Explainable AI-powered multimodal data integration for healthcare

Despite numerous efforts in the field of complex diseases to apply the findings of genetic association studies (GWAS) in clinical practice, replication of these findings in diverse populations often shows weaknesses or inconsistencies. This is due to the diversity of symptoms, degrees of severity and responses to treatment that individuals with the same complex disease present, despite sharing the same diagnosis. Until now, the classification of these patients into subtypes has been mainly based on a single approach, either genomics or clinical features, which has been insufficient to fully understand the triggers of the disease, limiting the application of personalised treatments.

In this context, there has been a growing interest in biomedicine in combining machine learning (ML) and artificial intelligence (AI) strategies to address complex diseases. This involves the integration of diverse knowledge domains, such as genetic data, clinical data, questionnaires, sociological data, environmental data and medical images, into a single learning system. However, very few of these systems have focused on unsupervised learning approaches, which generate new knowledge rather than simply reproducing existing knowledge. In addition, most current machine learning models are black boxes that lack a focus on interpretability and do not address the uncertainty inherent in biomedical systems.

To address this challenge, the method known as Phenotype to Genotype Many-to-Many Relationship Analysis (PGMRA), an approach based on machine learning and optimisation, has been developed to discover unsupervised relationships between genetic and phenotypic variables. In this project, we propose to expand this approach to a multimodal context in which genetic data and medical images are integrated. This marks a significant advance in jointly harnessing genetics and medical imaging for a more complete understanding of complex diseases.

Our hypothesis is that improved and personalised treatments for complex diseases can only be achieved with a multifaceted (multiview) and temporal study of the individual's trajectory. The multifaceted view is based on the study of multiple and diverse phenotypic measures called "phenome", the collective characterisation and quantification of groups of biological molecules that give rise to the structure, function and dynamics of an organism or organisms called "omics data", and the phenomic-multiomic associations resulting from the combination of these knowledge domains that will define the architecture of the disease. Moreover, the aggregation of multiple images (functional, structural, volumetric MRI) complemented with neuromodulation tools used in neurosciences (EEG, FNIRS, TMS), will allow to resolve the multifaceted description of the individual, and consequently, to determine a distributed architecture (subtypes) of the disease. Also, the longitudinal evolution of a complex disease architecture under different environmental conditions defines a person's risk trajectories and will help to define the dynamics of that architecture. ML and AI techniques applied to these trajectories will allow predicting person-centred risks and recommending the best treatments and decisions to prevent diseases.

The goal of this proposal is to provide a machine learning (ML) and artificial intelligence (AI) framework to encode information from multiple omics and medical images into structural data and systematically combine that information to provide multifaceted descriptions of complex objects or phenotypes. The ultimate goal is to unveil the genotypic-phenotypic architecture of complex diseases such as Alzheimer's or leukaemia through the integration of omics and neuroscience.

Supervisor(s) name(s), Affiliation(s), eMail address(es) for contact:

Prof. Coral del Val Muñoz
Department of Computer Science and Artificial Intelligence.
Andalusian Research Institute in Data Science and Computational Intelligence.
Instituto de Investigación Biosanitaria de Granada (ibs. GRANADA)
University of Granada
Email: delval@ugr.es

Proposed list of secondments (not mandatory, but recommended if known already):

- Prof. Dr. Silvio C. E. Tosatto (UNIPD)
- Technische Universität München

Main ArchiFun theme involved:

- ☐ Host-pathogen interactions;
- ☐ Mechanisms of bacterial resistance and cancer onsets;
- ☐ Neurodegenerative and autoimmune diseases;
- ☐ **Translational research in prevalent diseases; x**
- ☐ Physiology and ecology;
- ☐ Neurosciences and cognition.