ClinicaDL: an open-source deep learning software for reproducible neuroimaging processing
Elina Thibeau-Sutre1,2, Mauricio Díaz1,2, Ravi Hassanaly1, Alexandre Routier1, Didier Dormont1, Olivier Colliot1, Ninon Burgos1

1Joint co-authorship
1 Sorbonne Université, Institut du Cerveau - Paris Brain Institute, Inserm, CNRS, AP-HP, Hôpital Pitié Salpêtrière, Inria, Aramis project-team, Paris, France
2Inria, Service d’Expérimentation et de Développement, Paris, France

Introduction: In addition to facing a reproducibility crisis1, deep learning studies often include methodological flaws, especially in the field of neuroimaging2. This situation leads to the production of biased and overestimated results. To overcome this issue, we developed ClinicaDL: an end-to-end deep learning framework for deep learning users working on neuroimaging data that aims to prevent common pitfalls.

Methods: ClinicaDL allows its users to work with a great diversity of neuroimaging data sets as it interacts with a neuroimaging standard, BIDS. The (open-source) code is available on Github and the software is distributed as a python package uploaded to the PyPI index. ClinicaDL ensures its usability by proposing a well-maintained documentation and tutorials that can be run online to help newcomers to familiarize with the software. Continuous integration runs a series of tests on the main functionalities of the code and it prevents the introduction of new bugs and the detection of potential failures.

Results: ClinicaDL and its companion project Clinica6 allows performing an end-to-end neuroimaging analysis. While Clinica converts raw neuroimaging data sets to BIDS and stores preprocessed imaging data including structural MRI in a folder hierarchy called CAPS (Clinica Processed Structure), ClinicaDL includes a set of tools to prepare data for deep learning tasks (such as quality check, label definition, generation of synthetic data), architecture search, network training, as well as result inference, model evaluation and interpretation (Fig. 1).

Discussion: ClinicaDL implements a set of technical solutions to avoid the main methodological issues found in the literature, as well as saving all necessary information and parameters to guarantee the reproducibility of deep learning experiments. Some issues remain open for further contributions, such as exploiting optimally multi-gpu workspaces. This way, ClinicaDL helps deep learning researchers to produce reliable studies in neuroimaging.

Conclusion: ClinicaDL has been developed to answer three issues we identified in our field: (1) the use of neuroimaging data sets, (2) data leakage in validation and evaluation methods and (3) insufficient reproducibility.

References:

I hereby consent to the publication of this abstract under the CC-BY license