



Thesis Project Form

Title (tentative): Machine Learning based approaches to identify non-obvious relationships between clinical and biological features in Chronic Lymphocytic Leukemia (CLL)

Thesis advisor(s): Barla Annalisa, , Andrea Nicola Mazzarello (DIMES)

E-mail: Annalisa.Barla@unige.it

Address:

Phone: (+39) 010 353 6602

Description

Motivation and application domain

As Chronic Lymphocytic Leukemia (CLL) is characterized by an extremely heterogenous clinical and biological course, the proposed master thesis project will help to identify clusters of patients with similar features. Thus, possibly improving prognosis and novel ad hoc therapeutical approaches.

General objectives and main activities

The problem at hand is set within the clinical framework of characterizing Chronic Lymphocytic Leukemia (CLL). The DIMES research laboratory collects a dataset comprising primary CLL B-cell samples extracted from patients' blood. These cells underwent analysis to assess various biologically and clinically significant markers. Additionally, the primary cells were employed in vitro as either controls or were stimulated through membrane receptors that are both clinically and biologically relevant. The identical features measured ex-vivo were re-evaluated, and additional functional attributes, such as proliferation and apoptosis, were also incorporated.

The candidate will study the state-of-the art machine learning methods most suitable for analyzing the data at hand. The candidate will also devise and implement a machine-learning-driven pipeline capable of subtyping patients based on ex-vivo data. The proposed pipeline should also identify the in-vitro condition(s) that most closely correspond and correlate with the ex-vivo data.

Training Objectives (technical/analytical tools, experimental methodologies)

The candidate is expected to learn how to set up a machine learning pipeline, taking into account reproducibility, statistical robustness and performance assessment beyond an accurate and appropriate choice of an optimal algorithm for the task at hand.

Also, the candidate is expected to make good use of his/her background knowledge of the biomedical domain to validate the quality of the output of the machine learning pipeline.

Place(s) where the thesis work will be carried out: MaLGa Center

Additional information

Maximum number of students: 1