Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p27871-cd

Quantitative Characterization of Individual Algae Cells via Machine Learning based Clustering

Jonas Strohhofer
Munich University of Applied Sciences, Munich, Germany. E-mail: strohhofer.jonas@hm.edu

Nour Albaali
htw saar - University of Applied Sciences, Saarbrücken, Germany. E-mail: nour.albaali@htwsaar.de

Laurent Demaret
Munich University of Applied Sciences, Munich, Germany. E-mail: laurent.demaret@hm.edu

Timo Gehring
htw saar - University of Applied Sciences, Saarbrücken, Germany. E-mail: timo.gehring@htwsaar.de

Marcin Hinz
Munich University of Applied Sciences, Munich, Germany. E-mail: marcin.hinz@hm.edu

The industrial adoption of microalgae cultivation is currently limited by economic feasibility and the challenge of accurately predicting growth behavior. Conventional control philosophies often focus on isolated parameters rather than a holistic, predictive assessment of culture status. This study addresses this gap by extracting morphological information at the single-cell level to characterize population dynamics.
We used a pipeline for cell segmentation and feature extraction, capturing diverse properties including shape, color, texture, and intensity distribution. To evaluate the underlying structure of this data, three distinct clustering techniques were compared: density-based (DBSCAN), distance-based (K-Means), and distribution-based (GMM).
Our findings indicate that DBSCAN is not suited for this application, as the continuous nature of microalgal growth prevents the formation of discrete density-based clusters. In contrast, both K-Means and GMM successfully processed the data. By setting a cluster size of eight, we identified three highly consistent clusters across both algorithms, corresponding to distinct, identifiable phenotypes validated by domain knowledge. However, the remaining five clusters showed significant mismatches, highlighting the sensitivity of phenotype classification to the underlying mathematical assumptions. These results demonstrate that while machine learning offers a powerful tool for holistic culture monitoring, the choice of algorithm is critical for interpreting biological states.

Keywords: Microalgae, Image Analysis, Automated Phenotyping, Cell Segmentation, Growth Prediction.

1. Introduction

Microalgae offer a unique solution by converting CO_2 into high-value compounds such as feedstock for chemical industry, biochemicals, fuels, and polymers. Consequently, microalgae cultivation plays a key role in contributing to the UN Sustainable Development Goals (Gehring et al. (2026)). However, large-scale adoption is hindered by technological and economic barriers, requiring rapid optimization and scale-up strategies.

Contrary to traditional sunlight cultivation, artificial lighting and photobioreactors now enable precise parameter control (Gehring et al. (2025)). Yet, accurate growth prediction remains a bottleneck. Current assessments rely on a complex combination of sensors (e. g. pH, T, O_2, CO_2). This fragmented approach is laborious and thus expensive. To facilitate industrial scale-up, predictive control with minimal sensors and AI-assisted knowledge generation is desirable.

The use of machine learning (ML) for implementing cheap sensors is lagging behind and has not been adopted so far Freitas et al. (2025). In addition, existing measuring and control philosophies focus on individual parameters, avoiding

a holistic and predictive approach assessing the whole status of a microalgal culture.

Accordingly, automated and reliable growth rate predictions for the next 24 hours would represent a significant step forward, providing crucial data on biomass productivity potential. The modeling of such a complex system remains one of the main challenges for the design of predictive tools, which is lacking so far.

In this context, modern ML offers a powerful framework. By unraveling hidden dependencies and reducing *ad hoc assumptions*, it yields a system description simple enough to be measured yet rich enough for accurate modeling. This contribution bridges state-of-the-art imaging with unsupervised learning, aiming to recognize cell similarities based solely on data correlation rather than predefined phenotypes.

Such a data-driven approach paves the way for future investigation, making it possible to identify the most influential factors in growth and setting the basis for a robust analysis of the phenotypic transitions of microalgae systems.

A summarised visualisation of our methodology can be seen in figure 1.

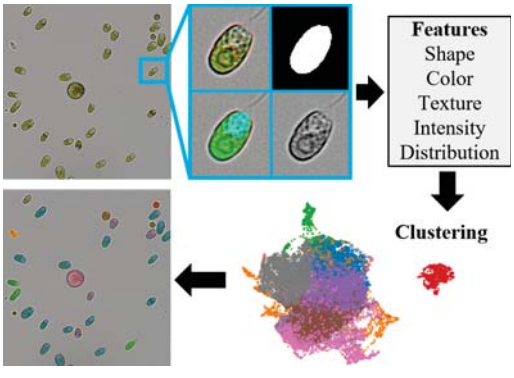


Fig. 1.: The overall workflow of our research: After segmenting the individuall cells features are extracted from the RGB and BW image, the mask and the area of the white/green parts of the cell. Those features are then clustered using K-Means, DBSCAN and GMM.

2. State of the Art

We review traditional algae phenotyping and established cell segmentation methods. We then discuss relevant quantitative features for microalgae and evaluate algorithms for clustering these features to identify distinct phenotypes.

2.1. *Single-cell Algae Phenotyping*

This section briefly reviews microalgae phenotyping and growth prediction methods, situating our approach within the context of current research. Traditional phenotyping relies on bulk metrics such as optical density, dry weight, or average cell counts, complemented by manual microscopic inspection performed by domain experts. Due to the dimension of the microalgae cells (typically $\leq$ 20 µm), an imaging via high resolution microscopy is required. With typical microalgae cell concentrations in the range of $10^6 - 10^8$ cells/mL a manual assessment falls short – even for specialists in the field – as for a representative evaluation thousands of cells have to assessed.

A further difficulty lies in the discrepancy between real system and microscopy images: While microscopy images represent a good instant picture of reality and what happened shortly before, their observation remain insufficient to predict future behavior and to directly derive a comprehensive modeling of growth phenomena.

CellProfiler Stirling et al. (2021) is as open-source tool for analyzing microscopy images. Although its companion tool, CellProfiler Analyst, can identify similar cells, it operates on a supervised basis: users must first manually divide cells into labeled classes for training.

As already pointed by Freitas et al. (2025), an important property of future algae analysis is to differentiate morphological properties between cells *without* being limited by predefined labels.

2.2. *Algae Segmentation*

To enable the clustering of algae into distinct phenotypes, quantitative information must first be extracted from each organism. The prerequisite for this analysis is to isolate individual cells within the microscopy images. While classical image

processing techniques, such as thresholding, often fail to generalize to new datasets, recent advancements in deep learning have demonstrated superior performance in separating overlapping instances. For example, AlgaeSeg-YOLO adapts the YOLO object detector for algae (Yang et al. (2024)). A major breakthrough was Cellpose by Stringer et al. (2021) which uses a U-Net architecture. Instead of predicting binary segmentation masks, it learned to predict vector flow fields, which helps the model to better generalize to unseen data. Recently, this method was combined with Meta's Segment Anything Model (SAM) (Kirillov et al. (2023)) to create Cellpose-SAM, leveraging the strengths of both architectures (Pachitariu et al. (2025)).

2.3. *Feature Extraction*

The clustering of algae relies on extracting discriminative quantitative information from microscopy images. Methodologically, feature representations fall into two primary categories: handcrafted features explicitly defined by domain experts based on biological traits and latent features automatically learned by neural networks.

For example, Soar et al. (2023) used a Convolutional Neural Network (CNN) to demonstrate that transfer learning can be effectively applied to the classification of microalgae. While such features can be adapted for clustering tasks, supervised deep learning approaches require large, labeled datasets. To address the lack of labels, Lu et al. (2019) proposed a self-supervised approach to extract features from fluorescent images. However, a significant drawback of deep learning representations remains their abstract nature, which therefore lacks direct biological interpretability.

In contrast, handcrafted features offer a biologically explainable alternative. Otálora et al. (2021) also used handcrafted features to distinguish different algae species. Established pipelines, such as CellProfiler typically extract a set of metrics which can be categorized into morphology, intensity, texture, and spatial distribution:

The spectrum of extractable quantitative features is vast. However, to keep the analysis interpretable, we limited our selection to those features which are already established in CellProfiler.

Since shape and optical characteristics indicate the health and growth conditions of *Dunaliella salina* Mai et al. (2017), we extended the standard feature set with custom color-based metrics. Grounded in domain expertise, this also included a targeted segmentation of the cell into chloroplast-associated (green) regions and optically bright, weakly pigmented (white) sub-regions, allowing for the extraction of biologically distinct intensity statistics beyond global cell averages.

For the visualization of the high-dimensional featureportfolio, Uniform Manifold Approximation and Projection (UMAP) is useful due to its ability to preserve both local and global structures (McInnes et al. (2018)). UMAP has also been validated as an effective visualization technique for cell-based data, offering a balance between computational speed and biological interpretability (Becht et al. (2018)).

2.4. *Clustering*

The overall aim of this paper is to identify specific sub-groups of algae in the microscopy images to be able to better describe the growth state of the population. In this study, we compare three established methods: Density-Based Spatial Clustering of Applications with Noise (DBSCAN), K-Means, and Gaussian Mixture Models (GMM). These algorithms all rely on different assumptions: GMM assumes that the data originates from a mixture of Gaussian distributions and that data points are independent of one another. K-Means expects a spherical cluster structure, a limitation often criticized when applied to complex real-world datasets. In contrast, DBSCAN offers the advantage of identifying the amount of clusters by itself. (Ros and Riad (2024))

While clustering microalgae is not a fundamentally new concept, its application in uncovering hidden structures within populations remains largely unexplored. Schröder et al. (2020) developed *MorphoCluster*, which employs clustering techniques primarily to annotate large-scale image datasets. Similarly, Coltelli et al. (2014) used Self-Organizing Maps to categorize algae; however,

their focus was on species differentiation rather than characterizing the species growth dynamics of a single population. Nelli et al. (2024) used clustering-techniques for separating cells from filaments just based on their morphology.

3. Data Acquisition and Dataset

Our study is based on *Dunaliella salina*, a marine halophilic microalgae of interest for producing value-added compounds such as beta-carotenoids and astaxanthin (Ahmed et al. (2017)). Cultivation is done e.g. in open lagoons in Australia (Schlipalius (1991); Masojídek et al. (2023)). *Dunaliella salina* (strain 184.80) was obtained from Culture Collection of Algae at Göttingen University SAG and cultivated in modified johnson medium (compare Borowitzka (1992)).

10 mL samples were harvested at optical density $OD_{680} = 0.2$ to 0.6, concentrated to $0.2 - 0.4$ mL by centrifugation. Addition of 5 µL 0.4% m/m formaldehyde solution was added to the algae concentrate directly prior microscopic imaging to prevent fast moving of cell flagella. 5 µl algae suspension were added on 10×10 mm slide and cover glass. Microscopic imaging was performed with a Keyence BZ-X1000 with a plan apochromat 40x objective.

The dataset consists of 261 RGB images (1832×1374) containing 16.141 filtered cells, exhibiting heterogeneity in shape and optical properties (compare leftmost image in Fig. 1).

4. Methodology

To extract relevant growth data, our workflow first segments individual cells from the background. We then extract features from each cell and group them. Finally, a domain expert evaluates and compares the clustering results.

4.1. *Segmentation and Postprocessing*

As already mentioned Cellpose-SAM is one of the most efficient existing methods for segmenting cells. It was applied to our dataset without the need for parameter tuning or additional training. To ensure data quality, the resulting segmentation masks underwent several post-processing steps. First, we removed cells touching the image borders to avoid analyzing incomplete shapes. Second, an exclusion zone was applied to the bottom-right corner to ignore areas covered by microscope annotations. Additionally, the pipeline filtered out segmented flagella as well as objects that did not meet specific minimum and maximum size requirements, effectively removing small debris and large artifacts.

4.2. *Features*

We base our feature-portfolio on the available features in CellProfiler. To be able to also include information on the color properties we added additional color features. This leads initially to a total of 110 handcrafted features for each individual cell. To improve clustering performance and reduce dimensionality, features with low variance were discarded. To mitigate multicollinearity, we conducted a pairwise correlation analysis. For any pair of features exhibiting a correlation coefficient greater than 0.9, one of the two features was removed to eliminate redundancy.

This lead to a total of 32 features to characterize the cell phenotypes. They can be categorized as follows: **Shape features** include Area, FormFactor, Extent, FeretDiameter$_{min}$, FeretDiameter$_{max}$, and Eccentricity; **Texture features** consist of Granularity$_1$, Granularity$_2$, InfoMeas$_1$, Entropy, Correlation, and Contrast. The **Distribution features** comprise RadialCV, FracAtD$_{3of4}$, Zernike$_{1Mag}$, and Zernike$_{2Mag}$, while **Intensity features** cover MassDisplacement, Intensity$_{Std}$, IntensityEdge$_{Std}$, MADIntensity, and IntensityEdge$_{mean}$. **Color features** are divided into global metrics (R$_{Std}$, G$_{Mean}$, G$_{Std}$, B$_{Mean}$, B$_{Std}$, B$_{Max}$) and segmented local metrics (Light$_{AreaRatio}$, Light$_{GMean}$, Light$_{GStd}$, Green$_{BMean}$, Green$_{BStd}$).

Overall the features describe each algae cell in terms of its shape, texture, intensity, color, and the distribution of intensities across its diameter.

Prior to clustering, all features are preprocessed using a *StandardScaler*. This step is needed for distance-based and density-based algorithms.

To visualize the highdimensional feature space, we used a UMAP projection ($n_{neighbors} = 5$,

$min_{dist} = 0.3$). The resulting manifold highlights a property of our dataset: instead of forming isolated subgroups, the data exhibits a continuous transition between various cell states. This observation aligns with the biological reality of algal development, which is characterized by gradual growth processes rather than abrupt stage transitions. However, the UMAP visualization also reveals a distinct sub-population, forming a small, separate island of data points. This visualization can be seen in figure 2a and figure 2b. For compactly representing the cell properties, we employ spider diagrams, enabling direct qualitative comparison of phenotypic patterns across clusters.

4.3. *Clustering*

To advance the monitoring of algal cultures beyond simple cell counts, a detailed assessment of phenotypic states is required. While domain experts can intuitively assess cellular states based on years of experience, these qualitative observations lack the scalability and reproducibility required for large-scale bioprocesses. Consequently, this implicit expertise remains fragmented, preventing its integration into a systematic knowledge base to advance the general understanding of algal culture behavior. The motivation behind this unsupervised clustering is to bridge this gap by exploring the underlying data structure to reveal phenotypic patterns. By identifying distinct clusters, we take an important step toward objectifying expert knowledge and translating it into reproducible data. We address this by applying three different unsupervised clustering algorithms to the extracted feature set. A domain expert subsequently evaluates the identified sub-groups to determine which method best captures the biological reality needed for growth state analysis.

Each algorithm was configured individually by optimizing its specific parameters. For K-Means, we used the Elbow method and chose a conservative value of eight clusters. This prevents overfragmenting the data and ensures that the resulting phenotypes remain biologically interpretable for experts. For GMM, we followed the Bayesian Information Criterion (BIC); since the results were not definitive, we also used eight clusters to allow for a direct comparison with K-Means. For DBSCAN, we conducted a grid search for ϵ and $min_{Samples}$ to find a balance between reducing noise and maintaining clear cluster separation.

5. Results

This chapter evaluates DBSCAN, K-Means, and GMM when applied to the available algae data. For each method, a preliminary hyperparameter optimization is done. The core analysis focuses on the phenotypic discussion of exemplary clusters. By correlating feature distributions with visual inspection, we assess each algorithm's ability to isolate biologically distinct states essential for monitoring algal growth.

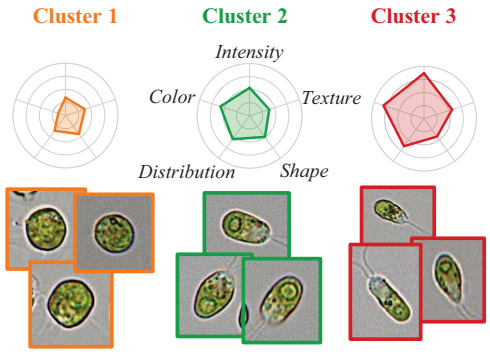
5.1. *DBSCAN*

DBSCAN determines the number of clusters implicitly based on the density parameters ϵ and $min_{Samples}$. Depending on the parameter configuration, the results converged into two extreme scenarios: either the formation of a single monolithic cluster containing the entire population, or severe over-fragmentation resulting in hundreds of micro-clusters which contain four or fewer cells.

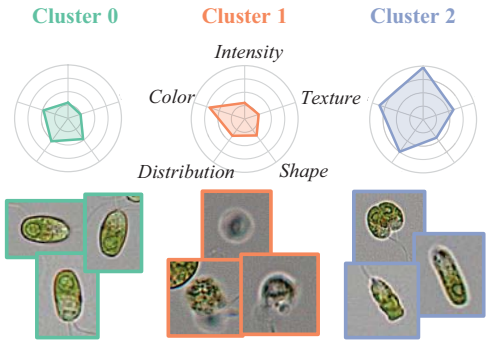
These results suggest that DBSCAN is not suited for the topology of our feature space. We hypothesize that this stems from the biological nature of algal growth, which represents a continuous spectrum rather than discrete, density-separated groups. The UMAP projection (compare fig. 3 also underscores that hypothesis. Consequently the algorithm bridges the boundaries between different algae cell states, effectively "walking along" the gradient of the data. Since the method proved to be either too restrictive (isolating near-duplicates) or too inclusive (merging distinct phenotypes), it failed to provide a biologically meaningful clustering required for the evaluation of growth phases.

5.2. *K-Means*

The optimal number of clusters k was evaluated using the Elbow method, which suggested a range between $k = 8$ and $k = 14$. To ensure biological interpretability and avoid over-fragmentation of



(a) **Cluster Examples K-Means:** Representative cell images for Clusters 1, 2, and 3 and their radial profiles illustrating the qualitative feature distributions.



(b) **Cluster Examples GMM:** Representative cell images for Clusters 0, 1, and 2 and their radial profiles illustrating the qualitative feature distributions.

Fig. 2.: Comparison of example clustering results between K-Means and GMM.

phenotypes, we selected $k = 8$ for our analysis. Example clusters can be seen in figure 2a.

The resulting clusters exhibited a heterogeneous size distribution. Three dominant clusters each contained more than 21.6% of the total population, collectively representing approximately two-thirds of all analyzed cells. The remaining data was distributed across two medium-sized clusters ($\approx$ 11%) and three minor clusters, each accounting for less than 4.2% of the dataset.

As the largest group *Cluster 2* represents the baseline phenotype. A feature analysis revealed intermediate values across all five categories (Shape, Texture, Intensity, Color, and Distribution), indicating a lack of extreme morphological traits. Domain experts described those cells in this cluster predominantly with an an oval morphology, frequently featuring a visible vacuole.

Accounting for 4.2% of the population, *Cluster 1* is defined by its distinct intensity profile. The features indicate that those cells have no bright regions and have a darker appearance. Visually, these cells are characterized by a round shape and the absence of a distinct vacuole.

Cluster 3 represents the smallest subpopulation (1.2%), primarily driven by specific color distribution patterns. We hypothesize that these cells represent a transient biological state, such as freshly divided cells or those initiating cell division as they appear to have elongated shapes.

5.3. *GMM*

The cluster selection was guided by BIC, which reached a plateau of stability starting at $k = 8$. Based on this, we selected a model with eight components for interpretation. This enables further comparison with the results of K-Means. Example clusters can be seen in figure 2b. The resulting clusters show a size distribution similar to K-Means: two dominant clusters encompass 22.4% and 23.9% of the data, while the remaining cells are distributed across four medium-sized groups (7.8%–16%) and two small clusters ($< 4.3\%$).

Cluster 0 is the largest one. It exhibits intermediate feature values across all categories. Visually, this group corresponds to the standard oval phenotype with a distinct vacuole and a heterogeneous (mixed) white-to-green intensity ratio, representing the typical healthy cell population.

Cluster 1 is a medium-sized sub-group (7.8%) which is defined by a highly specific feature profile, dominated by color metrics but exhibiting very low values for shape and texture. Visual inspection reveals that this group consists of smashed cells, and out-of-focus artifacts. Domain experts highlighted the identification of this cluster as an advantage of the GMM approach, as it allows for the automated exclusion of invalid data points prior to downstream growth analysis.

The smallest subgroup is *Cluster 2* (2.5%) which is characterized by high intensity values.

The cells in this cluster display irregular, ”odd” morphologies, ranging from elongated shapes to cells potentially undergoing division. While biologically difficult to classify, the algorithm successfully isolated these from the main population.

5.4. *Comparison*

Since DBSCAN did not yield biologically relevant results, the following comparison focuses exclusively on the clusters of K-Means and GMM. Figure 3 provides a qualitative overview, illustrating that while both algorithms identify some clusters similarly, their partitioning of the phenotyps differs in most sub-regions.

A quantitative analysis of the cluster overlaps reveals that three distinct phenotypes are identified with high consistency. The highest degree of congruency was observed for the subpopulation characterized by a complete lack of non-pigmented areas (K-Means *Cluster 1* vs. *GMM Cluster 3*). From the perspective of K-Means, these clusters share 99.3% of their data points, confirming that this feature is highly robust across different algorithmic assumptions.

Similarly, the clusters representing irregular shapes (K-Means *Cluster 3* vs. GMM *Cluster 2*) exhibit an overlap of 99%. This near-perfect agreement underscores that both algorithms isolate morphological outliers from the main population. In contrast, the larger, rounder cells (K-Means *Cluster 6* vs. GMM *Cluster 6*) show a more moderate overlap of 59.4%. This suggests that GMM places a higher weight on shape-based features, as 17% of these cells were assigned to GMM *Cluster 5* — a group also characterized as round but significantly smaller.

A major methodological advantage of GMM was observed regarding the ”artifact” cluster (*GMM Cluster 1*). While GMM successfully isolated debris and damaged cells into a single component, K-Means distributed these data points across almost all its clusters (with the exception of *Cluster 1*). From an analytical perspective, this provides a control metric, allowing for a quantified assessment of undeveloped or damaged cells within the population dynamics.

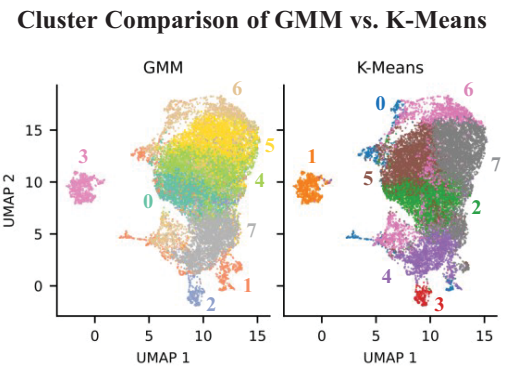


Fig. 3.: Qualitative comparison of the clusters.

The domain expert evaluated the GMM results as being better suited for a description of the algal population. While some phenotypes remain stable across both methods, GMM’s handling of artifacts and its sensitivity to shape features provide a more biologically picture of the culture. As transitions between physiological and phenotypic microalgae cell changes occur gradually rather than as a discrete process, we feel GMM particularly relevant for biotechnological monitoring and early detection of changes as prerequisite for model creation to accurately predict growth behavior.

6. Conclusion and Outlook

This study demonstrated that unsupervised clustering can effectively extract phenotypic states from single-cell data. While density-based methods (DBSCAN) failed due to the continuous nature of algal growth, both K-Means and GMM identified distinct subpopulations. Both algorithms found three similar clusters, where the distinct phenotype could be identified clearly, while mismatching on the other five found clusters.

Especially the detailed analysis of the different identified clusters and the borders between each clusters still need further investigation. Grounded in domain knowledge an expert described the GMM Clustering to be more useful for describing the state of the algae cultivation though further research needs to be put in into interpreting the found clusters and correlating them to the growth state and its forecast. Comparing clustering results over time to see how the clusters are changing

in size could help going into that direction. Further research could also include the exploration of other features which describe each cell. Also different clustering algorithm, e.g. hierarchical ones, could be used on the same data. Another idea is to do the clustering in an iterative manner, where phenotypes which could be clearly identified are excluded to get a better understanding of the intermediate cell states.

References

Ahmed, R. A., M. He, R. A. Aftab, et al. (2017). Bioenergy application of Dunaliella salina SA 134 grown at various salinity levels for lipid production. *Sci. Rep. 7*(1), 8118.

Becht, E., L. McInnes, et al. (2018). Dimensionality reduction for visualizing single-cell data using umap. *Nat. biotechnol.*.

Borowitzka, M. A. (Ed.) (1992). *Micro-algal biotechnology* (Repr ed.). Cambridge: Cambridge Univ. Press.

Coltelli, P., L. Barsanti, V. Evangelista, et al. (2014). Water monitoring: automated and real time identification and classification of algae using digital microscopy. *Environmental science. Processes & impacts 16*(11), 2656–2665.

Freitas, G. R., S. Badenes, R. Oliveira, and F. G. Martins (2025). A review of intelligent modeling for microalgae systems: Integrating data mining, machine learning, and hybrid approaches. *Processes 13*(9), 2956.

Gehring, T., Y. S. L. Choo, et al. (2026). Algae-driven transition from oil to renewable alternatives: Policy for scale and sustainability. *Sustainability & Circularity NOW 03*(CP).

Gehring, T., P. Maurer, M. Yildirim, et al. (2025). Floating light ball reactor: A scalable and flexible photobioreactor design for microalgae cultivation. *Sustainability & Circularity NOW 02*(CP).

Kirillov, A., E. Mintun, N. Ravi, et al. (2023). Segment anything.

Lu, A. X., O. Z. Kraus, S. Cooper, and A. M. Moses (2019). Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. *PLoS Comput. Biol. 15*(9), e1007348.

Mai, T., P. Nguyen, T. Vo, et al. (2017). Accumulation of lipid in dunaliella salina under nutrient starvation condition. *AJFN 5*(2), 58–61.

Masojídek, J., R. Lhotský, K. Štěrbová, et al. (2023). Solar bioreactors used for the industrial production of microalgae. *Appl. microbiol. and biotechnol. 107*(21), 6439–6458.

McInnes, L., J. Healy, and J. Melville (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint*.

Nelli, F., Z. Ge, L. Blackall, et al. (2024). Machine learning driven image segmentation and shape clustering of algal microscopic images obtained from various water types. *Computational and Structural Biotechnology Reports 1*, 100014.

Otálora, P., J. L. Guzmán, F. G. Acién, et al. (2021). Microalgae classification based on machine learning techniques.

Pachitariu, M., M. Rariden, and C. Stringer (2025). Cellpose-sam: superhuman generalization for cellular segmentation. *bioRxiv*.

Ros, F. and R. Riad (2024). *Feature and Dimensionality Reduction for Clustering with Deep Learning*. Cham: Springer.

Schlipalius, L. (1991). The extensive commercial cultivation of dunaliella salina. *Bioresour. Technol. 38*(2-3), 241–243.

Schröder, S.-M., R. Kiko, and R. Koch (2020). Morphocluster: Efficient annotation of plankton images by clustering. *Sensors (Basel, Switzerland) 20*(11).

Soar, J., O. S. Lih, L. H. Wen, et al. (2023). Deep image analysis for microalgae identification. In *Information Integration and Web Intelligence*, Volume 14416 of *Lecture Notes in Computer Science*, pp. 280–292. Cham: Springer.

Stirling, D. R., M. J. Swain-Bowden, et al. (2021). Cellprofiler 4: improvements in speed, utility and usability. *BMC bioinformatics 22*(1), 433.

Stringer, C., T. Wang, M. Michaelos, and M. Pachitariu (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods 18*(1), 100–106.

Yang, H., K. Lang, and X. Wang (2024). Identify and segment microalgae in complex backgrounds with improved yolo.