



BioISI - Biosystems & Integrative Sciences Institute

Developing machine learning tools for the analysis of genetic sequencing data

Place of work/: Instituto Nacional de Saúde Doutor Ricardo Jorge/Remote

Supervisors: Hugo Martiniano (external) and Francisco Couto(Internal)

Contact: (Email): hugo.martiniano@insa.min-saude.pt, fjcouto@edu.ulisboa.pt

In the last few years, the use of next generation sequencing (NGS) has resulted in vast volumes of data in both research and clinical diagnosis. One of the challenges in the analysis of NGS-derived data is the large number of genetic variants detected. While some are readily classified as disease-causing, most have uncertain clinical significance, meaning that, in most cases, the establishment of a genetic diagnosis is not possible.

This project aims to develop machine learning approaches to take advantage of integrated data from multiple publicly accessible data sources as a means to address this issue. Our approach is based on integrating sequencing data with data from multiple publicly accessible data sources, to produce a knowledge graph, and applying state-of-the-art graph learning methods to this knowledge graph, to improve the detection of genetic variants associated with a given disease or phenotype.

Depending on the interests of the student, the project can be more oriented to a machine learning or a genetics scope.

Students selected for this project, after thesis registration, are eligible to apply to the BioISI Junior Programme (supporting 8 students with a 6-month Scholarship(BII), being the selection criterium the academic merit of the candidates.



BioISI - Biosystems & Integrative Sciences Institute

Bibliography: (if applicable)