



Offre n°2024-07370

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

Type de contrat : Fixed-term contract

Niveau de diplôme exigé : Graduate degree or equivalent

Fonction : PhD Position

Contexte et atouts du poste

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).

Mission confiée

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**. Convolutional 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.

Principales activités

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.

Compétences

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

Avantages

- 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

Rémunération

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

Informations générales

- **Thème/Domaine** : Modeling and Control for Life Sciences Statistics (Big data) (BAP E)
- **Ville** : Talence
- **Centre Inria** : [Centre Inria de l'université de Bordeaux](#)
- **Date de prise de fonction souhaitée** : 2024-10-01
- **Durée de contrat** : 3 years
- **Date limite pour postuler** : 2024-05-03

Contacts

- **Équipe Inria** : [MONC](#)
- **Directeur de thèse** : Etchegaray Christele / christele.etchegaray@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

The successful 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.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Thank you to send:

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

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.