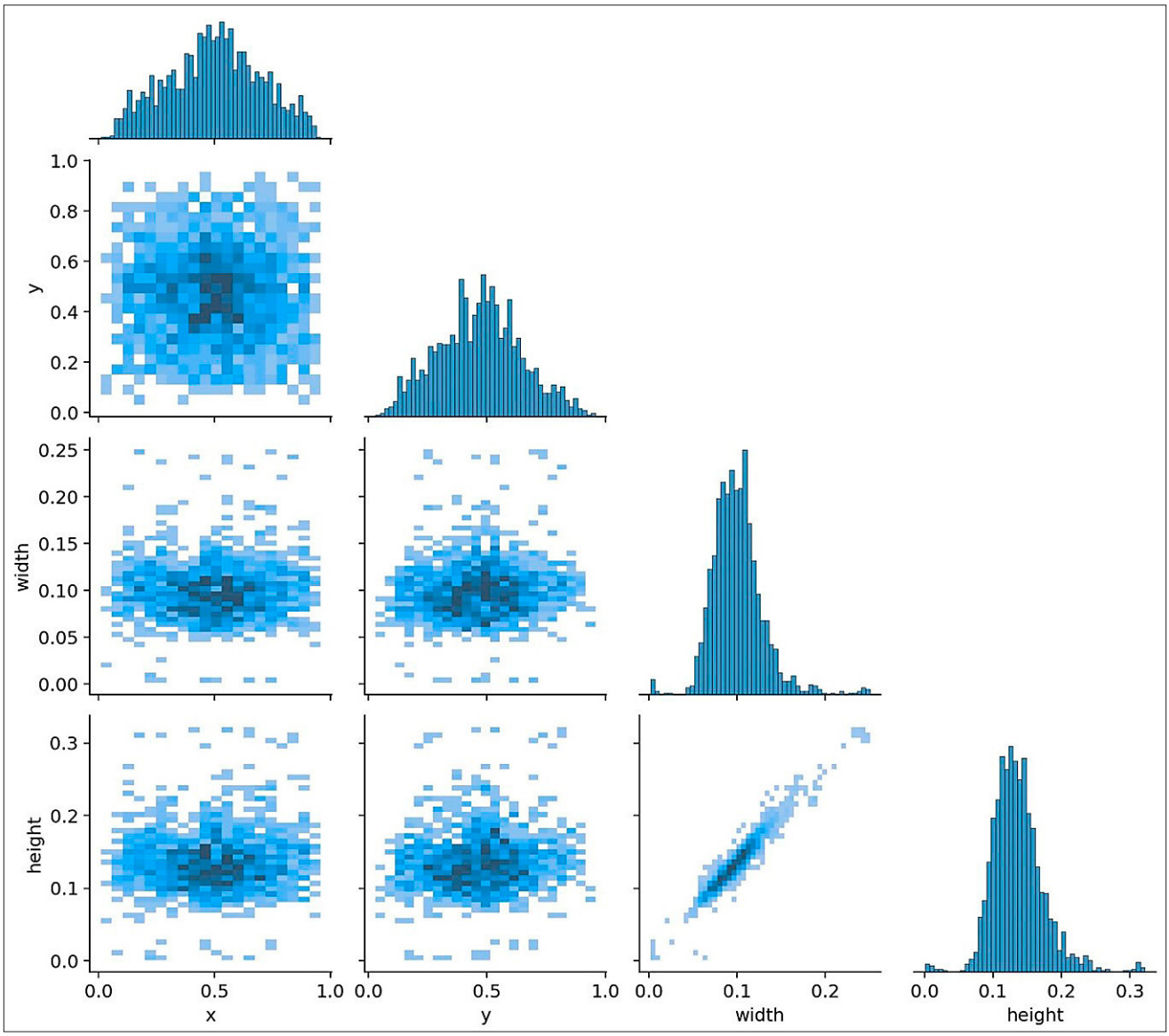


**Figure 5.** Distribution of height and weight of rectangles following the predictions made by YOLO v8

They mainly concern the possibility of using neural networks in bioindication to mark indicator organisms [18, 19], but they are also made in cosmetology to detect bacteria in products [20], in population ecology [21], or in water resources management, water safety, and rapid response to algal blooms. Research on microalgae using machine learning techniques offers significant improvements in environmental monitoring and monitoring of biomass production, which will improve the strategies to mitigate ocean acidification or commercial biofuel production [22].

The YOLO v8 network exhibited strong capabilities in identifying *Haematococcus pluvialis* and *Coelastrum sp.* objects, most likely because of the spherical-shaped cells or cell colonies. Generally, objects characterized by a round shape and a clear, compact outline are more easily recognized and classified accurately by neural networks. YOLOv8 can sometimes have difficulty correctly detecting objects with irregular, asymmetrical shapes the height-to-width distributions of which do not approximate a normal distribution, due to the loss fof unction of the YOLO network, and the bounding-box determination mechanism. An example corresponds to the differences in the evaluation of such metrics as precision, accuracy between diatoms with more or less spindly shapes and the more varied shapes of sedentary ciliates [18, 19].

The method employed by YOLOv8 to divide the image into a grid facilitates the prediction of the bounding-box centre for symmetrical objects, for mathematical reasons such as coordinate geometry, thereby improving evaluations of YOLO v8 grid metrics. However, despite the circular shape, which facilitates object recognition, slightly worse results were obtained for the presented algae *Hematococcus pluvialis* and *Coelastrum sp.* than for round, flat, coin-like objects [23]. This happened in the case of similar-sized sets of training, testing and validation images for the analyzed organism representatives, where both for the 2 carotenoid-containing algae species described in this paper and the species analyzed in the cited works, the total number of objects (labelled cases) used for teaching, testing and validating models, was about 1,000.

The size of cells within the same species is also an important element to consider in the context of the object recognition process. If microalgae cells were of identical size, the image annotation would be more precise. As a result, the obtained model would have a better chance of achieving higher detection accuracy. However, unicellular or colonial algae in nature differ significantly in terms of size, for example, *Haematococcus pluvialis* (~30 μm) [12], *Scenedesmus* (12–24 μm) [24] and *Coelastrum* sp. colonies (up to 50–100 μm) [14]. Extreme differences in microalgae morphology and size pose a serious challenge in accurate detection of microalgae. Additional background interference is also important, especially in environmental samples examined for monitoring and bioindication purposes, containing, in addition to other microorganisms, dead organic matter debris or microplastic elements.

In the case of *Haematococcus pluvialis* cultures, problems may also result from cell colour. The cultivation of this microalgae for astaxanthin production is a 2-step process. The first step is devoted to biomass accumulation under growth-friendly conditions (green step), and the second step is devoted to astaxanthin production under various stress conditions (red step). In each step, the cells can not only be different sizes, but also different colours [25]. This poses another challenge to the object recognition of YOLO v8.

## CONCLUSIONS

Deep machine learning and neural networks can greatly facilitate quantitative research and act as a valuable resource for analyzing microscopic images, for example, in bioindication or biomonitoring. The identification of *Hematococcus* and *Coelastrum* sp. presented in this work is just one of many examples where deep learning can be applied. In terms of recognition accuracy within a class and the extent to which YOLO v8 recognizes all members of the class, noticeably better results were obtained for the *Coelastrum* sp. object than for *Haematococcus sp.* This is confirmed by the accuracy, precision, and recovery parameters. This is due to the morphological variation of objects within a class (colour, size) and associated objects (background). The F1 score for the objects studied was >0.9. This is promising and indicates that the detection and classification of these objects can be supported. These results indicate that it is possible to determine and count not only single algal cells, but also colonies over a wide range of sizes, as in *Coelastrun sp.*

In microalgae research, Deep Learning Methods are employed for such tasks as image classification, object detection, and phenotype recognition. By leveraging large datasets of microalgae images, these models can automatically identify and classify microalgal species, morphological features, and cellular structures with high accuracy and efficiency. Future applications of neural networks from the YOLO family may include the detection of additional objects in microscopic samples and the segmentation of data elements in digital images of biological slices. This has the potential to play a major role in improving and automating the quantitative and qualitative control of biomass in terms of broad industrial application (aquaculture, food, health, energy and fuel industries), as well as bioindication within different fields of environment protection.

### Acknowledgements
The research was carried out as part of the task commissioned under the title "Politechnical Network VIA CARPATIA named after the President of the Republic of Poland Lech Kaczyński" financed from the Special Purpose Grant of the Minister of Science, contract number MEiN/2022/DPI/2578 action "PO SĄSIEDZKU - inter-university research internships and study visits.

## REFERENCES

1. Gamrat R, Puc M, Gałczyńska M, et al. Differences in the pollen content of varieties of Polish honey from urban and rural apiaries. Acta Univ Cibinien. Ser E: Food Technol. 2022;26(1):109–122. https://doi.org/10.2478/aucft-2022-0009
2. Choi H-J, Lee S-M. Effects of microalgae on the removal of nutrients from wastewater: Various concentrations of Chlorella vulgaris. Environ Eng Res. 2012;17(S1):S3-S8. https://doi.org/10.4491/eer.2012.17.S1.S3
3. Dębowski M, Zieliński M, Krzemieniewski M, et al. Możliwość namnażania biomasy glonów na bazie odcieku pochodzącego z odwadniania osadów pofermentacyjnych. Ann Set Environ Protect. 2013;15(2):1612–1623. ISSN 1506 218X1612-1622
4. Yang Y, Tang S, Chen JP. Carbon capture and utilization by algae with high concentration CO2 or bicarbonate as carbon source. Sci Total Environ. 2024;918:170325. https://doi.org/10.1016/j.scitotenv.2024.170325

5. Iglina T, Iglin P, Pashchenko D. Industrial CO2 capture by algae: A review and recent advances. Sustainability. 2022;14:3801. https://doi.org/10.3390/su14073801
6. Bhateria R, Dhaka R. Algae as biofuel. Biofuels. 2015;5(6):607–631. https://doi:10.1080/17597269.2014.1003701
7. Neeti K, Gaurav K, Singh R. The potential of algae biofuel as a renewable and sustainable bioresource. Eng Proc. 2023;37(1):22. https://doi.org/10.3390/ECP2023-14716
8. Szczepocka E, Szulc B, Szulc K, et al. Diatom indices in the biological assessment of the water quality based on the example of a small lowland river. Oceanol Hydrobiol Stud. 2014;43(3):265–273. https://doi.org/10.2478/s13545-014-0141-z
9. Commission Implementing Regulation (EU) 2020/1820 of 2 December 2020 authorising the placing on the market of dried Euglena gracilis as a novel food under Regulation (EU) 2015/2283 of the European Parliament and of the Council and amending Commission Implementing Regulation (EU) 2017/2470.
10. Le-Feuvre R, Moraga-Suazo P, Gonzalez J. Biotechnology applied to Haematococcus pluvialis Fotow: challenges and prospects for the enhancement of astaxanthin accumulation. J Appl Phycol. 2020;32:3831–3852 https://doi.org/10.1007/s10811-020-02231-z
11. Thiyagarasaiyar K, Goh BH, Jeon YJ, et al. Algae metabolites in cosmeceutical: An overview of current applications and challenges. Mar Drugs. 2020;18:323. https://doi.org/10.3390/md18060323
12. Pogorzelska E, Hamułka J, Wawrzyniak A. Astaksantyna – budowa, właściwości i możliwości zastosowania w żywności funkcjonalnej. Żywność. Nauka. Technologia. Jakość. 2016;1(104):5–16. http://dx.doi.org/10.15193/zntj/2016/104/097
13. Ota S, Kawano S. Three-dimensional ultrastructure and hyperspectral imaging of metabolite accumulation and dynamics in Haematococcus and Chlorella. Microscopy. 2019;68(1):57–68. https://doi.org/10.1093/jmicro/dfy142
14. Taye MM. Understanding of Machine Learning with Deep Learning: architectures, workflow, applications and future directions. Computers. 2023;12:91. https://doi.org/10.3390/computers12050091
15. Picińska-Fałtynowicz J, Błachuta J. Klucz do identyfikacji organizmów fitoplanktonowych z rzek i jezior dla celów badań monitoringowych części wód powierzchniowych w Polsce. Warszawa: Główny Inspektorat Ochrony Środowiska; 2012;83. ISBN 978-83-61227-05-2, OCLC 839100228
16. Adam C, Haryono A. Morphological study of Coelastrum cambricum from the Peat Water of Palangka Raya, Indonesia. Nusantara Sci Technol Proceed. 2022;25:21–28. https://doi.org/10.11594/nstp.2022.2504
17. Rauytanapanit M, Jancho K, Kusolkumbot P, et al. Nutrient deprivation-associated changes in green microalga Coelastrum sp. TISTR 9501RE enhanced potent antioxidant carotenoids. Mar Drugs. 2019;17:328. https://doi.org/10.3390/md17060328
18. Staniszewski M, Dziadosz M, Zaburko J, et al. Automatic System for Acquisition and Analysis of Microscopic Digital Images containing activated sludge. Sci Technol Res J. 2024;18(7):51–61. https://doi.org/10.12913/22998624/192503
19. Dziadosz M, Majerek D, Łagód G. Microscopic studies of activated sludge supported by Automatic Image Analysis based on Deep Learning Neural Networks, J Ecol Eng. 2024;25(4):360–369. https://doi.org/10.12911/22998993/185317
20. Jamka K, Wróblewska-Łuczka P, Adamczuk P, et al. Methodology for preparing a cosmetic sample for the development of Microorganism Detection System (SDM) software and artificial intelligence learning to recognize specific microbial species. Ann Agric Environ Med. 2021;28(4):681–685. https://doi.org/10.26444/aaem/144696
21. Wu J, Xu W, He J, et al. YOLO for Penguin Detection and counting based on remote sensing images. Remote Sens. 2023;15:2598. https://doi.org/10.3390/rs15102598
22. Codd GA, Morrison LF, Metcalf JS. Cyanobacterial toxins: Risk management for health protection. Toxicol Appl Pharmacol. 2005; 203(3):264–272. https://doi.org/10.1016/j.taap.2004.02.016
23. Walo M, Majerek D, Kuzmina T, et al. Bioindication of surface water supported by automatic image analysis using Deep Learning Neural Network – Cyclotella Case Study. J Ecol Eng. 2024;25(12):366–375. https://doi.org/10.12911/22998993/194179
24. Yan H, Peng X, Chen Ch, et al. YOLOx model-based object detection for microalgal bioprocess. Algal Res. 2023;74:103178. https://doi.org/10.1016/j.algal.2023.103178
25. Oslan SNH, Shoparwe NF, Yusoff AH, et al. A Review on Haematococcus pluvialis bioprocess optimization of green and red stage culture conditions for the production of natural astaxanthin. Biomolecules. 2021;11(2):256. https://doi.org/10.3390/biom11020256