

Postdoctoral Position in Machine Learning

Hosting body

This position is based at the research unit SESSTIM (Health Economic and Social Sciences and Medical Information Processing) at the Timone Faculty of Medical and Paramedical Sciences, Marseille, France. SESSTIM works to produce excellent, multidisciplinary and interdisciplinary research in social sciences and public health that can lead to changes in the various fields of predictive, personalised, pre-emptive and participatory medicine. SESSTIM researchers develop, or are associated with, research projects that attempt to provide answers to current challenges facing society and its populations, and contribute to methodological developments and advances. In terms of pathologies of interest, our research focuses mainly on cancer and infectious and communicable diseases. Our questions focus on individual, population and contextual factors. Our work targets populations in France, or more broadly in the North, but also in the South, mainly in the Mediterranean basin and sub-Saharan Africa.

Main responsibilities

The candidate will work in the multidisciplinary "Quantitative Methods and Medical Information Processing ([QuanTIM](#))" team, comprising researchers in epidemiology and public health, statisticians, biostatisticians, computer scientists and data scientists. More specifically, he/she will be assigned to a project involving the application and development of Artificial Intelligence techniques to data from cancer registries. The aim of the work will be **to develop or adapt a machine learning methodology in order to estimate excess mortality in the case of insufficiently stratified general population life tables**.

Activities

As part of the MIRACLE project (*Méthodologie et Intelligence artificielle pour IA recherche épidémiologique en CancéroLogiE sur bases de données*), funded by the French *Ligue contre le Cancer*, the candidate will contribute, for the benefit of patients and to decision-making in public health, to the enhancement of cancer databases, particularly the population-based ones. In this context, a key indicator measured in the general population is net survival, which represents the survival that would be observed in a hypothetical world where the studied disease is the only cause of death. By considering mortality due to other causes, which is derived from general population life tables stratified by certain variables, it enables comparisons to be made between populations and trends. However, using insufficiently stratified general population life tables leads to biased estimates of excess mortality. Different approaches have been considered and different models have been proposed to estimate excess cancer mortality for variables not directly observed in general population life tables (1-3). However, the existing models are based on certain assumptions that may be considered too strong given the needs and epidemiological questions. The candidate will familiarize himself/herself with various existing approaches and models, and then investigate the contribution of machine learning-based approaches based. He/she will **develop or adapt a methodology based on machine learning** (k-means, random forests or others) to **estimate excess mortality in the case of insufficiently stratified general population life tables**. The methodology developed should be adaptable to the situation where the number of variables not directly observed in the general population life tables is not limited. The candidate will **assess the performances** of these different methods through simulation studies. He/she will place particular emphasis on the **interpretation of the methods**, focusing on the **epidemiological interpretability** of the results obtained. He/she will **implement** the whole in an R package, preferably, or in another language depending on what is most suitable for practical application. In collaboration with other project investigators, he/she will write the article(s) on this work for publication in international peer-reviewed journals whether methodological or applied.

Competencies

Knowledge

- Strong theoretical and applied knowledge of machine learning techniques;
- Knowledge and skills in survival analysis;
- Expertise in model interpretation (e.g. SHAP technique);
- Proficiency in R and/or Python programming languages.

Know-How

- Autonomy, excellent organisational skills and thoroughness;
- Ability to work in a dynamic environment and meet deadlines;
- Capacity to listen, analyse and summarise;
- Capacity to work with multidisciplinary teams;
- Being a source of proposals.

Language skills

- English: scientific level (reading, writing, speaking).

Diploma level and experience

- PhD / Postdoctoral position with a specialisation in biostatistics, data sciences, mathematics or applied statistics.

Environment

Professional environment – Place of work

This work is carried out within the research unit SESSTIM, Health Economic and Social Sciences and Medical Information Processing, at the [Timone Faculty of Medical and Paramedical Sciences in Marseille](#).

Contract

Start date: As soon as possible, depending on administrative recruitment deadlines.

Duration: 12 months, with the possibility of extension.

Remuneration: Postdoctoral level; Aix-Marseille University salary scale.

Application:

Send to roch.giorgi@univ-amu.fr and nathalie.graffeo@univ-amu.fr your application file consisting of:

- A covering letter explaining how the applicant feels he/she can contribute to the project;
- A curriculum vitae;
- The thesis defence report.
- A letter (or letters) of recommendation would be an advantage.
- **Reference of the offer (to be indicated systematically):** MIRACLE-MLLT-24

Application deadline: April, 30th 2024.

Interviews will be conducted by visioconference or face-to-face in Marseille.

References

1. Touraine C, et al. More accurate cancer-related excess mortality through correcting background mortality for extra variables. *Stat Methods Med Res.* 2020;29(1):122-36.
2. Mba RD, et al. Correcting inaccurate background mortality in excess hazard models through breakpoints. *BMC Med Res Methodol.* 2020;20(1):268.
3. Rubio FJ, et al. On models for the estimation of the excess mortality hazard in case of insufficiently stratified life tables. *Biostatistics.* 2019;22(1):51-67.