PhD position in data science

Spatio-temporal modeling and AI tools for dose-response analysis and prediction of radiation-induced brain damage and cognitive impairment in patients treated with radiotherapy for glioblastoma.

Keywords : statistical learning, brain imaging, radiotherapy, probabilistic modelling, Bayesian statistics

Employer : Institut de Radioprotection et de Sûreté Nucléaire (IRSN)

Location : IRSN, Fontenay-Aux-Roses (92) – Ile de France – France

Benefits & working conditions : The position corresponds to a 3-year 100% PhD appointment with competitive salary and working conditions. It is co-funded in the framework of the ANR RADIO-AIDE (2022-2026). The PhD candidate will be physically based at the laboratory of epidemiology (LEPID at IRSN, Fontenay-Aux-Roses) with regular trips to INRIA Grenoble (Statify team).

The doctoral project will start at : October 2022

Applications should be sent to : Sophie Ancelet - sophie.ancelet@irsn.fr
Florence Forbes - florence.forbes@inria.fr

REQUIREMENTS AND ELIGIBILITY CRITERIA

Applicants must have :

- validated a MSc's degree (or an equivalent qualification in terms of credits)
- a background in statistics/machine learning/data science
- proficiency in modern high-level programming languages for data analysis (Python, R), probabilistic modelling and Bayesian statistics
- an interest in applications in epidemiology/public health.

Good writing and proficiency in oral and written English are also required.

The successful candidate is expected to be willing to work collaboratively in an interdisciplinary team of mathematicians, data scientists, neuroimaging experts, clinicians, and epidemiologists. He or she is also expected to be independent, highly motivated and able to communicate in scientific terms.

Age limit: 26 years old (unless exempted)

About the position

Background :

Radiotherapy (RT) is one of the most important treatments of brain tumours. However, its potential toxicity on the central nervous system is a highly relevant clinical issue [Soussain et al., 2009]. It is also part of the priority research questions in radiation protection, regarding the identification and the prevention of non-cancer side effects related to the use of ionizing radiation (IR) for therapeutic purposes [UNSCEAR, 2010]. Currently, the most frequent and threatening mid to long-term neurotoxic complication of brain RT is cognitive dysfunction mainly related to radiation-induced leukoencephalopathy (RIL). The main modifications observed in Magnetic Reasonance (MR) brain images following RT are diffuse supratentorial white matter hyperintensities (WMH), radionecrosis and brain atrophy (BA), resulting in ventricular dilatation. The symptoms related to cognitive impairments induced by brain RT are quite various (ex: dysfunctions in working memory, psychomotor speed, executive function and attention) and can significantly alter the quality of life of long-term survivors. Currently, the potential association between the spatio-temporal (ST) progression of radiation-induced brain injuries (as observed in MR brain images following RT) and the occurrence of cognitive impairment is poorly understood. This is partly due to complex interactions between tumour-related factors (or possible tumour recurrence) and treatment-induced neurotoxicity [Bompaire et al., 2018]. Most of previous studies did not consider jointly these two dimensions, despite their high level of correlation.

This thesis is part of the ANR research project RADIO-AIDE (*RADIation-induced neurOtoxicity assessed by ST modeling and Artificial Intelligence after brain raDiothErapy*) which will start in spring 2022 for 4 years (2022-2026). Led by IRSN (Coordination : Sophie Ancelet, thesis supervisor), this data science project brings together 7 partners (INRIA Statify team led by Florence Forbes (thesis director), INSERM-GIN, Centre Giovanni Borelli, ICM, AP-HP Pitié-Salpêtrière, Institut de Cancérologie Strasbourg Europe (ICANS, Strasbourg) and the start-up company Pixyl). RADIO-AIDE aims at developing spatio-temporal (ST) models and Artificial Intelligence (AI) tools to better understand the neurotoxic mechanisms underlying the initial localization and ST evolution of brain tissue modifications induced by RT in patients treated with RT for high-grade glioma (glioblastoma thereafter). The developed tools will allow to study the potential links between these MRI lesions and the occurrence of cognitive disorders, to individually predict the cognitive side-effects at early stage after brain RT and to provide to clinicians a usable academic software to perform an automatic processing of patient's MR brain images following brain RT, based on relevant image-based biomarkers of both RT-induced neurotoxicity and potential tumor progression.

Content :

This thesis aims at developing probabilistic ST models and AI tools to: a) estimate the potential associations between radiation-induced brain tissue modifications (as observed in MR brain images following RT), the dose of IR absorbed to the brain at a fine spatial scale (e.g. voxels of MR brain images) and the occurrence of cognitive impairment in patients treated with RT for glioblastoma ; b) predict the individual occurrence of cognitive impairment in these patients at an early stage after RT, with a quantification of the associated prediction uncertainties The work will be based mainly on data from the EpiBrainRad cohort [Durand et al. (2015)], which is one of the largest international prospective cohorts of patients treated with chemo-radiotherapy for glioblastoma and who have undergone standardised neuropsychological assessments. This cohort currently includes 224 adults treated for glioblastoma at the Hôpital de la Pitié-Salpêtrière (Paris) or at ICANS (Strasbourg). Since 2015, data are collected prospectively during the clinical follow-up of patients. For each patient, they include: a) clinical characteristics; b) multimodal MR brain images (collected before surgery, before RT and then every 2-3 months after RT), c) CT images used in the planning of brain RT treatment; d) results of an extensive battery of cognitive tests performed before RT and at 12, 24 and 36 months after RT.

The first task will consist in proposing probabilistic ST models and AI tools to estimate the potential associations between the initial localization and the ST progression of radiation-induced brain lesions, the dose of IR absorbed to the brain and the occurrence of cognitive impairment in patients treated with RT for glioblastoma. More precisely, the aim will be to develop, implement - on EpiBrainRad data and simulated data sets - validate and highlight the contributions of ST models and Bayesian statistical learning algorithms, potentially enriched with recent techniques from neural networks and/or AI, to estimate different dose-response relationships of interest. Furthermore, a comparison with previously segmented brain structures potentially involved in cognitive functions will allow to strengthen or generate new hypotheses concerning the radiosensitivity of specific anatomical and functional brain structures and their potential link with the occurrence of cognitive impairment. The proposed models will describe the spatial distribution of the absorbed dose to the brain at a finer spatial scale (i.e. voxel scale) compared to previous analyses, with the aim of improving, if possible, the accuracy of dose-response analyses. The proposed models will have to take into account the fact that two neighbouring voxels in a brain MR image have a higher probability of belonging to the same lesion range than two distant voxels. These spatial dependencies could be modelled with standard Markov Random Fields - such as Ising models or conditional autoregressive models [Besag, York and Mollie, 1991]. Functional regression models [Besse et al., 2005] could also be considered to describe the temporal dynamics of the progression of the brain tissue modifications between successive follow-up examinations of a same patient. In order to estimate the potential association between a set of segmentation maps and a dosimetric map, a first original methodological approach could consist in proposing different extensions to the GLLiM models - for "Gaussian Locally Linear Mapping" - based on a mixture of Gaussian distributions [Deleforge et al., 2015] in order to take into account the spatial and temporal dependencies of the described random systems, as well as a set of ancillary covariates of interest. The proposed model will then have to be extended or an alternative one will have to be proposed in order to describe the observed cognitive scores - for example by considering the difference in scores observed between two neuropsychological examinations - and to highlight, if it exists, a potential association with the segmented white matter hyperintensities (location, volume, etc.) and the absorbed IR dose at voxel scale. It will be essential to account for any information related to the original brain tumour and to any tumour recurrence, as well as to compare it with the brain structures potentially involved in cognitive functions.

Predicting the occurrence of cognitive impairment from brain MR images is a current methodological challenge. In the WP4 of the RADIO-AIDE project, the strengths and weaknesses of different AI tools (e.g. AlexNet CCN, Random Forest, manifold learning) for the individual prediction of cognitive scores following brain RT will be compared. One of the known limitations of these approaches is that they do not properly account for the uncertainties associated with the provided predictions. This is all the more unfortunate as the sample size of EpiBrainRad patients who underwent neuropsychological assessment at least 12 months after RT is relatively small (78 patients). The second part of the thesis will thus aim at accounting for these prediction uncertainties. Several approaches could be compared. One of them could be to implement the ST models previously defined in a Bayesian predictive framework in order to predict the temporal evolution of cognitive scores of interest. Another approach could be to implement recent AI uncertainty quantification techniques and/or to explore a Bayesian Deep Learning approach [Kendall A. and Gal Y., 2017]. The aim will be to highlight, if they exist, the potential benefits of using a Bayesian statistical learning approach in this context.

Interests :

The results and AI tools developed in this thesis will provide new knowledge to all the clinical fields interested in the potential neurotoxicity of brain radiotherapy (i.e.,neuro-oncology, neuroradiology, radiation oncology, and neuropsychology) and to radiation protection expertise. They will potentially have a strong impact for the clinical support of patients with brain tumor. Given the available data, they will allow to quantify the strength of the potential associations between radiation-induced brain tissue modifications (as observed in MR brain images following RT), the dose of IR absorbed to the brain at a fine spatial scale (e.g. voxels of MR brain images) and the occurrence of cognitive impairment in patients treated with RT for glioblastoma. The thesis will also provide new insights into brain structures that play a central role in cognition but that are particularly vulnerable to irradiation. Finally, the methodological approaches developed in the thesis will contribute to different scientific fields (e.g., statistics, probabilistic modelling, AI) related to joint modelling, prediction and analysis of heterogeneous, longitudinal and spatially structured data associated with different follow-up times for each patient.

References :

[1] Besag J., York J. and Mollie A. (1991) Ann Inst Stat Math; doi: 10.1007." BF00116466.
[2] Besse P.C. et al. (2005) Appl. Stoch. Models Bus. Ind; doi: 10.1002/asmb.539
[3] Bompaire F., Lahutte M. et al (2018) Supportive Care in Cancer; doi: 10.1007/s00520-018-4296-9
[4] Deleforge A., Forbes F. and Horaud R. (2015) Statistics and computing, doi : 10.1007/s11222-014-9461-5
[5] Durand T. et al. (2015) BMC Neurol; doi: 10.1186/s12883-015-0519-6.
[6] Kendall A. and Gal Y. (2017) NIPS 2017, arXiv:1703.04977
[7] Roca P., et al. (2020) Diagnostic and Interventional Imaging. doi.org/10.1016/j.diii.2020.05.009
[8] Soussain C., Ricard D. et al. (2009) The Lancet; doi: 10.1016/S0140-6736(09)61299-X
[9] United Nations Scientific Committee on the Effects of Atomic Radiation (2010) UNSCEAR 2010 Report.