

Master 2 Internship + PhD thesis

Statistical assessment of learning procedures in brain imaging

Keywords: deep learning, medical imaging, neuroimaging, statistics, validation

The topic

Machine learning has become the central tool for many different tasks in neuroimaging ranging from computer aided diagnosis and prognosis (classification/regression), segmentation of brain structures or image synthesis for instance. When publishing about a machine learning method, it is essential not only to assess its performance in an unbiased manner but also to report on how precise the performance estimates are. In other words, it is necessary to provide a comprehensive statistical assessment of the performance. Unfortunately, this is currently rarely done in the literature. There is thus a pressing need to raise awareness of the community on such problems and to provide clear guidelines on how this should be performed. This issue falls under the more general problem of reproducibility in ML for medical imaging and one can refer to (Colliot et al, 2024) to get the bigger picture.

There are two cases that should be distinguished: evaluation of a **trained model** and evaluation of a **learning procedure**. Please refer to (Varoquaux and Colliot, 2023) for more details about these two cases and for a discussion of related statistical issues. Both cases are important. The first case allows to statistically assess a model before putting it into production (for example for subsequent neuroscience studies or as a step towards future clinical use). In such a case, the statistical procedure is quite well established: the only source of variance comes from the testing set and one performs bootstrapping on the testing set. The second case aims at assessing a learning procedure. It is useful for instance to compare models and to be able to tell if one approach is significantly better than another one or if the difference is due to chance. How to properly perform such statistical assessment is still a matter of active research. Indeed, one needs to account for multiple sources of variance, not only coming from the testing set but also from the training set, from hyperparameter choices, from random seeds... How to do this remains an open problem even though some interesting directions have been proposed (Bouthillier et al, 2021).

Therefore, this project aims at studying statistical methods for assessing learning procedures for neuroimaging data analysis. It will include a back and forth between experimental aspects and theoretical aspects (new experimental results potentially leading to modification of statistical procedures which would in turn lead to new experiments).

To that purpose, the experiments will rely on ClinicaDL (<https://clinicadl.readthedocs.io/>), a software platform for reproducible deep learning in neuroimaging.

More specifically, the project will include the following aspects:

- Theoretical analysis of statistical approaches for assessing learning procedures
- Overview of existing practices and litterature
- Performing experiments on standard benchmark data

- Performing various experiments with ClinicaDL across different neuroimaging tasks (classification, regression, segmentation...) and various datasets (starting with datasets on Alzheimer's disease and then moving to include datasets with other pathologies in order to: i) measure the different sources of variances of learning procedures and see if the conclusions made for general ML benchmarks hold in our context; ii) assessing if the existing proposed statistical approaches are adapted across our tasks of interest (to that purpose, we will compare them to more comprehensive but computationally heavy procedures)
- If needed, enrich ClinicaDL with new models and datasets if the ones currently available are not sufficient
- Based on experimental results, propose revised statistical procedures that are adapted depending on the context (task, number of samples, type of network, dataset...)
- Implement the statistical procedures in ClinicaDL

This is an ambitious project. Only a fraction of the aforementioned work can be performed within the timeframe of an internship. This is why we propose to potentially continue the internship by a PhD.

- O. Colliot, E. Thibeau-Sutre, Brianceau C, and N. Burgos, "Reproducibility in medical image computing: what is it and how is it assessed?," 2024. <https://openreview.net/forum?id=3fIXW9mFfn>
- X. Bouthillier, P. Delaunay, M. Bronzi, A. Trofimov, B. Nichyporuk, J. Szeto, N. Mohammadi Sepahvand, E. Raff, K. Madan, V. Voleti et al., "Accounting for variance in machine learning benchmarks," Proceedings of Machine Learning and Systems, vol. 3, pp. 747–769, 2021. <https://arxiv.org/abs/2103.03098>
- G. Varoquaux and O. Colliot, "Evaluating machine learning models and their diagnostic value," To appear in Machine Learning for Brain Disorders, Springer, HAL preprint, vol. hal-03682454, 2023 <https://hal.archives-ouvertes.fr/hal-03682454>

A vibrant scientific, technological, clinical and ethical environment

You will work within the ARAMIS lab (www.aramislab.fr) at the Paris Brain Institute (<http://www.icm-institute.org>), one of the world top research institutes for neurosciences. The institute is ideally located at the heart of the Pitié-Salpêtrière hospital, downtown Paris.

The ARAMIS lab, which is also part of Inria (the French National Institute for Research in Computer Science and Applied Mathematics), is dedicated to the development of new machine learning and statistical approaches for the analysis of large neuroimaging and clinical data sets.

To perform large scale experiments, you will have access to the Jean Zay supercomputing infrastructure which comprises about 2,000 V100 GPUs and about 400 latest generation A100 GPUs.

The internship/thesis will be directed by Olivier Colliot (Research Director). The intern will be interacting frequently with engineers and PhD students working on the ClinicaDL and Clinica software platforms, in particular Camille Brianceau (engineer) and Ravi Hassanaly (PhD student). The project is also part of a large scale international collaboration on this topic involving in particular the German Cancer Research Center at Tübingen and the Soda Team at Inria Saclay.

Your profile

- Master or engineering degree with a specialization in machine learning and/or statistics
- Excellent background in statistics and machine learning
- Excellent programming skills in Python
- Good writing skills
- Good relational and communication skills
- Ability to collaborate efficiently with other team members

Starting date

- Ideally, we would like to recruit a Master's intern starting no later than April 1st 2024. However, other starting dates can be considered. Moreover, we will also consider applications for a PhD thesis without internship thus do not hesitate to apply even if you already have an internship in another lab.

Contact

Olivier Colliot - <http://www.aramislab.fr/perso/colliot/> - olivier.colliot@cnrs.fr