

DataLab Group

Instituto de Ciencias Matemáticas (ICMAT-CSIC)

roi.naveiro@icmat.es

nuria.campillo@csic.es

Madrid

25 de septiembre de 2021

This document presents all the information relative to the project that the DataLab group of ICMAT presents for the AIHUB JAE Intro scholarships call.

Project supervisors

Roi Naveiro, roi.naveiro@icmat.es

Nuria Campillo, nuria.campillo@csic.es

AIHUB Research group

DataLab Group, Institute of Mathematical Sciences (ICMAT-CSIC).

Location of the center where the scholarship will be enjoyed

ICMAT-CSIC, Campus Cantoblanco UAM, C/ Nicolás Cabrera, 13-15, 28049 Madrid, España

Project title

Generative models and Bayesian optimization for *de novo* design with applications to CB2R ligands discovery.

Project Description

Drug design with specific properties entails high research and development costs. Therefore prediction of such properties based on relevant features is essential in such activity. From a computational point of view, this entails two core activities: the identification of drug like molecules that serve as potential candidates and the evaluation of these candidates through some performance measure. The usual approach to identify promising candidates goes as follows: chemical experts identify a set of possible candidates, and quantitative structure–activity relationship models (QSAR models, typically machine

learning ones) are built to screen the candidates and select the most promising ones which will be later synthesized and biologically evaluated. One of the goals of the project is to develop methods to identify promising molecules automatically, reducing the intervention of chemical experts. For that, generative models (such as variational autoencoders) capable of creating drug-like molecules will be trained and optimized in order to generate molecules with optimal performance. Bayesian optimization methods will be used for optimization purposes.

As underlying case study, we shall consider the case of cannabinoid CB2 receptors. These have been instrumental in the treatment of numerous diseases and lack the negative effects of CB1 receptors but their development is extremely costly. On the whole, we aim at developing a decision theoretic QSAR approach to CB2 receptor ligands design based on AI predictive models and a utility model which takes into account the costs and benefits associated with research and development in drug discovery.

Requirements

Knowledge of machine learning and statistics is essential. Working knowledge of Python and/or R required. Knowledge of TensorFlow and/or PyTorch is recommended. No chemical knowledge is required.