

As depicted in Fig. 2b, during dataset construction, we initiate the process by preprocessing the original images using a 2×2 tiling approach, which aids in improving the model's accuracy in detecting smaller microalgae, such as *Cosmarium* and *Haematococcus Pluvialis*. Subsequently, the tiled image set undergoes a Mosaic augmentation to restore the original image size and introduce more background context, thereby enhancing the model's performance. An overall precision improvement of approximately 6% is observed, as demonstrated in the performance comparison graph in Fig. 2b. Upon completion of all data preprocessing and augmentation steps, the final dataset, as illustrated in the bar chart in Fig. 2a, is divided into three distinct sets: training (85%, comprising 4276 images), validation

(10%, with 484 images), and testing (5%, encompassing 243 images).

Based on the microalgae dataset, we train a microalgae species detection network (MSDN) for 100 epochs with batch size 16 using the YOLOv5 architecture.³⁴ The training process is monitored, and details can be found in Fig. S2.† The detecting results of the trained network on the test image set and its evaluation matrices for each microalgae species are given in Fig. 2c. Since each object in the analysed image may belong to a different class, only true positive (TP), background false negative (FN) and background false positive (FP) are meaningful in the confusion matrix. The precision ($TP/(TP + FP)$) and recall ($TP/(TP + FN)$) of the model for each

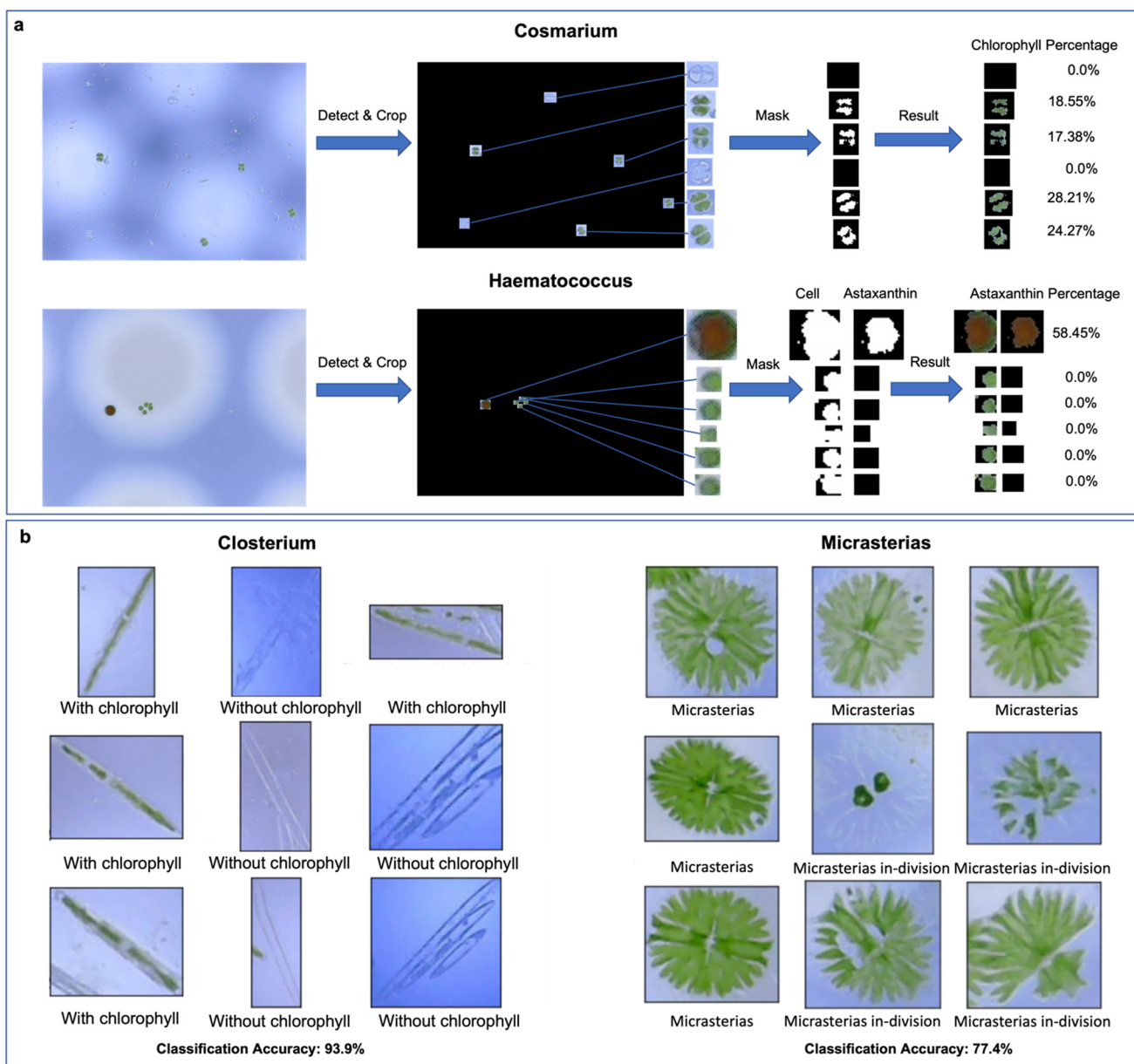


Fig. 3 Additional functions of the AIMP for monitoring the growth stage of microalgae. a) Further classification of *Cosmarium* and *Haematococcus Pluvialis* detected by the microalgae detection model by a colour extraction function. b) Further classification of *Closterium* and *Micrasterias* using two specific trained classifiers based on the EfficientNet-Lite0 architecture.



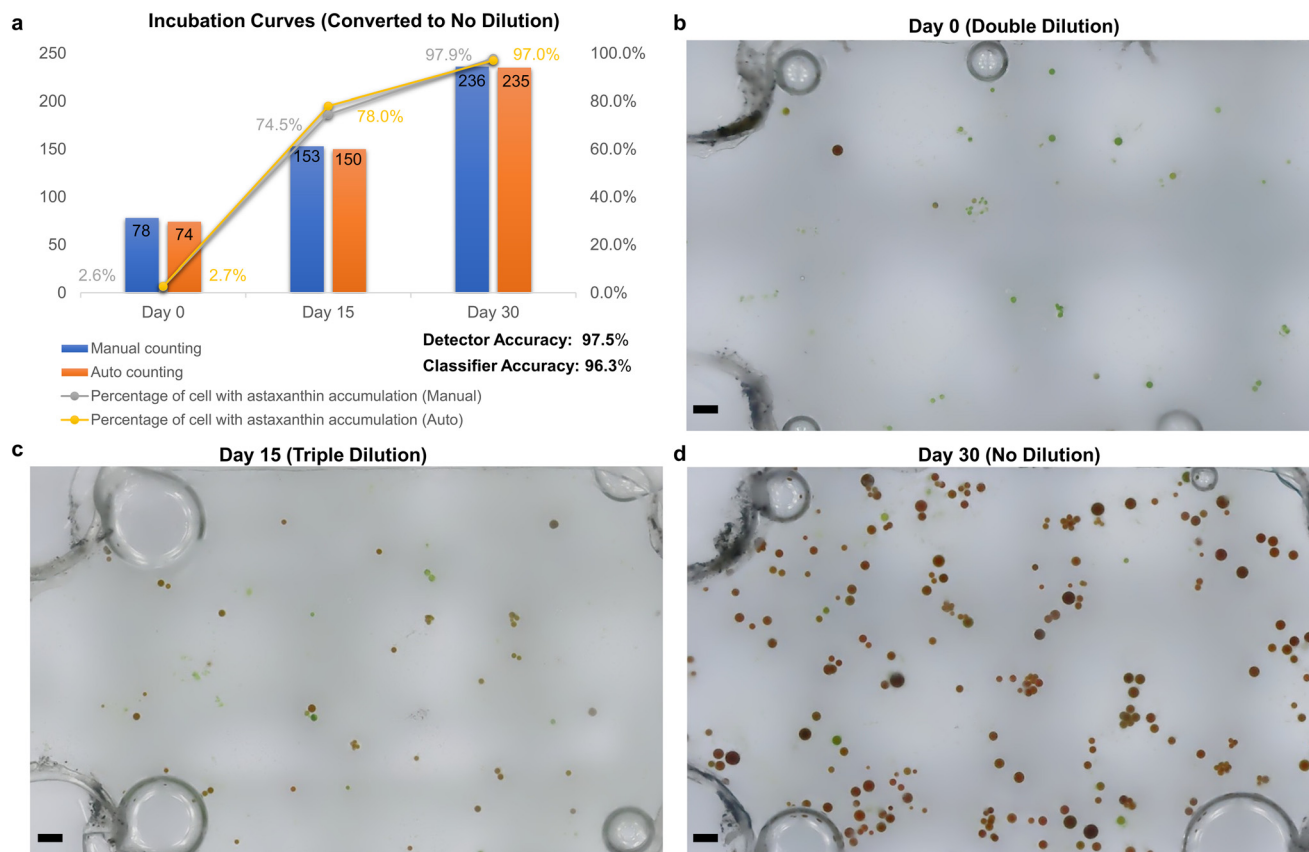


Fig. 5 Monitoring the accumulation of astaxanthin in *Haematococcus Pluvialis*. a) Culture experiment data and comparison graph of manual and AIMP obtained statistics. Example images of a single observation chamber at b) day 0 (samples are diluted twofold before adding to the chip), c) day 15 (samples are tripled diluted before adding to the chip), and d) day 30 (samples are not diluted). Scale bars are 200 μm .

accuracy and the subsequent dichotomous accuracy can be derived using eqn (3). A comparison of these two sets of data shows that the AIMP achieves 96.3% accuracy in further dichotomising based on astaxanthin accumulation detection.

$$\text{Precision}_C = \frac{\sum_{i=0,15,30}^3 \frac{|\text{Ratio_Manual_Count@Day}_i - \text{Ratio_Auto_Count@Day}_i|}{\text{Percentage_Manual_Count@Day}_i}}{3} \quad (3)$$

Fig. 5b–d show example images of a single observation chamber collected using the automatic stitching function of the AIMP at Day 0, Day 15 and Day 30, respectively. The *Haematococcus Pluvialis* microalgae are initially very small ($\sim 10 \mu\text{m}$) but grow bigger ($\sim 70 \mu\text{m}$) after 30 days of culture. In our microfluidic chip, the height of the observation chamber is $\sim 150 \mu\text{m}$. Therefore, larger microalgae are less likely to overlap within the chamber, as opposed to smaller ones. Hence, when working with relatively small microalgae, we take measures to dilute and agitate them, effectively reducing the occurrence of overlap. Samples used on Day 0 are diluted twofold and samples used on Day 15 are triple diluted to reduce the overlap of *Haematococcus Pluvialis* cells.

The samples used on Day 30 are not diluted because *Haematococcus Pluvialis* cells at this time have become larger and less prone to overlap. To mitigate the effect caused by the unavoidable overlapping of microalgae, many images of microalgae with slight overlap are intentionally included in

our training dataset, even for the larger ones. This deliberate inclusion helps ensure the model's robustness and ability to handle various scenarios.

Conclusion

In summary, we develop the AIMP to overcome the significant time and labour cost associated with conventional methods for microalgae detection and monitoring. Using a low-cost USB portable microscope and the mini XYZ motorised stage, the AIMP achieves the low-cost and portability requirements without sacrificing critical functionality. The use of laser-cut microfluidic chips instead of microfabricated microfluidic



- 31 G. Zheng, Y. Wang, Z. Wang, W. Zhong, H. Wang and Y. Li, *Mar. Pollut. Bull.*, 2013, **72**, 231–243.
- 32 J. Wang, G. Wang, M. Chen, Y. Wang, G. Ding, Y. Zhang, Y. Kang and X. Pan, *Algal Res.*, 2019, **42**, 101593.
- 33 G. Jaworski, J. Talling and S. Heaney, *Br. Phycol. J.*, 1981, **16**, 395–410.
- 34 G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael, J. Fang, Z. Yifu, C. Wong and D. Montes, *Zenodo*, 2022, DOI: [10.5281/zenodo.7347926](https://doi.org/10.5281/zenodo.7347926).
- 35 D. H. Fudholi, S. Rani, D. M. Arifin and M. R. Satyatama, *Sci. J. Inform.*, 2021, **8**, 111–118.
- 36 M. N. Ab Wahab, A. Nazir, A. T. Z. Ren, M. H. M. Noor, M. F. Akbar and A. S. A. Mohamed, *IEEE Access*, 2021, **9**, 134065–134080.
- 37 A. Hosikian, S. Lim, R. Halim and M. K. Danquah, *Int. J. Chem. Eng.*, 2010, **2010**, 391632–391643.
- 38 R. R. Ambati, P. Siew Moi, S. Ravi and R. G. Aswathanarayana, *Mar. Drugs*, 2014, **12**, 128–152.

