ClinicaDL: an open-source deep learning software for reproducible neuroimaging processing

Elina Thibeau-Sutre[†,1], Mauricio Díaz[†,1,2], Ravi Hassanaly[1], Alexandre Routier[1], Didier Dormont[1], Olivier Colliot[1], Ninon Burgos[1]

[†]Joint co-autorship

[1] Sorbonne Université, Institut du Cerveau - Paris Brain Institute, Inserm, CNRS, AP-HP, Hôpital Pitié Salpêtrière, Inria, Aramis project-team, Paris, France

[2]Inria, Service d'Expérimentation et de Développement, Paris, France

Introduction: In addition to facing a reproducibility crisis[1], deep learning studies often include methodological flaws, especially in the field of neuroimaging[2]. 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 Clinica[6] 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).

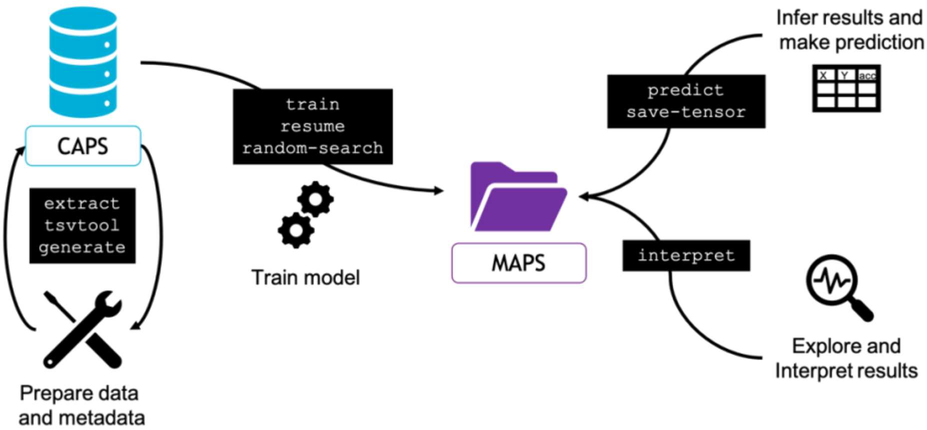

Figure 1. ClinicaDL main functionalities. `extract`, `tsvtool` and `generate` functionalities read and write in the ClinicA Processed Structure (CAPS) which contains neuroimaging data preprocessed by Clinica pipelines. ClinicaDL writes its own output, the Model Analysis and Processing Structure (MAPS), which contains the results of the training phase as well as inference on new data or the results of interpretability methods.

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:

1. Hutson, M., "Artificial intelligence faces reproducibility crisis," Science **359**(6377), (2018).
2. Wen, J., Thibeau-Sutre, E., et al. "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," Med. Image Anal. **63** (2020).
3. MONAI Consortium., [MONAI: Medical Open Network for AI], Zenodo (2020).
4. Pérez-García, et al. "TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning," Comput. Methods Programs Biomed. **208**, 106236 (2021).
5. Gorgolewski, K. J., et al., "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments," Sci. Data **3**, 160044 (2016).
6. Routier, A., et al., "Clinica: An Open-Source Software Platform for Reproducible Clinical Neuroscience Studies," Front. Neuroinformatics **15**, 689675 (2021).