

Framework for the detection and monitoring of bloom-forming cyanobacteria in Pampean shallow lakes through microscopic image processing

David Gonzalo Romero ¹  Sol Ailen Montenegro ^{2,3}  Silvina Mariana Rosa ² 
Lilen Yema ³  Cristian Martinez ¹ 

¹ Departamento de Informática, Universidad Nacional de Salta (Fac. Cs. Exactas-UNSa), Salta, Argentina. ² Laboratorio de Biología de Protistas, Departamento de Biodiversidad y Biología Experimental (DBBE, FCEN-UBA), Buenos Aires, Argentina. ³ Laboratorio de Limnología, Instituto de Ecología, Genética y Evolución (IEGEBA, UBA-CONICET), Buenos Aires, Argentina.

Abstract. Cyanobacterial blooms are large-scale proliferations with profound ecological, public health, and economic impacts. For monitoring purposes, the World Health Organization defines alert levels based on biomass thresholds, with biovolume as the most widely recommended reference parameter. However, manual estimation of this metric is labor-intensive, time-consuming, and highly dependent on increasingly scarce taxonomic expertise. This study aims to develop a framework for automating the calculation of metrics for monitoring potentially toxic, bloom-forming filamentous cyanobacteria from microscopic images. Our approach integrates computer vision and digital image processing techniques for automated quantification. A custom dataset was curated, consisting of 26 high-resolution images (2048 x 1536 pixels) primarily featuring *Dolichospermum circinale*, collected from El Paraíso lagoon (Buenos Aires, Argentina). Manual laboratory counts were conducted as a baseline to estimate cell density and morphological dimensions. Subsequently, a Python-based pipeline was implemented to perform automated detection, segmentation, and quantification for each image. Comparative analysis between approaches demonstrated high consistency, with no statistically significant differences in results. Future work, which is currently under development, focuses on incorporating additional species and refining metrics, including the implementation of deep learning techniques (YOLO v11), with the ultimate goal of developing a robust, scalable, and low-cost web tool to enhance public health surveillance in local water bodies.

Keywords: biovolume, cyanobacteria, framework, machine learning

Framework para la detección y monitoreo de cianobacterias formadoras de floraciones en lagunas pampeanas mediante procesamiento de imágenes microscópicas

Resumen. Las floraciones de cianobacterias son crecimientos masivos de estos microorganismos con un profundo impacto ecológico, sanitario y económico. Para su monitoreo, la Organización Mundial de la Salud establece niveles de alerta basados en umbrales de biomasa, siendo el biovolumen el parámetro de referencia más recomendado. Sin embargo, el cálculo manual de esta métrica es laborioso, lento y dependiente de la disponibilidad de expertos, actualmente escasos. El objetivo de este trabajo es desarrollar un framework capaz de automatizar el cálculo de métricas para el monitoreo de cianobacterias filamentosas formadoras de floraciones potencialmente tóxicas a partir de imágenes microscópicas. Nuestra propuesta aborda la cuantificación automatizada de cianobacterias mediante la integración de técnicas de visión por computadora y procesamiento digital de imágenes. Se construyó un dataset propio compuesto por 26 imágenes de alta resolución (2048×1536 píxeles), con predominio de *Dolichospermum circinale*, obtenidas a partir de muestras recolectadas de la laguna El Paraíso (Buenos Aires). Como referencia, se realizó el recuento manual en laboratorio, registrando parámetros para estimar densidad celular y dimensiones morfológicas. Posteriormente, se implementó un pipeline en Python que realiza tareas de detección, segmentación y recuento/cuantificación automático sobre cada imagen. La comparación entre ambos enfoques evidenció concordancia en los resultados, sin diferencias numéricas significativas. En etapas subsiguientes, actualmente en desarrollo, se incorporarán nuevas especies de cianobacterias y se trabajará en el ajuste de las métricas, incluyendo técnicas de aprendizaje profundo (YOLO v11), con el objetivo de desarrollar una herramienta web robusta, escalable y de bajo costo que contribuya al monitoreo de la calidad sanitaria de cuerpos de agua locales.

Palabras clave: aprendizaje automático, biovolumen, cianobacteria, framework

1 Introducción

Las cianobacterias (phylum Cyanobacteria) son un grupo de microorganismos fotosintéticos oxigénicos que se encuentran principalmente en cuerpos de agua continentales (lagos, lagunas, embalses, ríos). Con un total de 6198 especies agrupadas en más de 200 géneros (Guiry & Guiry, 2026), estos microorganismos pueden generar sobrecrecimientos masivos denominados floraciones cuando las condiciones ambientales son propicias. Estas floraciones, que suelen estar dominadas por una o pocas especies, representan un importante problema ambiental, sanitario y económico debido a la potencialidad de algunas especies de cianobacterias de producir toxinas. En las últimas décadas, las floraciones han aumentado en intensidad y frecuencia debido a la eutrofización antrópica y al cambio climático, y la tendencia indica que seguirán aumentando globalmente (Chorus & Welker, 2021; Aguilera et al. 2023). Para monitorearlas, la Organización Mundial de la Salud (OMS) ha desarrollado una serie de valores guías y de alerta a partir de valores umbrales de abundancia, teniendo en cuenta la relación toxina/biomasa. Originalmente utilizaba valores de abundancia expresados como células/mL (además de valores de clorofila y toxinas) (Chorus & Bartram, 1999), y en la actualización más reciente (Chorus & Welker, 2021) se reemplazaron por valores de biovolumen (mm^3/L), ya que la biomasa es un parámetro más preciso para evaluar el riesgo. En el mismo sentido Bonilla & O'Farrell (2023) sugieren que los monitoreos de cianobacterias deben estar basados en el uso del biovolumen, ya que constituye la mejor variable indicadora de biomasa de estos organismos. Sin embargo, el cálculo manual de esta métrica es laborioso, consume tiempo y depende de la disponibilidad de expertos, actualmente escasos.

En la última década, el uso del Machine Learning (ML) aplicado a la identificación taxonómica de algas y cianobacterias dio lugar a publicaciones con importantes resultados que destacan el potencial de este enfoque para el monitoreo automatizado. Mosleh et al. (2012) en Malasia y Schulze et al. (2013) en Alemania, han explorado para ello el uso de redes neuronales artificiales para clasificar algas a nivel de género utilizando técnicas de segmentación por detección de bordes y extracción de características geométricas y texturales. Es de destacar que la gran mayoría de estos trabajos se focalizan en el reconocimiento (clasificación, identificación y detección) y no así en la cuantificación dada la complejidad técnica que se requiere para calcular el biovolumen (Madkour et al. 2023). Gandola et al. (2016) desarrollaron un sistema semi-automatizado con alta dependencia de expertos para la cuantificación de géneros filamentosos recolectados en lagos de Italia. Baek et al. (2020) han presentado una solución basada en Deep Learning (DL) para la detección, identificación y recuento de cianobacterias de muestras en ríos en Corea del Sur, observándose dificultades a nivel de recuento en la densidad, superposición celular y calidad de la imagen, entre otras. Recientemente, Ventura et al. (2025) proponen un método que combina técnicas clásicas de procesamiento de imágenes con DL para segmentar células de microalgas con alta precisión que, si bien demostró buenos resultados en una única especie, su capacidad de generalización en contextos multiespecie con morfología diferentes es aún incierta. Es importante notar que los modelos propuestos en la literatura se han desarrollado sobre organismos de otras regiones geográficas, que no necesariamente reflejan la biodiversidad de Argentina. Ante este vacío, nuestro trabajo presenta el primer framework diseñado para automatizar el cálculo de métricas en el monitoreo de

cianobacterias filamentosas potencialmente tóxicas a partir de imágenes microscópicas provenientes de lagunas pampeanas. Para iniciar este estudio, se seleccionó una muestra de una laguna con dominancia de la especie *Dolichospermum circinale*, debido a que es una de las más frecuentes en las floraciones tóxicas en Argentina (Aguilera et al. 2018; O'Farrell et al. 2019) ya que además posee una morfología que facilita el reconocimiento de las células. Este desarrollo, enmarcado en un proyecto interdisciplinario de impacto socio-tecnológico, busca optimizar la vigilancia sanitaria de los cuerpos de agua mediante una herramienta adaptada a las condiciones locales, superando las limitaciones de los métodos manuales y la dependencia de analistas expertos.

2 Materiales y Métodos

Para la construcción del dataset, se tomaron muestras de agua de la laguna pampeana El Paraíso (39.065559, -62.747130, Villarino, provincia de Buenos Aires). Se realizó un análisis cuantitativo de cianobacterias mediante el método de Utermöhl (1958) con un microscopio óptico invertido (Olympus CKX31) a un aumento de 400X, utilizando una cámara con un volumen (V) de 2 cm³ y un área (A) de 530,93 mm². Se recorrió la cámara hasta encontrar al menos 25 individuos, registrando la presencia de cianobacterias formadoras de floraciones, y tomando una imagen de alta resolución (2048×1536 píxeles) en cada uno de los casos. Se midieron las dimensiones (diámetro) de las células y se contó la cantidad de células por filamento en cada campo. Para el cálculo de la densidad celular de la muestra (N), se utilizó la siguiente fórmula:

$$N=CA/aSV \quad (1)$$

donde C es el número de células contadas, S el número de campos contados y a corresponde al área de cada campo; el volumen de las células se estimó mediante el enfoque geométrico de Hillebrand (1999), considerando a las células de *Dolichospermum circinale* como una esfera. Finalmente, el biovolumen de la muestra se estimó multiplicando el volumen celular promedio por la densidad celular.

Posteriormente, se implementó un pipeline en Python que realiza tareas de detección, segmentación y recuento automático sobre cada imagen. El mismo se ejecutó sobre una computadora portátil Intel core i5 1.7 Ghz con 16 GB de RAM y disco sólido de 256 GB, cuyo flujo de trabajo se compone de 5 etapas (Figura 1):

- Preprocesamiento: debido a la complejidad morfológica de esta especie, particularmente la presencia de aerotopos, el preprocesamiento resultó esencial. Se implementó una conversión a escala de grises y un filtro Gaussiano con kernel 9x9 y $\sigma=2$ para mitigar la textura intracelular sin degradar el gradiente de la pared celular.
- Detección de posibles cianobacterias: para la detección de candidatos, se empleó la Transformada de Hough Circular, optimizando parámetros según la morfología de la especie, una resolución de acumulador $dp=1,1$ para alta sensibilidad; una distancia mínima ($minDist=30$) como restrictor biológico para evitar redundancias; un rango biométrico de radios ($20 \leq r \leq 60$) para

excluir otros géneros y detritos; y umbrales de Canny e identificación ($param1=45$, $param2=30$) optimizados tras un proceso de validación empírica.

- Refinamiento espacial: el pipeline integró un refinamiento espacial mediante la aplicación del algoritmo DBSCAN ($\epsilon=364$, $min_samples=6$). Esta técnica permitió discriminar células reales de artefactos aislados (burbujas o sedimentos) al exigir una asociación espacial densa, característica intrínseca de los filamentos, eliminando elementos aislados que constituyen ruido de fondo.
- Segmentación focalizada: a partir de los clústeres identificados, se define una Región de Interés (ROI) en el área de mayor concentración celular. En esta fase, se realiza un re-muestreo con una resolución de acumulador incrementada ($dp=1,3$) lo cual permitió recuperar células solapadas omitidas en la vista global.
- Cuantificación y generación de métricas de biovolumen: integra un factor de conversión basado en la escala detectada para transformar unidades de píxeles a micrómetros (μm), según el tipo de especie se elige la forma geométrica más próxima; en este caso de esfera, se toman las medidas lineales requeridas dado por los radios detectados (percentiles) para calcular el volumen celular y se lo multiplica por la abundancia para obtener el biovolumen, proporcionando una métrica conservadora y robusta alineada con protocolos internacionales (Bonilla & O'Farrell, 2023).

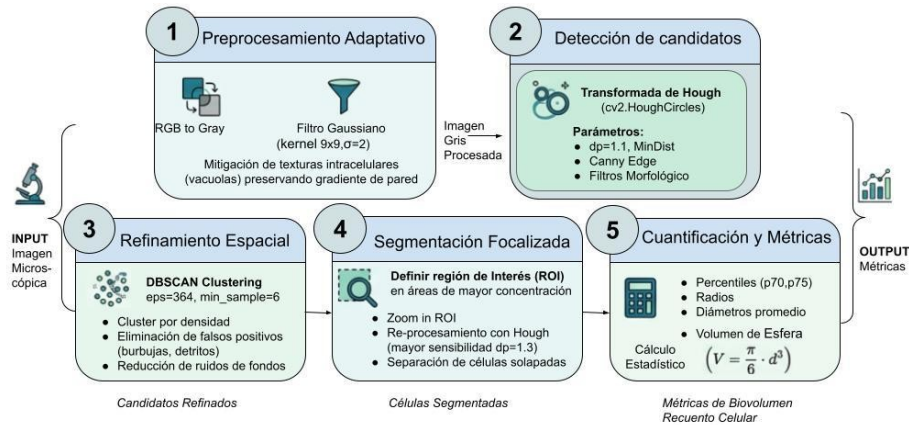


Fig. 1. Workflow del análisis automatizado de cianobacterias.

3 Resultados y análisis

El dataset propio se conformó con 26 imágenes de alta resolución, seleccionadas a partir de los campos donde se registró la presencia de *Dolichospermum circinale* sobre un total de 197 campos inspeccionados, utilizando una cámara con un volumen de 2 cm^3 y un área de $530,93\text{ mm}^2$, se registraron un total de 705 células con un volumen celular promedio de $136,732\text{ }\mu\text{m}^3$ estimando la densidad celular en $4840,00\text{ cel/ml}$ y el biovolumen en $0,662\text{ mm}^3/\text{L}$. Para mayor detalle de los datos recolectados en el

laboratorio y de las métricas generadas, se adjunta un enlace a una planilla¹. La eficacia del flujo de trabajo propuesto se ilustra en la Figura 2 de izquierda a derecha. Durante la etapa de preprocesamiento, se logró una reducción significativa del ruido (Fig. 2.a y 2.b) preservando la integridad de los bordes celulares ya que un suavizado excesivo eliminaría los bordes, mientras que uno insuficiente generaría falsos positivos (Gonzalez & Woods, 2018). Para la detección inicial (Fig. 2.c y 2.d), la Transformada de Hough Circular operó con parámetros optimizados de resolución ($dp=1,1$), gradiente y distancia mínima entre centros ($minDist=30$), esto se aprecia en el centro de las imágenes con la cantidad de células detectadas inicialmente. Luego se observó una mejora sustancial en la detección durante las etapas subsiguientes (Fig. 2.e y 2.f); específicamente, al identificar la Región de Interés (ROI) basada en la alta densidad celular, característica intrínseca de la morfología de los filamentos. Para ello se aplicó el algoritmo DBScan y se buscó el área de mayor concentración, implementado mediante un bucle con radios de rango dinámico. En dicha región, el re-muestreo con parámetros de mayor sensibilidad ($dp=1,3$) permitió recuperar células solapadas que habían sido omitidas en el análisis global y al mismo tiempo descartar los artefactos aislados (burbujas o sedimentos) como se observa a la derecha de las imágenes. Finalmente, una vez detectadas las células, el algoritmo cuantificó las métricas biométricas necesarias para el cálculo del biovolumen.

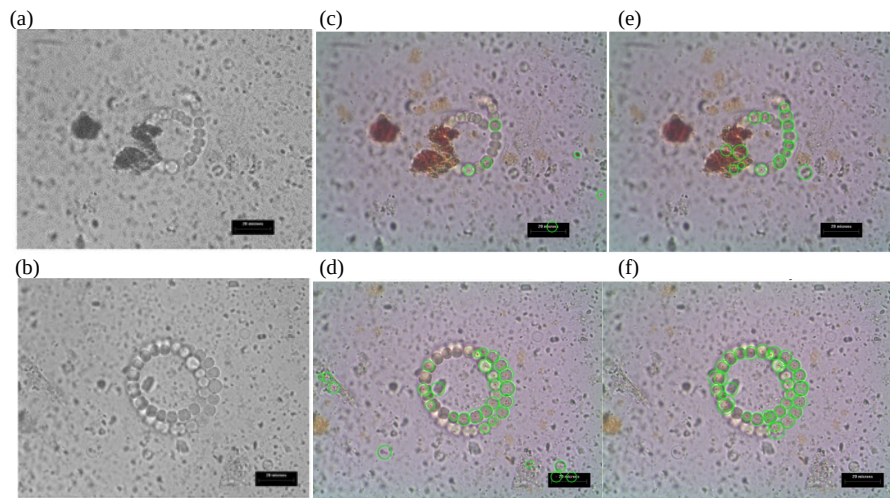


Fig. 2. Imágenes del dataset, mostrando, de izquierda a derecha, el preprocesamiento, la detección celular inicial y la segmentación celular focalizada sobre la región de interés ROI. Las escalas representan $20\ \mu m$.

En la Tabla 1 se presentan las métricas obtenidas manualmente en el laboratorio y mediante nuestra propuesta automatizada, referidas a la longitud, ancho celular promedio, volumen celular promedio, recuento celular y biovolumen.

1 Enlace: https://github.com/davidgonzalomero/jaiio2026_framework

Tabla 1. Comparación de resultados: método manual en laboratorio vs. automático

Método	Long. cel. prom. (h)	Ancho cel. prom. (d)	Vol. cel. prom. (μm^3)	Dens. celular	Biovol. (mm^3/L)	Tiempo de proc.
Manual	6,6	6,14	136,73	4840,00	0,662	1 día
Automatizado	6,48	6,48	142,47	6326,36	0,918	03:18 min

La comparación entre ambos enfoques evidenció concordancia en los resultados obtenidos con ambas metodologías, ya que las métricas de recuento celular (6326,36 *cel/ml*) y de biovolumen (0,918 mm^3) estimadas por el algoritmo fueron similares a las obtenidas mediante el análisis manual en laboratorio (4840,00 *cel/ml* y 0,662 mm^3 respectivamente). No se encontraron diferencias significativas ni en el diámetro celular (p-valor = 0,3879) ni en la densidad celular (p-valor = 0,5026), ambas variables utilizadas como base para el cálculo del volumen celular y, con ello, del biovolumen. Esto representa una leve sobreestimación del algoritmo de 0,256 mm^3 a nivel de biovolumen y del 30% en términos de densidad celular lo cual indica un enfoque conservador ante condiciones de riesgo. Comparado con otros trabajos, Ventura et al. (2025) lograron una aproximación del 80%, resultando en una subestimación del 20% medido en término de segmentación, paso técnico previo necesario para el cálculo del biovolumen. Baek et al. (2020), por su parte, han logrado una aproximación cercana al 85% con una diferencia expresada en números absolutos de células, siendo el error de 23 células por muestra realizada, aunque tampoco establecen comparaciones en términos de biovolumen. Finalmente, es importante mencionar que, si bien el algoritmo demostró buenos resultados en una única especie, aún no tiene la capacidad de estimar métricas en contextos multiespecie con morfologías diferentes, siendo esta nuestra línea de trabajo actual que se mencionará brevemente a continuación.

4 Conclusiones y trabajo futuro

Se propuso un framework para automatizar el cálculo de métricas para el monitoreo de cianobacterias a partir de imágenes microscópicas. Los resultados alcanzados sobre un dataset propio son prometedores, ya que la comparación entre enfoque automático y manual no mostraron diferencias significativas sobre las métricas de recuento y de biovolumen sobre *Dolichospermum circinale*.

Actualmente, se está trabajando en la incorporación de nuevas especies de cianobacterias y la culminación de pruebas numéricas pendientes. En este sentido hemos desarrollado un modelo de aprendizaje profundo pre-entrenado YOLO v11 con ajuste fino de parámetros e hiperparámetros, a efectos de no iniciar con un entrenamiento desde cero por cuestiones de tiempo y costo computacional (GPU, TPU). Para estos nuevos experimentos se ha ampliado el dataset con imágenes de muestras con otras cinco especies de cianobacterias formadoras de floraciones en la Argentina (*Raphidiopsis mediterranea*, *Planktothrix agardhii*, *Sphaerospermopsis aphanizomenoides*, *Anabaenopsis cunningtonii* y *Anabaenopsis elenkinii*), como se muestra en la Figura 3, donde el etiquetado, el preprocesado y el entrenamiento de modelos basados en DL son necesarios.

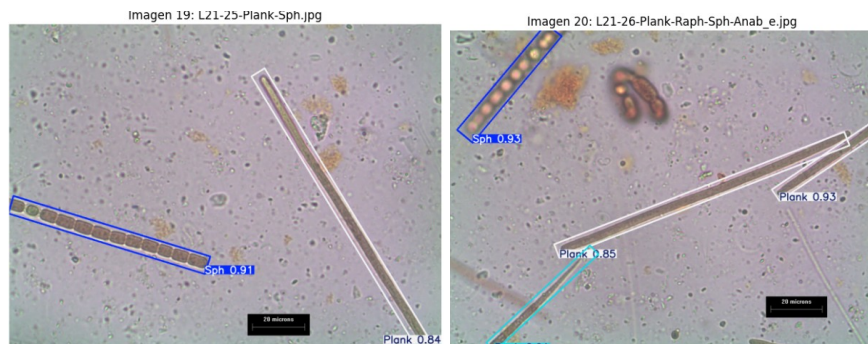


Fig. 3. Imágenes de muestras de lagunas pampeanas con diferentes tipos de cianobacterias. Las escalas representan 20 μm .

El pipeline continúa con la segmentación celular dentro de los *bounding boxes* detectados por el modelo para luego cuantificar y generar las métricas anteriormente indicadas. De esta manera, se espera que el pipeline mejorado forme parte de una herramienta web robusta, escalable y de bajo costo, superando las limitaciones observadas en herramientas semi-automáticas o de asistencia (Zohary et al., 2016; Vadrucchi et al., 2018) para su aplicación en el monitoreo sanitario de cuerpos de agua locales.

Referencias

- Aguilera, A., Haakonsson, S., Martin, M.V., Salerno, G.L. and Echenique, R.O. (2018). *Bloom-forming cyanobacteria and cyanotoxins in Argentina: a growing health and environmental concern*. *Limnologia* 69, 103–114. <http://dx.doi.org/10.1016/j.limno.2017.10.006>
- Aguilera, A., Almanza, V., Haakonsson, S., Palacio, H., Benítez Rodas, G., Capelo-Neto, J., Urrutia, R., Aubriot, L. and Bonilla, S. (2023). *Cyanobacterial bloom monitoring and assessment in Latin America*. *Harmful Algae* 125, 102429. <https://doi.org/10.1016/j.hal.2023.102429>.
- Baek, S., Pyo, J., Pachepsky, Y., Park, Y., Ligaray, M., Ahn, C., Kim, Y., Chun, J. and Cho, K. (2020). *Identification and enumeration of cyanobacteria species using a deep neural network*. *Ecological Indicators* 115, 106395. <https://doi.org/10.1016/j.ecolind.2020.106395>.
- Bonilla, S. y O'Farrell I. (2023). *La importancia de usar el biovolumen en estudios de fitoplancton y monitoreo ambiental de cianobacterias*. *Ecología Austral* 33:558-566. <https://doi.org/10.25260/EA.23.33.2.0.2148>.
- Chorus, I. and Bartram J. (Eds.). (1999). *Toxic Cyanobacteria in Water: A Guide to their Public Health Consequences*. Monitoring and Management (1st ed.). CRC Press. Organización Mundial de la Salud.
- Chorus, I. and Welker, M. (2021). *Toxic Cyanobacteria in Water* (2nd edition). CRC Press, Boca Raton (FL), on behalf of the World Health Organization, Geneva, CH.
- Gandola, E., Antonioli, M., Traficante, A., Franceschini, S., Scardi, M. and Congestri, R. (2016). *ACQUA: Automated cyanobacterial quantification algorithm for toxic filamentous genera using spline curves, pattern recognition and machine learning*. *Journal of Microbiological Methods* 124, 48-56. <https://doi.org/10.1016/j.mimet.2016.03.007>.

- Gonzalez, R. and Woods, R. (2018). *Digital Image Processing* (4th ed.). Pearson. (Base teórica de filtros y gradientes).
- Guiry, M. and Guiry, G. (2026). *AlgaeBase*. World-wide electronic publication, National University of Galway. <https://www.algaebase.org>. Último acceso el 31/03/2026.
- Hillebrand, H., Dürselen, C., Kirschtel, D.B., Pollinger, U. and Zohary, T. (1999). *Biovolume calculation for pelagic and benthic microalgae*. *Journal of Phycology* 35:2. <https://doi.org/10.1046/j.1529-8817.1999.3520403.x>.
- Madkour, D., Shapiai, M., Mohamad, S., Aly, H., Ismail, Z. and Ibrahim, M. (2023). *A Systematic Review of Deep Learning Microalgae Classification and Detection*. *IEEE Access* 11, 57529-57555. <https://doi.org/10.1109/ACCESS.2023.3280410>.
- Mosleh, M., Manssor, H., Malek, S., Milow, P. and Salleh, A. (2012). *A preliminary study on automated freshwater algae recognition and classification system*. *BMC Bioinformatics* 13 (17), S25. <https://doi.org/10.1186/1471-2105-13-S17-S25>.
- O'Farrell, I., Motta, C., Forastier, M., Polla, W., Otaño, S., Meichtry, N., Devercelli, M. and Lombardo, R. (2019). *Ecological meta-analysis of bloom-forming planktonic Cyanobacteria in Argentina*. *Harmful Algae* 83, 1–13. <https://doi.org/10.1016/j.hal.2019.01.004>.
- Schulze, K., Tillich, U., Dandekar, T. and Frohme, M. (2013). *Planktovision - an automated analysis system for the identification of phytoplankton*. *BMC Bioinformatics* 14, 115. <https://doi.org/10.1186/1471-2105-14-115>.
- Utermöhl, H. (1958). *Zur vervollkommnung der quantitativen phytoplankton-methodik: Mit 1 Tabelle und 15 abbildungen im Text und auf 1 Tafel*. Internationale Vereinigung für theoretische und angewandte Limnologie: Mitteilungen 9 (1), 1-38.
- Vadrucci, M., Roselli, L., Castelluccia, D., Di Festa, T., Donadei, D., Florio, M., Longo, E., D'Arpa, S., Maci, F., Ranieri, S., Spinelli, M., Pastorelli, A. and Ungaro, N. (2018). *PhytoNumb3rs: An easy-to-use computer toolkit for counting microalgae by the Utermöhl method*. *Ecological Informatics* 46, 147-155. <https://doi.org/10.1016/j.ecoinf.2018.06.007>.
- Ventura, L., Savardi, M., Adami, N., Prein, G., Santin, A. and Benini, S. (2025). *Enhancing DL-based Cell Segmentation of Microalgae with Classical Image Processing Priors*. *Proceedings of EUSIPCO*, 651-655. ISBN: 978-9-46-459362-4.
- Zohary, T., Shneor, M. and Hambright, K. (2016). *PlanktoMetrix – a computerized system to support microscope counts and measurements of plankton*. *Inland Waters*, 6(2), 131–135. <https://doi.org/10.5268/IW-6.2.965>.