

# MÁSTER EN NUEVAS TECNOLOGÍAS ELECTRÓNICAS Y FOTÓNICAS

TRABAJO DE FIN DE MÁSTER - CURSO 2026-2027

## PROPUESTA

|                              |                                                                                                     |
|------------------------------|-----------------------------------------------------------------------------------------------------|
| <b>Título:</b>               | <i>Predicción nivel de riesgo biológico en embalses mediante inteligencia artificial y genética</i> |
| <b>Título en inglés:</b>     | <i>Predicting biological risk levels in reservoirs using artificial intelligence and genetics</i>   |
| <b>Tutor/es:</b>             | José A. López Orozco; Raul Fernández Fernández                                                      |
| <b>Correos-e:</b>            | <a href="mailto:jalo@ucm.es">jalo@ucm.es</a> ; <a href="mailto:raufer06@ucm.es">raufer06@ucm.es</a> |
| <b>Lugar de realización:</b> | Departamento de Arquitectura de Computadores y Automática                                           |

### Resumen:

Uno de los riesgos biológicos más destacables en embalses es la existencia de toxinas en el agua. Una de las toxinas que mayormente aparecen son las cianotoxinas (producidas por cianobacterias como por ejemplo Microcistis, Aphanizomenon o Anabaena). Sin embargo, la existencia de cianobacterias en el agua no implica necesariamente que éstas produzcan las toxinas puesto que existen multitud de factores medioambientales y del entorno que pueden llevar a su producción y liberación en el agua.

En este Trabajo fin de Máster (TFM) se propone utilizar métodos de inteligencia artificial, como por ejemplo redes neuronales, para encontrar qué factores medioambientales y de entorno pueden provocar la generación de estas sustancias biológicas nocivas.

### Metodología:

Para asegurar la viabilidad del trabajo en el plazo de un TFM, nos centraremos únicamente en la toxina denominada microcistina. Además utilizaremos datos de acceso público de muestras biológicas las cianobacterias potencialmente productoras (i.e. que contengan los genes relacionados con la producción de la toxina) y que indiquen si hay microcistinas o no en las muestras. Los objetivos concretos son:

1.- Revisión de repositorios públicos (como por ejemplo el European Nucleotide Archive o el NCBI/GenBank) que contengan secuencias metagenómicas de muestras de agua dulce para la identificación de las cianobacterias y microcistinas.

2.- Procesamiento bioinformático de los datos. Se utilizará el software SqueezeMeta, que es un pipeline automático y completo para el análisis de metagenómica y metatranscriptómica, y que automatizará todo el proceso de análisis, incluyendo el co-ensamblaje de genes relacionados. El objetivo es obtener perfiles taxonómicos (i.e. qué especies hay en la muestra y su abundancia) y funcionales (qué genes relacionados con la producción de la toxina buscada están expresados).

3.- Construcción de la base de datos. Esta base de datos incluirá la información metagenómica encontrada junto con metadatos ambientales asociados a las muestras (por ejemplo, pluviometría, temperatura, pH, concentración de nutrientes, etc.) y datos del entorno (otros genes o marcadores encontrados destacables), con la finalidad de servir de fuente para el predictor de la probabilidad de riesgo de aparición de toxinas en las muestras de agua.

4.- Predictor (evaluación del riesgo). Se probarán diferentes técnicas (p.e. métodos de selección de características y agrupamiento, redes neuronales o redes bayesianas) para encontrar un predictor que teniendo en cuenta la existencia de genes de cianobacterias productoras y otros aspectos ambientales y del entorno, genere una posible estima del riesgo de la aparición de la toxina.

#### **Conocimientos previos recomendados:**

Conocimientos básicos de programación, y de tratamiento y análisis de datos.

#### **Bibliografía:**

- Jüttner F, Watson SB (2007). Biochemical and ecological control of geosmin and 2-methylisoborneol in source waters. *Appl Environ Microbiol*, 73(14), pp.4395-406. Doi: 10.1128/AEM.02250-06.
- Tamames, J., & Puente-Sánchez, F. (2019). SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Frontiers in Microbiology*. Sec. Systems Microbiology. Volume 9. Doi:10.3389/fmicb.2018.03349
- Otten, T. G., et al. (2023). Innovative Genetic Methods for Predicting Microcystin Toxin Risk: Temporal Dynamics of the mcyE Gene in Drinking Water Reservoirs. *Water Research*.
- Wang, Z., et al. (2022). Revealing Physiochemical Factors and Zooplankton Influencing Microcystis Bloom Toxicity in a Large-Shallow Lake Using Bayesian Machine Learning. *Environmental Science & Technology*.
- Modern, K., et al. (2024). Toward a Predictive Understanding of Cyanobacterial Harmful Algal Blooms through AI Integration of Physical, Chemical, and Biological Data. *ACS EST Water*.
- Claudia Fournier, Raúl Fernandez-Fernandez, Samuel Cirés, José A. López-Orozco, Eva Besada-Portas, Antonio Quesada (2024). LSTM networks provide efficient cyanobacterial blooms forecasting even with incomplete spatio-temporal data. *Water Research*, Volume 267. Doi: 10.1016/j.watres.2024.122553
- Fang, S., et al. (2026). Co-application of pan-genomics and machine learning approaches to investigate the regulatory and evolutionary basis of microcystins production. *Harmful Algae*
- National Library of Medicine. URL: <https://www.nlm.nih.gov/>
- KEGG: Kyoto Encyclopedia of Genes and Genomes. URL: <https://www.genome.jp/kegg/>