

# HERRAMIENTA GRÁFICA PARA LA CARACTERIZACIÓN DE CULTIVOS DE MICROALGAS BASADA EN REDES NEURONALES ARTIFICIALES

Pablo Otálora<sup>1</sup>, José Luis Guzmán<sup>1</sup>, Francisco Gabriel Acién<sup>2</sup>, Manuel Berenguel<sup>1</sup>

<sup>1</sup>Departamento de Informática, Universidad de Almería, ceiA3, CIESOL, Ctra. Sacramento s/n, 04120 Almería, España. Tel: +34 950 214133; {p.otalora,joseluis.guzman,beren}@ual.es

<sup>2</sup>Departamento de Ingeniería Química, Universidad de Almería; facien@ual.es

## Resumen

*En este trabajo se aborda el problema de caracterización de muestras de microalgas mediante modelos de redes neuronales artificiales. Se presenta una interfaz gráfica combinada con los modelos que permite predecir la composición de las especies *Chlorella vulgaris* y *Scenedesmus almeriensis* a partir de datos relativos a la muestra, como pueden ser imágenes de las células o propiedades descriptivas de estas. Los resultados del trabajo proporcionan una herramienta completa para abordar el problema sin necesidad de conocimientos del lenguaje de programación, de forma rápida, sencilla e intuitiva.*

**Palabras clave:** Microalgas, redes neuronales, interfaz gráfica, biotecnología.

## 1 Introducción

El proceso de producción de microalgas está cobrando cada vez más importancia en los últimos años, debido a su utilidad en gran variedad de aplicaciones relacionadas con los principales problemas de la sociedad. Principalmente, se trata de un proceso útil para el tratamiento de aguas residuales, la mitigación de emisiones de  $CO_2$  o la producción de biocombustibles o productos de alto valor, para el ámbito de la nutrición humana, alimentación animal o cosmética [2, 16, 21]. Las microalgas son organismos fotosintéticos con la capacidad de crecer y reproducirse en ambientes sin necesidad de agua limpia, con una gran tasa de cultivo por superficie y con una amplia variedad de condiciones [11].

Sin embargo, para asegurar la competitividad del proceso es necesario maximizar su productividad a través de la optimización del diseño y el control del proceso [3]. Esta solo puede ser alcanzada logrando una caracterización suficiente del proceso, la cual no es trivial, debido a su marcado carácter biológico [9]. Existen una gran cantidad de variables que influyen en el proceso, siendo una de las fundamentales la especie de microalga que se cultiva. Existen más de 30000 especies conocidas, de las cuales menos de 20 son explotadas

comercialmente [10]. La decisión sobre la especie que se cultiva se basa en el producto que se desee obtener del proceso. Además, las condiciones óptimas de producción varían en función de la especie cultivada, por lo que conocer la composición del cultivo es un paso fundamental para lograr la correcta caracterización del proceso [17].

Tradicionalmente, la monitorización de la composición se realiza usando un microscopio óptico, manualmente y basándose en las diferencias morfológicas entre las especies. Como es evidente, no es un proceso trivial, pues requiere mano de obra muy cualificada y mucho tiempo, además de ser imposible de automatizar [4, 5]. En trabajos previos se han desarrollado métodos más automatizados, basados en espectroscopía de absorción o citometría de flujo [19, 22]. Ambos métodos presentan claras ventajas frente al análisis manual, pero son relativamente costosos. También se han empleado métodos basados en técnicas de captación y procesamiento de imágenes para el análisis de imágenes ópticas [17].

Por otro lado, debido al incremento de la capacidad computacional, y sobre todo, al inmenso y creciente volumen de datos que se genera y se gestiona en los últimos años, los modelos de aprendizaje computacional basados en datos están siendo usados en casi cualquier ámbito [13]. Estas técnicas permiten obtener modelos muy precisos, a menudo más allá de la comprensión humana, haciendo uso únicamente de datos, sin necesidad de poseer un conocimiento exhaustivo del sistema modelado. Concretamente, las redes neuronales artificiales están demostrando muy buenos resultados para la obtención de modelos de todo tipo.

Para el problema de caracterización de la composición del cultivo, se han desarrollado modelos de red neuronal basados en datos de absorción de luz o procesamiento de imágenes para identificación y clasificación de distintas especies [6, 8, 18]. En [15] se presentan dos modelos de red neuronal basados en imágenes de las células de las microalgas, así como en rasgos descriptivos de estas, en combinación con el dispositivo *FlowCAM* [1]. Ambos modelos fueron desarrollados haciendo uso de un gran volumen de datos, y validados con

muestras mixtas, demostrando ser una herramienta bastante útil para afrontar este problema.

Sin embargo, el modelo descrito es un programa codificado en MATLAB. Si se desea emplear como una herramienta de fácil empleo y que no requiera conocimientos de informática, es muy recomendable el desarrollo de una interfaz que permita al usuario caracterizar cualquier mezcla a partir de los datos que el dispositivo de adquisición proporciona. Con esto en mente, este trabajo tiene como objetivo el desarrollo de una herramienta gráfica a partir de los modelos ya disponibles, con el fin de proporcionar una metodología útil, sencilla y completa para la caracterización de mezclas.

## 2 Materiales y métodos

En esta sección se presentarán las especies de microalgas consideradas por la herramienta, así como las herramientas empleadas para la adquisición de datos y el desarrollo de los modelos.

### 2.1 Especies de microalgas

Las especies empleadas para el desarrollo de los modelos en [15], y por tanto las que la herramienta será capaz de caracterizar son *Chlorella vulgaris* y *Scenedesmus almeriensis*. Se trata de dos de las especies más empleadas a nivel comercial, debido a su elevada tasa de crecimiento y su tolerancia a amplios rangos de pH, oxígeno disuelto y temperatura. Son especies óptimas para su cultivo en reactores abiertos.

*Chlorella vulgaris* (Figura 1) es una de las especies más producidas, con gran interés en el ámbito de la nutrición humana como suplemento dietético. En cuanto a *Scenedesmus almeriensis* (Figura 2), se trata de una especie menos extendida pero muy útil como fuente de luteína en ese mismo ámbito [7, 14].

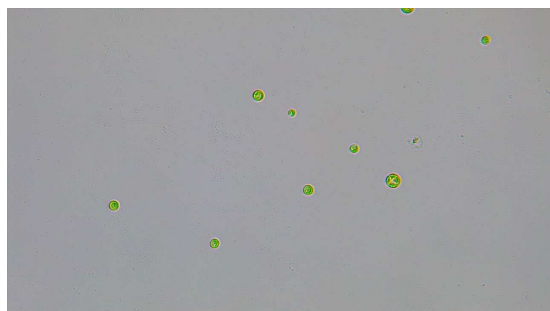


Figura 1: Muestra de *Chlorella vulgaris*.

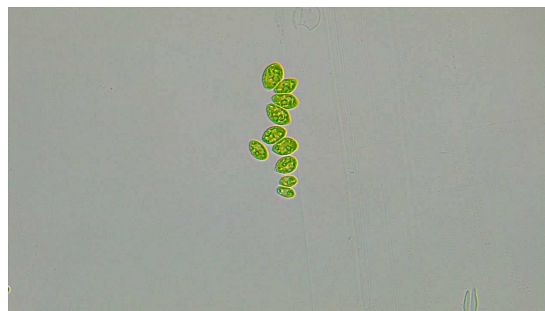


Figura 2: Muestra de *Scenedesmus almeriensis*

### 2.2 FlowCAM

Los modelos desarrollados en [15] emplean como entradas imágenes individuales de las células de las muestras o rasgos descriptivos de estas, por lo que la herramienta gráfica deberá recibir del usuario alguno de estos tipos de datos. En el trabajo citado estos fueron adquiridos mediante *FlowCAM* [1]. Se trata de un dispositivo de análisis de partículas capaz de detectar y analizar cada una de las células de un fluido en movimiento.

El dispositivo impulsa la muestra mediante una pequeña bomba hacia una cámara con una lente microscópica en la cual toma imágenes a nivel microscópico con cierta frecuencia. El software del dispositivo analiza las imágenes, detectando y segmentando cada célula individual, y obteniendo una serie de propiedades o rasgos descriptivos de estas. Las salidas que devuelve la herramienta son un *collage* con todas las imágenes individuales detectadas en formato *.tif* como la mostrada en la Figura 3, y una hoja de datos con las propiedades de cada célula en formato *.xlsx*.

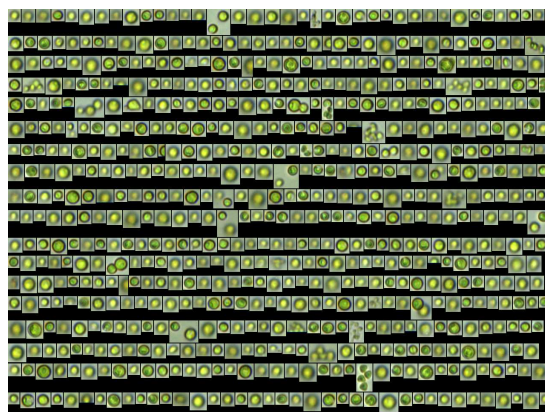


Figura 3: *Collage* de imágenes individuales proporcionado por *FlowCAM*.

De cara al uso de la herramienta gráfica, será imprescindible el uso de uno de estos tipos de datos. Ambos pueden ser obtenidos por *FlowCAM*, aunque existen otros dispositivos de de-

tección y segmentación de imágenes de células. Sin embargo, es imprescindible que el formato de estos datos sea en *collages .tif*. En cuanto a las propiedades de las células, si se desea usar estos datos para la caracterización de la muestra, sí que deberán ser obtenidos mediante *FlowCAM*, pues se emplearán concretamente los rasgos proporcionados por el dispositivo.

### 2.3 Toolboxes de MATLAB

Todo el desarrollo de este trabajo ha sido realizado en MATLAB. Los modelos obtenidos en [15] fueron construidos y entrenados haciendo uso del “*Deep Learning Toolbox*” [12]. Este toolbox permite diseñar las estructuras de las redes neuronales capa por capa, establecer las opciones de entrenamiento, llevar este a cabo y, por supuesto, emplear las redes neuronales entrenadas para posteriores predicciones. Para este trabajo, permitirá realizar las predicciones con los modelos previamente desarrollados.

Para trabajar con las imágenes proporcionadas por *FlowCAM* se hizo uso del “*Image Processing Toolbox*” [20]. Este conjunto de funciones permite una gran variedad de técnicas para el procesamiento, análisis y visualización de imágenes. Durante este trabajo, es fundamental para segmentar y transformar los *collages* que la herramienta recibirá como entrada.

Finalmente, la herramienta de MATLAB más empleada durante este trabajo ha sido el “*App Designer*”. Esta permite construir de manera sencilla interfaces gráficas con distintos elementos, tales como figuras, botones o tablas. En estas se pueden desplegar funciones que permitan al usuario interactuar con algunos de los elementos constituyentes, y mostrando los resultados o la información pertinente por otros.

## 3 Redes neuronales empleadas

Toda la herramienta está basada en los modelos de clasificación de cepas desarrollados y descritos en [15]. Ambos modelos están basados en redes neuronales artificiales, pero emplean como entradas distintos tipos de datos y presentan estructuras diferentes. Los dos modelos distinguen entre las mismas especies de microalgas y fueron entrenados y validados con las mismas muestras.

### 3.1 Modelo basado en rasgos

En primer lugar, se desarrolló un modelo de clasificación empleando las propiedades proporcionadas por *FlowCAM* como entradas. Estas son 30 variables numéricas independientes entre sí, resultando

como salidas dos variables con valores de 0 a 1, que se pueden considerar el “grado de similitud” de dicha entrada con cada una de las salidas. Las variables de salida son las mismas para ambas redes.

De este modo, la estructura de este primer modelo es la representada en la Figura 4. El modelo consta de una capa de entrada de 30 nodos (uno por cada entrada), una capa oculta de 25 nodos, y una capa de salida de 2 nodos (uno por cada salida). El modelo fue entrenado con muestras puras demostrando una precisión del 96.6%, y fue complementado con un umbral de clasificación para la salida “*Chlorella*”. Los resultados de validación de este modelo con muestras mixtas se muestran en la Figura 5. En esta gráfica se presenta la concentración predicha de *Chlorella vulgaris* frente a la real para las distintas mezclas validadas. Como se observa, se trata de un modelo muy preciso para todo el espectro de mezclas.

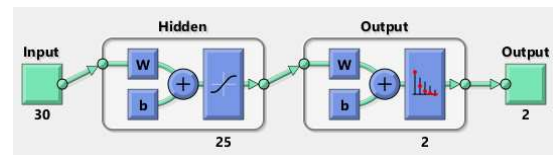


Figura 4: Estructura de la red neuronal basada en rasgos.

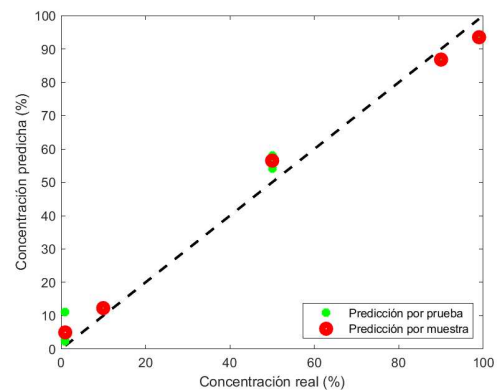


Figura 5: Resultados de validación de la red neuronal basada en rasgos.

### 3.2 Modelo basado en imágenes

El segundo modelo emplea como entradas las imágenes de las células individuales. Puesto que *FlowCAM* proporciona los *collages* con dichas imágenes, es preciso llevar a cabo un proceso de segmentación previo. De este modo, se pueden obtener las imágenes individuales con las que trabaja la red. Estas deben estar normalizadas a un tamaño de  $227 \times 227 \times 3$ .

En cuanto a la estructura de la red, se trata de una estructura ya existente ligeramente modificada y reentrenada, la de la red *AlexNET*. Se trata de una red neuronal convolucional de 25 capas, con el mismo tamaño de entrada que las imágenes normalizadas, y con un tamaño de la capa de salida reducido a 2, correspondiente con el número de clases identificadas.

La precisión del modelo durante su entrenamiento con muestras puras fue del 98.4%. Al igual que en el caso anterior, al modelo se le aplicó un umbral de clasificación para la salida “*Chlorella*”, obteniendo los resultados de validación mostrados en la Figura 6. Esta gráfica, muestra la concentración predicha de *Chlorella vulgaris* frente a la real para las mezclas analizadas haciendo uso de este modelo. Se puede apreciar que los resultados en este caso también se aproximan mucho a la realidad, resultando un modelo muy preciso.

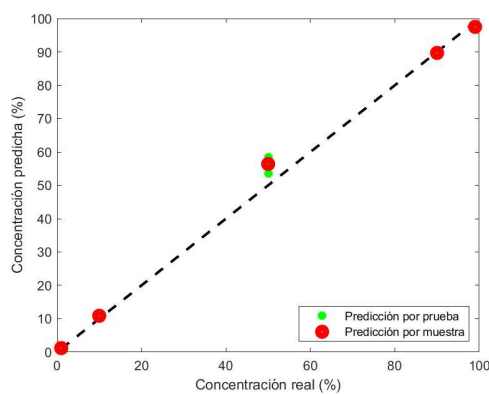


Figura 6: Resultados de validación de la red neuronal basada en imágenes.

#### 4 Herramienta gráfica

El objetivo de la herramienta es que permita, de manera sencilla, que el usuario pueda introducir los datos relativos a las muestras en cualquiera de las dos formas, bien sea un *collage .tif* o las propiedades obtenidas por *FlowCAM* en formato *.xlsx*. A partir de estos datos, el modelo debe presentar de forma gráfica y clara la composición de la muestra analizada. La Figura 7 muestra la interfaz completa de la herramienta.

De este modo, la herramienta tiene que, en primer lugar, dejar que el usuario seleccione el tipo de dato que va a usar. Esto se realiza mediante un grupo de botones de opciones que permiten la selección de una única opción (1). Una vez seleccionado, el usuario puede pulsar el botón “Add new dataset” (2) para que el programa abra el explorador de archivos y deje seleccionar los datos de entrada. Si se va a emplear el modelo basado

en rasgos, el explorador detectará los archivos *.xlsx* de cada carpeta para que el usuario seleccione los que quiere usar. Si por el contrario se desea predicción basada en imágenes, el usuario debe seleccionar la carpeta que contenga todos los archivos *.tif*. La ventana de texto (3) mostrará el path completo del archivo, o de la carpeta que contiene las imágenes.

Tras esto, el usuario puede pulsar el botón “Classify” (4) para que la herramienta dé comienzo a la clasificación. En función de la opción seleccionada, se carga un modelo u otro. Se verifica que el formato de los datos de entrada es correcto para cada uno de los casos, y se cargan dichos datos. En el caso del modelo basado en imágenes, dicha carga lleva consigo la segmentación de los datos, por lo que es un proceso más lento. Una vez cargados, se realiza la predicción con el modelo de red neuronal y se aplican los umbrales de clasificación. La ventana de texto (5) proporciona información sobre el proceso, indicando si se están cargando los datos, si se está realizando la clasificación o si ha finalizado el proceso. En el caso de la red basada en imágenes, también muestra el porcentaje de datos cargados.

Por último, la tabla (6) muestra la composición de cada especie en tanto por ciento, mientras que la figura (7) presenta un histograma con dichas composiciones. El tamaño de las imágenes de cada especie también será proporcional a su composición. De este modo, de forma muy sencilla el usuario puede conocer la composición de ambas especies para cualquier muestra a partir de los datos pertinentes.

La propia Figura 7 muestra el estado final de la interfaz tras realizar una clasificación de una muestra mixta basada en rasgos. El resultado cuando la clasificación se basa en imágenes es muy similar. El proceso completo desde que se seleccionan los archivos hasta que se muestran los resultados toma más o menos tiempo en función del volumen de datos empleados y del modelo que se use. Para una muestra de 50000 células, la clasificación basada en rasgos tomó menos de 20 segundos, mientras que empleando las imágenes tardó más de cinco minutos. La diferencia de tiempo es significativa, pero es lógica teniendo en cuenta la complejidad de los datos y el preprocesamiento de estos.

Mientras que la clasificación basada en rasgos es compatible únicamente con los datos adquiridos por *FlowCAM*, ya que son las propiedades que el modelo previo emplea en las predicciones, la clasificación basada en imágenes es más fácil de extrapolar a otros dispositivos de adquisición, siempre que sean capaces de segmentar las imágenes de

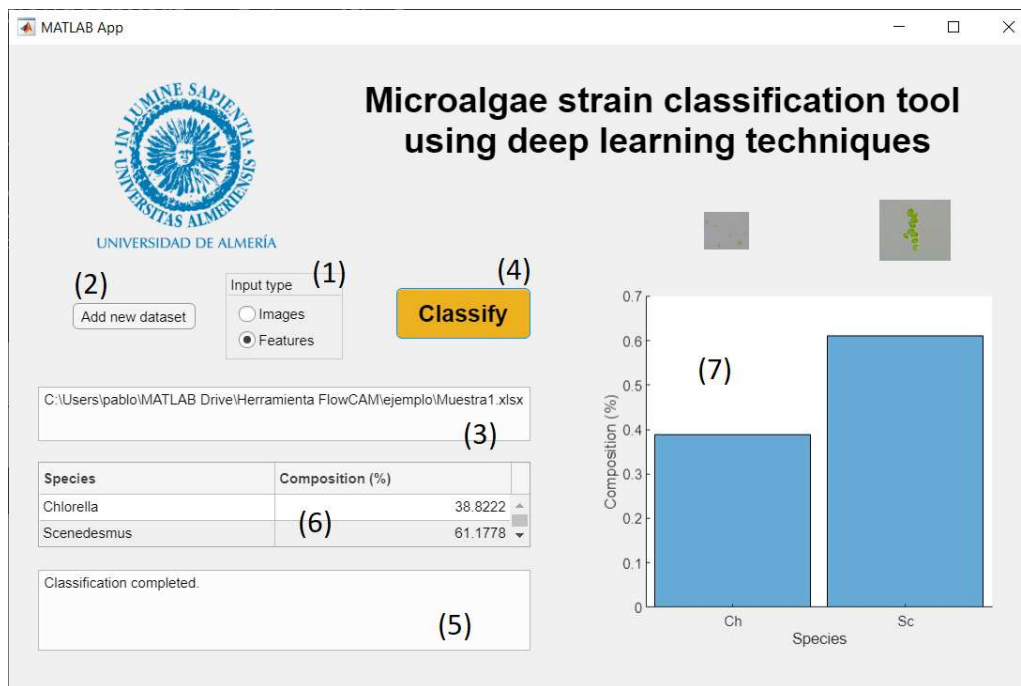


Figura 7: Interfaz de la herramienta para clasificación basada en rasgos.

las células e integrarlas en un *collage .tif*, lo cual supone una clara ventaja en términos de flexibilidad.

## 5 Conclusiones

En este trabajo se presenta una interfaz gráfica para la clasificación de cepas de microalgas basada en redes neuronales artificiales, empleando como entradas imágenes de las células de la muestra o rasgos descriptivos de estas. Los modelos empleados son muy precisos y permiten una caracterización correcta validada con mezclas muy diversas, siendo capaces de distinguir las especies *Chlorella vulgaris* y *Scenedesmus almeriensis*. A partir de estos, la herramienta proporciona una interfaz sencilla para que, únicamente introduciendo el archivo con los datos, se presente al usuario los resultados de la clasificación de forma automática y con la mínima intervención por parte de este.

Se trata de una herramienta fundamental para la implementación de estos modelos en un proceso de producción de microalgas, pues no se puede depender de los conocimientos del entorno de programación por parte de los usuarios para caracterizar cualquier muestra. Es imprescindible que esta caracterización pueda ser llevada a cabo por cualquier usuario, y la interfaz propuesta lo permite.

De cara a futuros trabajos, puede resultar muy interesante ampliar tanto los modelos como la propia interfaz a más especies y géneros de mi-

croalgas, para conseguir una herramienta más completa y que pueda ser libremente empleada en cualquier proceso de producción de microalgas, partiendo siempre de dispositivos de adquisición de datos apropiados.

## Agradecimientos

Este trabajo ha sido financiado con el Proyecto del Plan Nacional DPI2017-84259-C2-1-R del Ministerio de Ciencia, Innovación y Universidades.

## English summary

### GRAPHICAL TOOL FOR CHARACTERISATION OF MICROALGAE CULTURES BASED ON ARTIFICIAL NEURAL NETWORKS

## Abstract

*This work addresses the problem of characterization of microalgae samples through artificial neural network models. It presents a graphical interface in combination with the models that allows the prediction of the species composition of Chlorella vulgaris and Scenedesmus almeriensis from data related to the sample, such as cell images or descriptive properties of the cells. The results of the work provide a powerful tool to approach the problem without the need of any programming language knowledge, in a fast, simple and intuitive way.*

**Keywords:** Microalgae, neural networks, graph-



ical interface.

## Referencias

- [1] FlowCAM <sup>®</sup> Manual, 2011.
- [2] N. Abdel-Raouf, A. A. Al-Homaidan, and I. B.M. Ibraheem. Microalgae and wastewater treatment. *Saudi Journal of Biological Sciences*, 19(3):257–275, jul 2012.
- [3] F. G. Acien, J. M. Fernández Sevilla, and E. Molina Grima. Microalgae: The basis of mankind sustainability. In *Case Study of Innovative Projects - Successful Real Cases*, chapter 7, pages 123–140. 2017.
- [4] E. Álvarez, M. Moyano, Á. López-Urrutia, E Nogueira, and R. Scharek. Routine determination of plankton community composition and size structure: A comparison between FlowCAM and light microscopy. *Journal of Plankton Research*, 36(1):170–184, 2014.
- [5] P. Coltelli, L. Barsanti, V. Evangelista, A. M. Frassanito, V. Passarelli, and P. Gualtieri. Automatic and real time recognition of microalgae by means of pigment signature and shape. *Environmental Sciences: Processes and Impacts*, 15(7):1397–1410, jul 2013.
- [6] I. Correa, P. Drews-Jr, S. Botelho, M. S. De Souza, and V. M. Tavano. Deep learning for microalgae classification. *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, Decem:20–25, 2017.
- [7] J. M. Fernández-Sevilla, F. G. Acien Fernández, and E. Molina Grima. Biotechnological production of lutein and its applications. *Applied Microbiology and Biotechnology*, 86(1):27–40, mar 2010.
- [8] B. M. Franco, L. M. Navas, C. Gómez, C. Sepúlveda, and F. G. Acien. Monoalgal and mixed algal cultures discrimination by using an artificial neural network. *Algal Research*, 38(January):1–7, 2019.
- [9] J. L. Guzmán, F. G. Acien Fernández, and M. Berenguel. Modelling and control of microalgae production in industrial photobioreactors (in Spanish). *Revista Iberoamericana de Automática e Informática Industrial*, 18(1):1–18, 2021.
- [10] K. Heimann and R. Huerlimann. Microalgal classification: Major classes and genera of commercial microalgal species. *Handbook of Marine Microalgae: Biotechnology Advances*, pages 25–41, 2015.
- [11] A. Hernández-Pérez and J. I. Labbé. Microalgas, cultivo y beneficios (Microalgae, culture and benefits, in Spanish). *Revista de Biología Marina y Oceanografía*, 49(2):157–173, 2014.
- [12] M. Hudson, B. Martin, T. Hagan, and H. B Demuth. *Deep Learning Toolbox™ User’s Guide*, 1992.
- [13] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, jul 2015.
- [14] R. A. Kay and L. L. Barton. Microalgae as Food and Supplement. *Critical Reviews in Food Science and Nutrition*, 30(6):555–573, jan 1991.
- [15] P. Otálora, J.L. Guzmán, F.G. Acien, M. Berenguel, and A. Reul. Microalgae classification based on machine learning techniques. *Algal Research*, 55:102256, 2021.
- [16] J. K. Pittman, A. P. Dean, and O. Osundeko. The potential of sustainable algal biofuel production using wastewater resources. *Biore-source Technology*, 102(1):17–25, jan 2011.
- [17] S. Promdaen, P. Wattuya, and N. Sanevas. Automated microalgae image classification. *Procedia Computer Science*, 29:1981–1992, 2014.
- [18] P. Qian, Z. Zhao, H. Liu, Y. Wang, Y. Peng, S. Hu, J. Zhang, Y. Deng, and Z. Zeng. Multi-Target deep learning for algal detection and classification. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020-July:1954–1957, 2020.
- [19] A. Reul, M. Muñoz, B. Bautista, P. J. Neale, C. Sobrino, J. M. Mercado, M. Segovia, S. Salles, G. Kulk, P. León, W. H. D. van de Poll, E. Pérez, A. Buma, and J. M. Blanco. Effect of CO<sub>2</sub>, nutrients and light on coastal plankton. III. Trophic cascade, size structure and composition. *Aquatic Biology*, 22:59–76, nov 2014.
- [20] C. M. Thompson and L. Shure. *Image Processing Toolbox™ User’s Guide*, 1995.
- [21] C. V. González-López, F. García-Cuadra, N. Jawiarczyk, J. M. Fernández-Sevilla, and F. G. Acien-Fernández. Valorization of Microalgae and Energy Resources. In *Sustainable Mobility, IntechOpen*. 2020.

- [22] J. Wang, J. Zhao, Y. Wang, W. Wang, Y. Gao, R. Xu, and W. Zhao. A new microfluidic device for classification of microalgae cells based on simultaneous analysis of chlorophyll fluorescence, side light scattering, resistance pulse sensing. *Micromachines*, 7(11):1–17, nov 2016.



© 2021 by the authors.  
Submitted for possible  
open access publication  
under the terms and conditions of the Creative Commons Attribution CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>).