Inria

Offer #2024-07370

PhD Position F/M Statistical Learning on Flow Cytometry Data for the early characterization of Acute Myeloid Leukemia (IDP 2024)

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Context

Acute Myeloid Leukemia (AML) is an aggressive form of bone marrow cancer characterized by the proliferation of immature blood cells. The typical treatment is intensive chemotherapy that starts as early as possible. For some patients, this treatment turns out to be ineffective. Alternative treatment and/or inclusion in a clinical trial could be proposed if only these patients could be identified from the diagnosis.

A recent study (Itzykson et al., 2021) proposed a therapeutic decision tool based on cytogenetic and molecular biomarkers (chromosomal abnormalities, mutations). It is able to classify patients in three groups based on the adequacy of intensive chemotherapy (favorable, adverse or intermediate). Unfortunately, these biomarkers are obtained too late to inform the initial therapeutic decision.

In this PhD thesis, the goal is to develop statistical learning approaches for flow cytometry data obtained at diagnosis, in order to predict the cytogenetic and molecular prognosis markers for each patient.

The work is based on the collaboration with the team of Pierre-Yves DUMAS (PU-PH) at Bordeaux University Hospital Center, and implies the Regional Data Registry DATAML (Didi et al., 2024).

Assignment

The first goal is to go beyond the manual treatment of flow cytometry data performed by the clinicians by establishing a **data preprocessing algorithm**. Flow cytometry data appear as large dimensional tables where, for each patient, tens of thousands of cells are individually characterized by two markers of size and granularity, and 10 markers for expression in surface proteins. A first task will focus on cell outliers filtering using a strategy based on unsupervised clustering techniques such as Self-Organizing Maps (Van Gassen et al., 2015). This work will lead to the development of a R library.

The second goal is to develop **deep-learning models for the prediction of the presence of mutations**. Convolutive Neural Networks will be adapted to the specificities of flow cytometry data (e.g robustness with respect to markers permutation), extending the previous work from (Hu et al., 2020). The effect of some settings in the data preprocessing or cell subsampling will be investigated. Interpretability of the predictions will be assessed by permutation methods. Possible further development will aim at predicting a mutation rate (regression) rather than a binary mutation status (classification).

The third goal is to supplement the previous approach to build a **model for the prediction of the chemotherapy-adequacy group from flow cytometry data**. This stratification arise from the combination of a 3-class cytogenetic risk group with some mutation landscapes. First, there will be exploration of strategies for combining mutation models. A second task will focus on the prediction of the cytogenetics risk group (classification). A third task will consist in building a decision tree approach to combine these models. The resulting model will then be validated on an independent dataset.

References :
· Didi et al., 2024. Artificial intelligence-based prediction models for acute myeloid leukemia using real-life data: A DATAML registry study. Leuk Res., 136:107437
· Hu et al., 2020. A robust and interpretable end-to-end deep learning model for cytometry data. Proceedings of the National Academy of Sciences, 117(35), 21373-21380.
· Itzykson et al., 2021. Genetic identification of patients with AML older than 60 years achieving longterm survival with intensive chemotherapy. Blood 138, 507–519.
· Van Gassen et al., 2015. FlowSOM: Using Self-Organizing Maps for visualization and interpretation of cytometry data. Cytometry Part A, 87(7), 636-645.

Main activities

Main activities:

- Perform a bibliography review on learning methods for flow cytometry data
- Develop and validate programs for data processing based on unsupervised techniques (clustering)
- Develop and validate programs for machine and deep learning approaches for classification and regression tasks
- Write reports
- Present the works' progress to partners, to the scientific community.

Skills

Technical skills and level required: Experience in programming (Python and/or R), machine and/or deep learning

Languages: At least intermediate level in English

Relational skills: Adaptability

Other valued appreciated: Integrity, willingness to learn, method

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

- 2100€ / month (before taxs) during the first 2 years,
- 2190€ / month (before taxs) during the third year

General Information

- **Theme/Domain** : Modeling and Control for Life Sciences
  Statistics (Big data) (BAP E)
- **Town/city** : Talence
- **Inria Center** : [Centre Inria de l'université de Bordeaux](#)
- **Starting date** : 2024-10-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2024-05-03

Contacts

- **Inria Team** : [MONC](#)
- **PhD Supervisor** :
  Etchegaray Christele / christele.etchegaray@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

The successfull candidate will have a background in applied mathematics or in mathematical engineering, ideally with an experience in machine and/or deep learning. We are looking for a candidate with appetite for interdisciplinary work with clinicians.

> **Warning** : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Thank you to send:
- CV

- Cover letter
- Master marks and ranking
- Support letter(s)

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.