



**HAL**  
open science

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

Elina Thibeau-Sutre, Mauricio Diaz, Ravi Hassanaly, Alexandre M Routier,  
Didier Dormont, Olivier Colliot, Ninon Burgos

► **To cite this version:**

Elina Thibeau-Sutre, Mauricio Diaz, Ravi Hassanaly, Alexandre M Routier, Didier Dormont, et al..  
ClinicaDL: an open-source deep learning software for reproducible neuroimaging processing. 3IA  
Doctoral Workshop, Nov 2021, Toulouse, France. hal-03423072v2

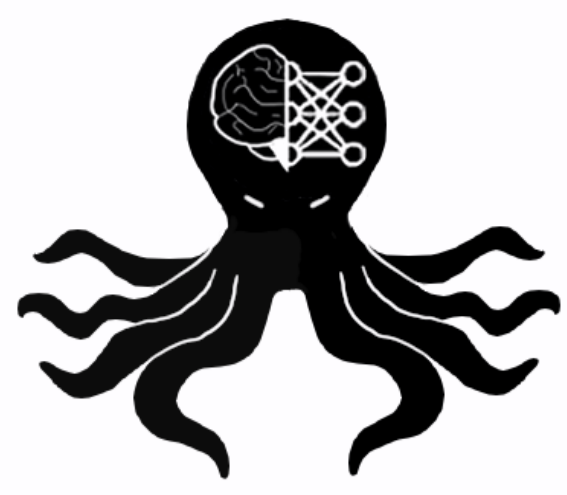
**HAL Id: hal-03423072**

**<https://hal.science/hal-03423072v2>**

Submitted on 24 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



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



Elina Thibeau-Sutre<sup>1</sup>, Mauricio Diaz<sup>1</sup>, Ravi Hassanaly<sup>1</sup>, Alexandre Routier<sup>1</sup>, Didier Dormont<sup>1,2</sup>, Olivier Colliot<sup>1</sup>, Ninon Burgos<sup>1</sup>

<sup>1</sup>Sorbonne Université, Institut du Cerveau, Inserm, CNRS, AP-HP Pitié-Salpêtrière, Inria Équipe-projet ARAMIS, Paris, France  
<sup>2</sup>AP-HP Pitié-Salpêtrière, Département de Neuroradiologie, Paris, France



elina.thibeausutre@icm-institute.org

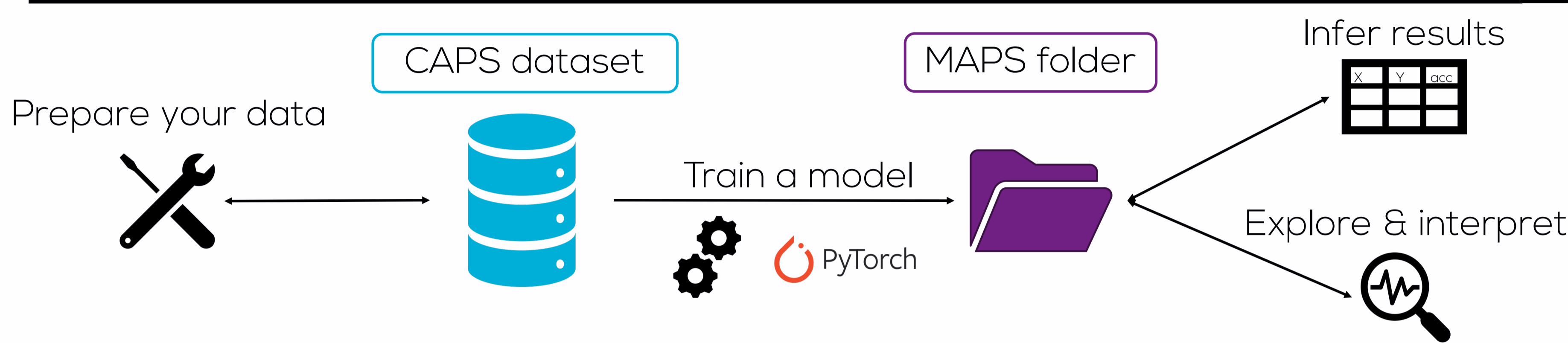
@AramisLabParis

Deep learning has become one of the most used data analysis technique for medical image analysis. Unfortunately, this recent massive use of deep learning has also been associated with **methodological flaws** in many studies which results are contaminated by data leakage.

Moreover, the whole deep learning community faces a **reproducibility crisis** that discredits its results. Hence there is an urgent need in publishing open-source software, data sets and scripts that allow reproducing the methodologies described in deep learning studies.

Finally, deep learning users who are not neuroimaging specialists have difficulty in accessing **properly formatted and pre-processed data sets**. This issue has been partly tackled by a data set format established by the community: the Brain Imaging Data Structure (BIDS).

## Software overview



## MAPS

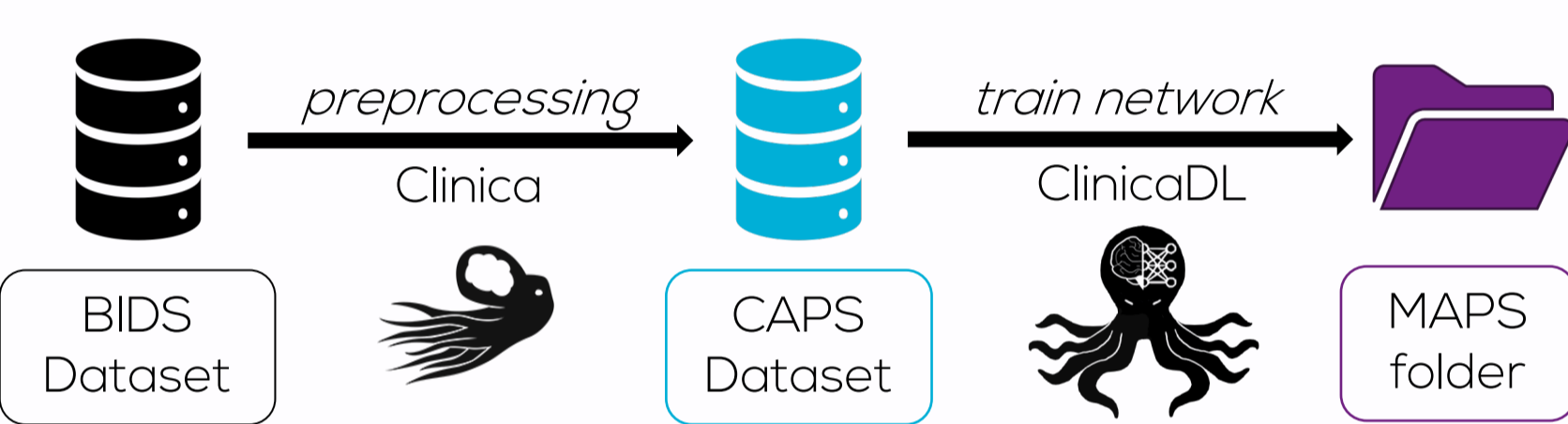
The Model Analysis and Processing Structure (MAPS) contains the results of all operations performed on a model.



## Motivations & Solutions

### Use of neuroimaging

The software ecosystem



**Problem**  
Diversity of formats  
- DICOM  
- ANALYZE  
- NIFTI

**Solution**  
Use of a standard  
**BIDS**

**Problem**  
Complex set of preprocessing tools  
STK, FSL, ...

**Solution**  
Unique library  
**Clinica**

**Problem**  
High dimensional data

**Solution**  
3D data decomposition  
3D patches, 2D slices, 3D ROI

### Reproducibility

**Problem**  
Achieve transparency

**Solution**  
Share usable code  
Continuous integration, Versioning, Maintain documentation

**Problem**  
Reproduce experiments

**Solution**  
Store hyperparameters in JSON file  
Fix random seed

### Rigorous validation

**Problem**  
Many possible scenarios of data leakage

1. Absence of an independent test set
2. Biased split
3. Late split
4. Biased transfer learning
5. Biased ensemble learning

**Example of solution**  
Check intersection with training data  
train+validation.tsv contains the list of all participants seen during training and eventual pretraining(s)

## Conclusion

ClinicaDL is an open-source software for deep learning processing on neuroimaging data. With this software, we solve the three main issues encountered by deep learning users who are not specialist of the neuroimaging domain:

- (1) the data management and preprocessing of neuroimaging data sets,
- (2) the contamination of results by data leakage,
- (3) the lack of reproducibility of deep learning experiments.

Moreover, thanks to abstract templates, a great flexibility is given to the users.

## Useful links

HAL, GitHub, Read the Docs

Preprint HAL, GitHub, Documentation

