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

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

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

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

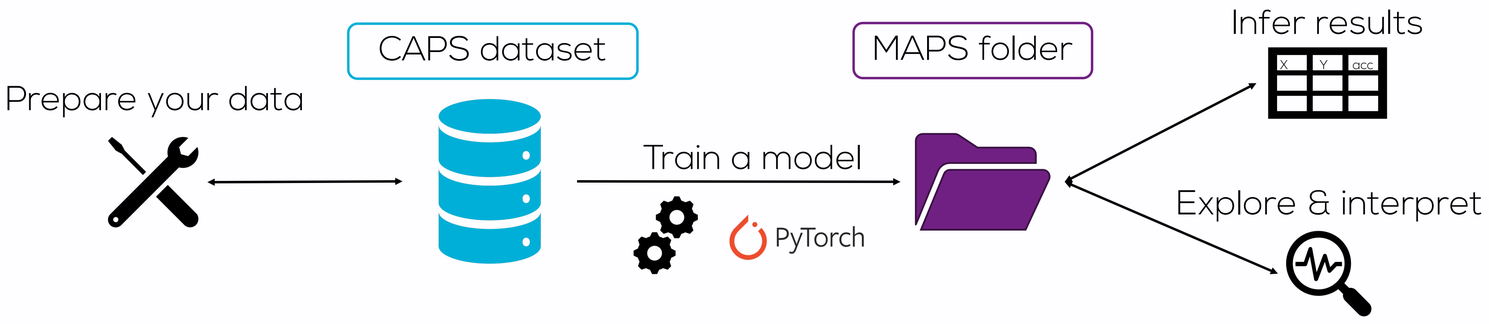
elina.thibeausutre@icm-institute.org @AramisLabParis

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 preprocessed 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

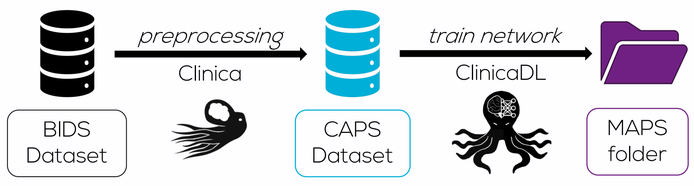
The Model Analysis and Processing Structure (MAPS) contains the results of all operations performed on a model.

MAPS
- Training logs
- Model weights
- Evaluation on test sets
- Experiment parameters
- Model interpretation

• 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

Solution

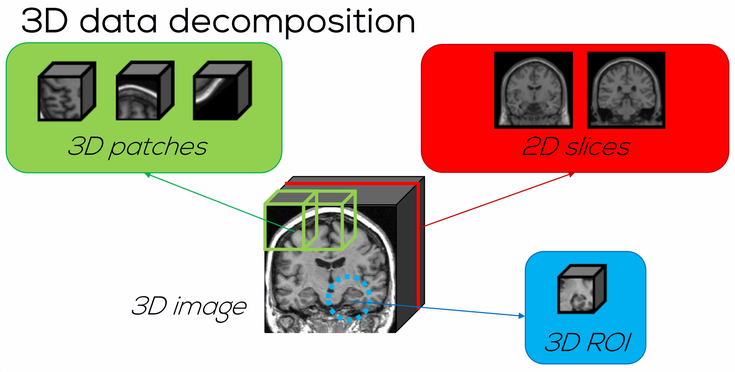
Unique library Clinica

Problem

High dimensional data

Solution

3D data decomposition



Reproducibility

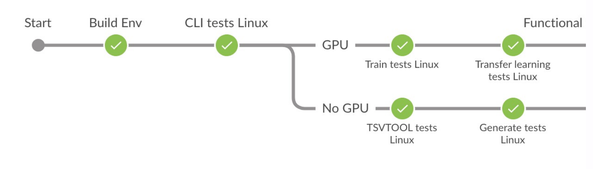
Problem

Achieve transparency

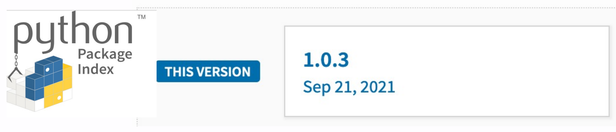
Solution

Share usable code

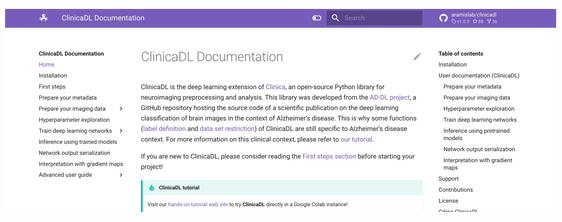
Continuous integration



Versioning



Maintain documentation



Problem

Reproduce experiments

Solution

Store hyperparameters in JSON file

```
preprocessing_dict:
    preprocessing:          "t1-linear"
    mode:                   "image"
    use_uncropped_image:    false
  file_type:
    pattern:                "*space-MNI152NLin2009cSym_desc-Crop_res-1x1x1_T1w.pt"
    description:            "T1W Image registered using t1-linear and cropped (matrix size 169×208×179, 1 mm isotropic voxels)"
    needed_pipeline:        "t1-linear"
  mode:                     "image"
  network_task:             "classification"
  caps_directory:           "/gpfswork/rech/bgc/commun/data/labels_list/session_3/ADNI_NIFD_caps_linear.tsv"
  tsv_path:                 "/gpfswork/rech/bgc/commun/data/labels_list/session_3/ADNI_NIFD_tsv_linear.tsv"
```

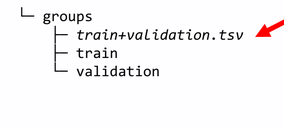
Fix random seed

Rigorous validation

Problem

Many possible scenarios of data leakage

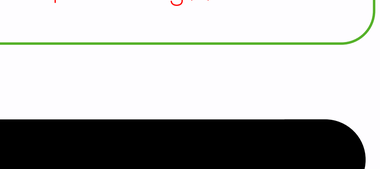
1. Absence of an independent test set



2. Biased split

3. Late split

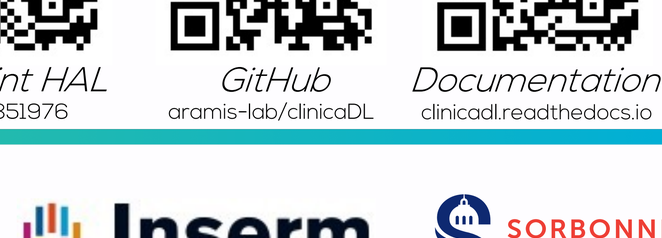
4. Biased transfer learning



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

Preprint HAL
hal-03351976

GitHub
aramis-lab/clinicaDL

Documentation
clinicadl.readthedocs.io