Offer #2021-04274

Post-Doctoral Research Visit F/M Statistical learning for the detection and classification of circulating tumor cells in the bloodstream based on their biophysical characteristics

Contract type : Fixed-term contract

Level of qualifications required : PhD or equivalent

Fonction : Post-Doctoral Research Visit

About the research centre or Inria department

The Inria Lille - Nord Europe Research Centre was founded in 2008 and employs a staff of 320, including 280 scientists working in fourteen research teams. Recognised for its outstanding contribution to the socio-economic development of the Hauts-De-France région, the Inria Lille - Nord Europe Research Centre undertakes research in the field of computer science in collaboration with a range of academic, institutional and industrial partners.

The strategy of the Centre is to develop an internationally renowned centre of excellence with a significant impact on the City of Lille and its surrounding area. It works to achieve this by pursuing a range of ambitious research projects in such fields of computer science as the intelligence of data and adaptive software systems. Building on the synergies between research and industry, Inria is a major contributor to skills and technology transfer in the field of computer science.

Context

This project is established in the framework of a consortium with technologists, physiciens, biologists and statisticians. It aims to develop a breakthrough approach for the diagnosis and evaluation of cancer metastasis from blood samples by the first smart MEMS (Micro-Electro-Mechanical-Systems) equipment performing high content biophysical characterization of cells in flow for their identification by statistical learning.

The consortium is a complementary partnership gathering 4 laboratories LIMMS and SMMiL-E (CNRS-IIS, Lille, Tokyo), IEMN (Junia Engineering School), CANTHER (INSERM, University of Lille), MODAL (INRIA) with collaboration with a french company (Asygn) :

LIMMS/CNRS-IIS and SMMiL-E (Dominique Collard and Mehmet Cagatay Tarhan): developed bioMEMS microtechnologies allowing to measure physical characteristics of individual cells in flow

- Pr. Dominique Collard was LIMMS director in Tokyo for more than 10 years, today he is with the mirror structure SMMiL-E in Lille to apply bioMEMS technology in the field of cancer
- Dr. Mehmet Cagatay Tarhan : has 15 years of experience (including 12 in the University of Tokyo) in microsystems for biology and microfluidic, he joint SMMiL-E Lille with a talent grant from Hauts de France region

CANTHER (Chann Lagadec) : where Cells will be provided

- Dr. Chann Lagadec is a biologist specializing in stem and tumor cells (breast cancer), and has established protocols to evaluate their metastatic potential

INRIA-MODAL (Sophie Dabo-Niang) : aims at developing the statistical methodology to identify the most relevant physical characteristics of Cancer tumor cells (CTCs) in order to identify and classify them according to their metastatic potential.

- Sophie Dabo-Niang is specialized in statistical/machine learning for the representation and analysis of random events with temporal or spatial variation (Functional, Spatial Data).

ASYGN : French company located in Grenoble, providing electronic products for high-performance sensor applications and also custom developments.

The consortium project is organized in 3 work packages addressing the MEMS equipment (WP1), Signal processing chain (WP2) and the data analysis (WP3) by statistical/machine learning as shown in the graphical abstract available here.

Electrical and mechanical cell data characteristics :

Most of the data collected in WP1 and WP2 are of continuous nature (temporal curves of amplitudes, rigidity, viscosity,..., with different compression levels,...), raw data (cell size, max of viscosity, max of rigidity; phase, contrast,...) or images data (cell images during compression).

Electrical and mechanical characteristics (multiparametric biophysical phenotypes) of breast cancer cell lines are acquired by BioMEMS. Measurements are currently available [6] for 2 cell lines with known metastatic potential. Other cell lines are currently being analyzed and will be available at the beginning of this postdoc project. All these data will be useful for cell characterization. Additional prostate cancer cell lines and patient CTCs (accepted CTCFIND protocol) will be available for evaluation of statistical algorithms.

- **How to characterize cells**

In fact, the **MEMS sensor** (see WP1 in the graphical abstract) compresses the passing cells. Various electrodes configurations can be designed to measure either electrical or mechanical cell characteristics under no or controlled deformation. A first multifrequency electrical measurement will reveal the cell size, membrane capacitance and sub element response. Downstream electrodes will be actuated in real time according to size enabling the cell mechanical and electrical characterisations under controlled deformation. At the microchannel outlet, an actuated deflector sorts the cell toward distinct reservoirs according to their classification.

The research work on the MEMS aims in finding optimal device design, especially for the moving electrode to reach practicable throughput (100-500 cells / second) with the proper signal generation. Innovative microfluidic development will provide fluid handing in this open channel configuration, tune and control the transporting flow, manage the cell passage in the different sensing point and the final cells collection in the different reservoirs.

The **signal processing chain** pilots the overall operations of the sensor including the challenging real time sensor position tuning, the – in flow - data treatment for cell classification and sorting. The architecture features challenging time-domain data management to detect and pool asynchronous events relative to a single cell among a constant flow of data generated from new coming cells. Embedded data processing will compute in real time the key cell characteristics to tune the sensor configuration (actuated elements) and to class the cell within the **statistical/machine learning** multi parameters landscape (as demonstrated in WP3 part of the graphical abstract).

Assignment

This postdoc project aims at developing a statistical learning method for the detection and classification of circulating tumor cells in the bloodstream based on their biophysical characteristics. This proposal is part of a innovative approach for the diagnosis and evaluation of cancer metastasis from blood samples.

Single cell characterization methods are essential for cancer studies. Cancer tumor cells (CTCs) in the blood are very heterogeneous and present in very small numbers.
Flow cytometry, the standard approach, evaluates a cell population using fluorescent biomarkers. Unfortunately, this method is not adapted to identify cell subpopulations with similar labelling properties or that lose their properties such as CTCs. Metastatic progression is characterized by changes in the shape and structural integrity of cancer cells. Therefore, the mechanical and electrical properties of these cells reflect their mutations ([1]) and are of great potential to better understand their malignant transformation ([2-4]) and thus identify and classify them.

Assignments :

With the help of Sophie Dabo-Niang (leading WP3) and other statisticians of MODAL involved, the recruited person will develop the statistical/machine learning methods enabling the statistical cell identification (classification and prediction) from their biophysical characteristics. The methods will be trained from large-scale, heterogeneous, continuous and spatial datasets including high numbers of cells' electrical and mechanical properties. The statistical analysis will rely on canonical correlation analyses, classification, regression methods for complex heterogenous, large dimension data. Machine learning methods, particurlarly based on Functional Data Analysis (SVM, Random Forest, Boosting classification Trees, Knn classification methods) will be used tp predict the cell type. The mathematical/Statistical work package computes heterogenous data and continuous curves as significant advantages for real time classification avoiding information loss arising in more classical multivariate methods that only consider punctual data. The challenges are, on the one hand, to build mathematical learning tools from a large set of complex, heterogeneous and different dimensional data (some data will be curves, thus high dimensional, others are real valued or images of the cells during compression, ...) adapted to the objective of the accuracy of the prediction of the metastatic potential of a cell and, on the other hand, to implement them efficiently.

For a better knowledge of the proposed research subject

1. Lam, W. A., Rosenbluth, M. J. & Fletcher, D. A. (2007) *Blood* 109, 3505–3508.
2. Suresh, S. *et al. (2005). Acta Biomaterialia* 1, 15–30
3. Takayama, Y., *et al. (2018). Micromachines* 9, 275, 17.
4. Rezard, Q., *et al. (2020).* 24th *Conf. on Miniaturized Syst. for Chemistry & Life Sci.*, 805-806.
5. Dabo-Niang, S.*, et al. (2016). Journal of Multivariate Analysis 147, 168-182 (2016).*
6. Tarhan, M. C. *et al. (2016). Sci. Rep.* 6, 28001

Collaboration:

The recruited person will be in connection with the different partners.

Responsibilities :

The person recruited is responsible for the development and implementation (Python, R) of the mathematical methods for the cell prediction and will take initiatives for the choice of the programming language.

Main activities

Main activities :

- Develop the statistical/machine learning methods (mainly FDA SVM, Random Forest, Boosting classification Trees, Knn classification methods) enabling the statistical cell identification from their biophysical characteristics
- Write articles for publication
- Develop R/Phyton sofware and documentation

Additional activities :

- Present the works' progress to partners, scientific events
- Analyse the requirements of partners

Skills

Technical skills and level required : Experience in statistical modelling (in particular FDA) and package development

Languages : French, English

Relational skills : Autonomy, Rigor, Passionate about Innovation, Application

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- 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

Gross monthly salary (before taxes) : 2 653 €

General Information

- **Theme/Domain :** Optimization, machine learning and statistical methods
  Statistics (Big data) (BAP E)
- **Town/city :** Villeneuve d'Ascq
- **Inria Center :** Centre Inria de l'Université de Lille
- **Starting date :** 2024-06-01
- **Duration of contract :** 1 year, 6 months
- **Deadline to apply :** 2024-03-31

Contacts

- **Inria Team :** MODAL
- **Recruiter :**
  Dabo Sophie / Sophie.Dabo@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 candidate should hold a PhD in statistics, machine learning. We are looking for an highly motivated candidate with an outstanding potential and a strong background in statistics and a deep interest in applications and package development.

Experience in FDA (Functional Data Analysis) classification/clustering and applications and previous Python or R package development are highly recommended.

The candidates should be able to work effectively as part of a multidisciplinary team, and to develop and pursue independent ideas. The successful candidate is expected to conduct innovative research at the highest international level.

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

CV + application letter + recommendation letters + List of publications

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.