

# Clinica: an open source software platform for reproducible clinical neuroscience studies

Alexandre Routier<sup>1,2</sup>, Jérémy Guillon<sup>1,3</sup>, Ninon Burgos<sup>1,3</sup>, Jorge Samper-Gonzalez<sup>1,3</sup>, Junhao Wen<sup>1,3</sup>, Sabrina Fontanella<sup>1,3</sup>, Simona Bottani<sup>1,3</sup>, Thomas Jacquemont<sup>1,3</sup>, Arnaud Marcoux<sup>1,3</sup>, Pietro Gori<sup>1,3</sup>, Pascal Lu<sup>1,3</sup>, Tristan Moreau<sup>1,3</sup>, Michael Bacci<sup>1,3</sup>, Stanley Durrleman<sup>1,3</sup>, Olivier Colliot<sup>1,4,5</sup>

<sup>1</sup> Inria Paris, Aramis project-team, Paris, France

<sup>2</sup> Sorbonne Université, Inserm, CNRS, Institut du Cerveau et la Moëlle épinière (ICM), FrontLab, Paris, France

<sup>3</sup> Sorbonne Université, Inserm, CNRS, Institut du Cerveau et la Moëlle épinière (ICM), Paris, France

<sup>4</sup> Sorbonne Université, Inserm, CNRS, Institut du Cerveau et la Moëlle épinière (ICM), AP-HP, Paris, France

<sup>5</sup> AP-HP, Departments of Neuroradiology and Neurology, Hôpital de la Pitié-Salpêtrière, Paris, France

## Introduction

Neuroimaging plays an important role in clinical neuroscience studies. Carrying out such studies involves many data analysis steps including image preprocessing, extraction of image-derived measurements and statistical analysis. A similar issue is faced by researchers in machine learning for neuroimaging: various steps are needed to extract features that are then fed to advanced learning algorithms. Often, these steps are performed by researchers through ad-hoc combinations of tools from widely used community software such as FreeSurfer or SPM. Such handicraftsman approach makes it difficult to: transmit tools and knowledge; reapply tools to new studies due to heterogeneous organization of inputs; merge results of several studies due to heterogeneous organization of outputs; reproduce results. Major progresses have been made with the Nipype pipelining system [Gorgolewski et al., 2011], the BIDS data organization standard [Gorgolewski et al., 2016] and the subsequent adoption of BIDS by various software resulting in BIDS Apps [Gorgolewski et al., 2017]. Nevertheless, carrying out a multimodal neuroimaging study remains challenging due to the need to combine tools from various software which each have their specific philosophy.

To address these issues, we present Clinica (<http://clinica.run/>), a software that aims at making clinical neuroscience easier and more reproducible.

## Methods

Clinica contains three main parts. 1) Feature extraction pipelines for different neuroimaging modalities (currently T1 MRI, diffusion MRI and PET). These pipelines allow to extract different types of features for each modality: voxel-based features (i.e. tissue density maps from T1, SUVR maps from PET, FA/MD/RD/AD maps from diffusion MRI), regional features (i.e. average of a given measurement within a region), anatomical meshes (surfaces of anatomical structures, white matter tracts), graphs (anatomical or functional connectivity networks). The pipelines are written in Python, based on the Nipype library and combine different softwares (currently SPM, FreeSurfer, MRtrix, FSL, PETPVC and ANTs). 2) Statistics and basic machine learning tools that take the different types of features as input. This part currently includes surface-based statistics based on SurfStat and basic classification methods from Scikit-learn. 3) I/O tools. This part currently includes tools for dataset management as well as tools to curate publicly available datasets (ADNI, AIBL and OASIS) and convert them into the BIDS standard.

Clinica uses the BIDS standard for inputs. The outputs of the different pipelines are stored under a specific BIDS-inspired structure called CAPS (ClinicA Processed Structure). In the future, Clinica could alternatively rely on BIDS-derivatives, to which we aim to contribute.

## Results

Clinica is distributed under the terms of the MIT License.

The target audience is mainly of two types:

- neuroscientists/clinicians conducting clinical neuroscience studies involving multimodal imaging, typically not experts in image processing for all of the involved imaging modalities. They will benefit from a unified set of tools covering the complete set of steps in a study (from raw data to statistical analysis).

- researchers developing advanced machine learning algorithms, typically not experts in brain image analysis. They will benefit from: standardized and fully automatic feature extraction; tools to convert public datasets into BIDS; baseline classification algorithms to which they could compare their results. Overall, we hope that Clinica will allow them to perform reproducible evaluations of their method.

We demonstrated that Clinica can be used in these two situations in [Bertrand et al., 2017] and [Samper et al., 2017], respectively.

## Conclusions

Clinica is an open-source software platform for clinical neuroscience. We hope that it will help researchers to: spend less time on data management and processing, easily share data and results, make their research more reproducible.

## References

Bertrand, A., Wen, J., Rinaldi, D., Houot, M., Sayah, S., Camuzat, A., Fournier, C., Fontanella, S., Routier, A., Couratier, P., Pasquier, F., Habert, M.-O., Hannequin, D., Martinaud, O., Caroppo, P., Levy, R., Dubois, B., Brice, A., Durrleman, S., Colliot, O., Le Ber, I., Predict to Prevent Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis (PREV-DEMALS) Study Group. (2017). Early Cognitive, Structural, and Microstructural Changes in Presymptomatic C9orf72 Carriers Younger Than 40 Years. *JAMA Neurol.*

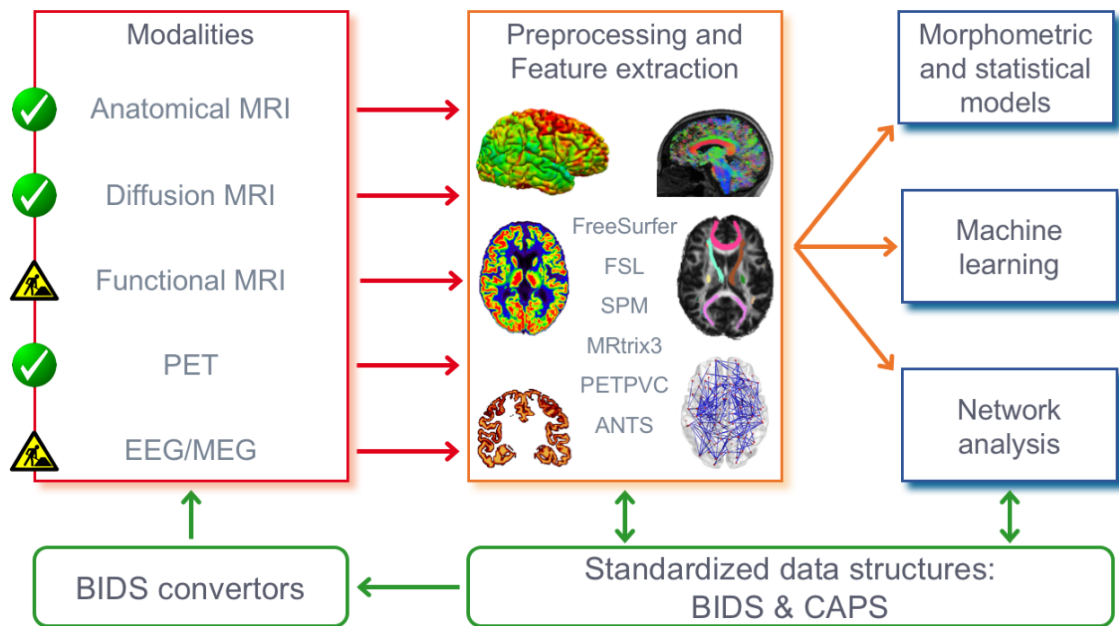
Samper-Gonzalez, J., Burgos, N., Fontanella, S., Bertin, H., Habert, M.-O., Durrleman, S., Evgeniou, T., Colliot, O. (2017). Yet another ADNI machine learning paper? Paving the way towards fully-reproducible research on classification of Alzheimers disease, in: *Machine Learning in Medical Imaging, Lecture Notes in Computer Science*. Presented at the International Workshop on Machine Learning in Medical Imaging, Springer, pp. 5360.

Gorgolewski, K.J., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L. and Ghosh, S.S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, 5.

Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O. and Handwerker, D.A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3, p.160044.

Gorgolewski, K.J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capot, M., Chakravarty, M.M., Churchill, N.W., Cohen, A.L., Craddock, R.C., Devenyi, G.A. and Eklund, A. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS computational biology*, 13(3), p.e1005209.

## Figures



**Figure 1** Clinica, an open source software platform for reproducible clinical neuroscience studies. Clinica contains three main parts: 1) Feature extraction pipelines for different neuroimaging modalities. 2) Statistics and basic machine learning tools. 3) I/O tools that include tools to curate publicly available datasets (ADNI, AIBL and OASIS) and convert them into the BIDS standard. Clinica uses the BIDS standard for inputs and outputs of the different pipelines are stored under a specific structure called CAPS (Clinica Processed Structure).