Propuesta de Trabajo Fin de Máster
Año académico 2025-2026
MÁSTER EN CIENCIA DE DATOS PARA CIENCIAS EXPERIMENTALES

Proyecto Nº 11
Título: Assessing functional host-microbiome-diet interactions using deep learning models.
Departamento/ Laboratorio: Microbiome and Clinical Physiology
Director: Rafael Valdés Mas   **Correo electrónico:** rvaldesmas@unav.es
Resumen: The **human gastrointestinal tract (GIT) and its microbiome form a metabolically active interface** crucial for digestion, immune regulation, and pathogen defense, with dietary signals shaping gut mucosal function, microbiome composition, and their interactions. Disruptions in this host-microbiome-diet interplay are linked to diseases such as Inflammatory Bowel Disease, celiac disease, and cancer. While DNA sequencing provides limited functional insight, metaproteomics offers a deeper view of the dietary exposome and active microbial and host functions. However, challenges in bacterial species resolution and functional protein identification persisted until recently. **We have developed a metagenome-informed metaproteomics (MIM) methodology capable of simultaneously identifying the majority of intestinal proteins originating from diet, host, and bacterial microbiome**. This methodology relies on classical approaches, including fragment ion matching, mass tolerances, and statistical scoring models, to identify peptides and proteins. These methods efficiently detect modifications and perform open searches but lack the adaptive learning capabilities of deep-learning-based tools, which can model complex patterns and generalize from large datasets.   In this project, **we aim to apply deep-learning-based end-to-end database searches for tandem mass spectrometry**. Unlike classical database searches, deep-learning-based proteomics approaches are not strictly bound by traditional false discovery rate (FDR) correction methods, which may lead to the exclusion of low-scoring but real peptides. Deep learning models may identify peptides by learning spectral patterns, **enhancing sensitivity to low-abundance proteins**. Additionally, deep learning can process raw MS/MS spectra directly without strict database constraints, potentially enabling the detection of novel peptides or PTMs that traditional searches might overlook. Furthermore, dietary intake quantification often relies on subjective methods such as questionnaires or food diaries, which are prone to recall bias, altered compliance, and insufficient resolution. Our approach may improve sensitivity for detecting low-abundance dietary proteins, enhancing our current MIM methodology for dietary assessment and paving the way for a generalizable framework for dietary quantification.

OPTATIVAS RECOMENDADAS   1. Análisis de datos de alto rendimiento   2. Machine Learning   3. Advanced topics in Machine Learning