



Recent advances in the open-source ClinicaDL software for reproducible neuroimaging with deep learning

Ravi Hassanaly, Camille Brianceau, Mauricio Diaz, Sophie Loizillon, Elina Thibeau-Sutre, Nathan Cassereau, Olivier Colliot, Ninon Burgos

► To cite this version:

Ravi Hassanaly, Camille Brianceau, Mauricio Diaz, Sophie Loizillon, Elina Thibeau-Sutre, et al.. Recent advances in the open-source ClinicaDL software for reproducible neuroimaging with deep learning. SPIE Medical Imaging, Feb 2024, San Diego, United States. hal-04419141

HAL Id: hal-04419141

<https://hal.science/hal-04419141v1>

Submitted on 26 Jan 2024

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.



Distributed under a Creative Commons Attribution 4.0 International License

Recent advances in the open-source ClinicaDL software for reproducible neuroimaging with deep learning

Ravi Hassanaly^a, Camille Brianceau^a, Mauricio Diaz^b, Sophie Loizillon^a, Elina Thibeau-Sutre^d,
Nathan Cassereau^c, Olivier Colliot^a, and Ninon Burgos^a

^aSorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

^bInria, SED, Paris, France

^cCNRS, IDRIS, Orsay, France

^dDepartment of Applied Mathematics, Technical Medical Centre, University of Twente, Enschede, The Netherlands

ABSTRACT

In this paper, we present ClinicaDL, an open-source software platform that aims at enhancing the reproducibility and rigor of research for deep learning in neuroimaging. We first provide an overview of the software platform and then focus on recent advances. Features of the software aim at addressing three key issues in the field: the lack of reproducibility, the methodological flaws that plague many published studies and the difficulties using neuroimaging datasets for people with little expertise in this application area. Key existing functionalities include automatic data splitting, checking for data leakage, standards for data organization and results storing, continuous integration and integration with Clinica for preprocessing, amongst others. The most prominent recent features are as follows. We now provide various data augmentation and synthetic data generation functions (both standard and advanced ones including motion and hypometabolism simulation). Continuous integration test data are now versioned using DVC (data version control). Tools for generating validation splits have been made more generic. We made major improvements regarding usability and performance. We now support multi-GPU training and automatic mixed precision (to exploit tensor cores). We created a graphical interface to easily generate training specifications. We allow tracking of experiments through standard tools (MLflow, Weights&Biases). We believe that ClinicaDL can contribute to enhance the trustworthiness of research in deep learning for neuroimaging. Moreover, its functionalities and coding practices may serve as inspiration for the whole medical imaging community, beyond neuroimaging.

Keywords: Software Platform, Deep Learning, Neuroimaging, Reproducibility, Validation

1. INTRODUCTION

Research in deep learning for medical imaging has exploded in the past years, as shown by the tremendous increase in the number of publications.^{1,2} This has led to impressive results on various medical imaging tasks. However, at the same time, the field is also plagued with reproducibility and rigor issues. Reproducibility is the ability to reproduce published results and is the cornerstone of scientific progress.^{3,4} Reproducibility issues have been pointed in various fields,⁵⁻⁷ including deep learning for medical imaging.⁸ Lack of rigor is also often present with many examples of data leakage and invalid validation procedures in published papers.^{9,10}

Scientists typically develop code independently, which may not be of high quality, nor easily reusable nor sufficient for the reproducibility of papers. Moreover, deep learning researchers often lack specific knowledge in medical imaging. This may lead them to lose time with data management and preprocessing. Even worse, lack of domain knowledge may lead to flawed validation procedures and thus invalid published results, as shown in the case of neuroimaging.¹⁰ This is why there is a need to provide open-source tools to share efforts and give

Further author information:

Ravi Hassanaly: E-mail: ravi.hassanaly@icm-institute.org

Ninon Burgos: E-mail: ninon.burgos@cnrs.fr

access to this knowledge and methods to other researchers. ClinicaDL¹¹ is part of this larger initiative. It is an open-source software platform for reproducible deep learning studies in neuroimaging. In particular, it aims at helping towards addressing three issues: the lack of reproducibility of published articles, the methodological flaws that result in biased or overestimated results in the literature and finally the difficulty to handle neuroimaging datasets for researchers with little knowledge in this application area.

In this paper, we first provide an overview of ClinicaDL (Section 2) and then focus on novel features that we recently implemented (Section 3).

2. SOFTWARE OVERVIEW

ClinicaDL’s main features include: experiments and data preparation, model training, inference and prediction, and interpretation algorithms. The combination of ClinicaDL and Clinica^{12*} allows performing an end-to-end neuroimaging analysis, from the download of raw data sets to the interpretation of trained networks. An overview of the software platform is provided in Figure 1. More details on ClinicaDL functionalities can be found in the documentation[†]. All the deep learning part uses PyTorch¹³ as back-end. The software is distributed as a Python library with the Python index package. In this section, we focus on how ClinicaDL aims to address three recurrent issues of deep learning applied to neuroimaging: the lack of reproducibility, the methodological flaws and the difficulties using neuroimaging datasets.

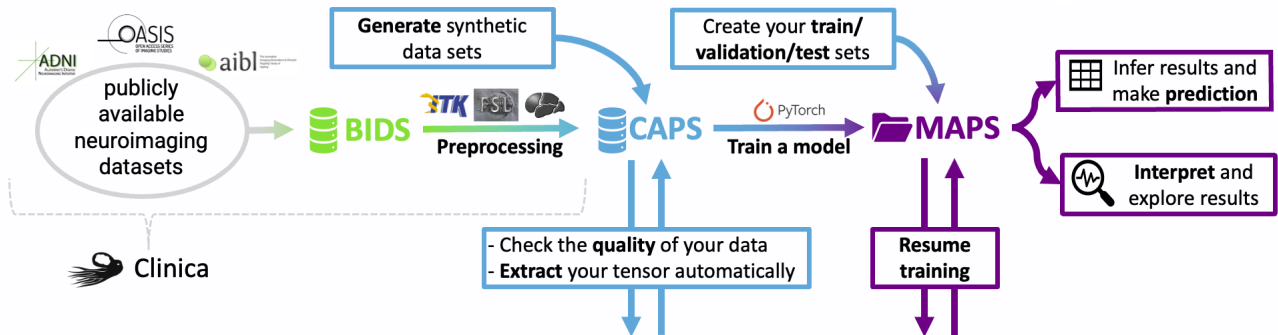


Figure 1. **Overview of the ClinicaDL software platform.** Data preprocessed by Clinica¹² are stored in a BIDS-like structure called CAPS. ClinicaDL further allows preparing data experiments and creating training, validation and test sets. Various models can then be trained and results are stored in a unified structure called MAPS. Finally, ClinicaDL allows running inference, saving predictions and interpreting the results.

Easy use of neuroimaging data. ClinicaDL has been developed as the deep learning extension of Clinica,¹² an open-source software platform that aims to unify neuroimage processing pipelines and data management. Clinica contains a set of automated pipelines to process different imaging modalities using standard software tools such as SPM,¹⁴ Freesurfer,¹⁵ or ANTs.¹⁶ It also includes converters to organize several open-source neuroimaging datasets such as ADNI¹⁷ or the UKBiobank¹⁸ under the brain imaging data structure¹⁹ (BIDS). A first step to make it easier to use neuroimaging datasets and to make experiments reproducible is to be part of the effort made by the BIDS community to standardize the organization of the datasets. The Clinica Processed Structure (CAPS) is part of this effort: this BIDS-like structure has the added benefit of considering all datasets as longitudinal and always using compressed NIFTI files. ClinicaDL can automatically read in a CAPS and load images that have been processed by Clinica for training and inference. ClinicaDL also allows extracting 2D slices, 3D patches or regions of interest as PyTorch tensors from 3D brain volumes to facilitate training.

*www.clinica.run

†<https://clinica.readthedocs.io/en/latest/>

Reproducibility. Using open-source software for deep learning experiments improves reproducibility by essence as the code is shared and usable. However, good coding practices can further enhance reproducibility. A first step is to have a robust continuous integration to ensure that the code does not break when adding new features. A second point is to automatize the installation of the software into a separate environment and keep control of the installed dependencies. Finally, ClinicaDL pushes reproducibility further by making it possible to relaunch an experiment with the same parameters since all of them are saved in the result directory.

Rigorous validation. ClinicaDL includes a set of pipelines and tests to avoid users making the most common methodological mistakes. For instance, the experiment preparation pipelines consist in making rigorous splits stratified by age and sex at the subject level. At inference, ClinicaDL automatically checks for various types of data leakage (e.g., wrong data split at the slice or visit level, absence of independent test set, late split, biased transfer learning, biased ensemble learning¹¹).

3. NEW FEATURES

In this section, we present in more detail ClinicaDL’s new features, which have been designed and implemented since the last journal publication¹¹ and conference presentation.²⁰ These new features are summarized in Figure 2. Three of them concern the topics described in Section 2 (easy use of neuroimaging data, reproducibility and validation). We also added features related to usability (making the platform more user-friendly and adding deep learning features) and performance.

Easy use of neuroimaging. We added various functionalities for data augmentation and synthetic data generation. ClinicaDL now supports TorchIO²¹ data augmentation. Other generation pipelines have been implemented to generate motion artifacts²² and hypometabolic data.²³ These generation pipelines can be used for data augmentation but also to validate models on synthetic data.

Reproducibility. We made major improvements to the continuous integration. We added versioning of test data used for continuous integration with data version control (DVC)[‡]. We also added non-regression tests for some pipelines and unitary tests for some critical functionalities of the software. There is now the possibility to fix the seed to improve the reproducibility of the results. This will, for example, determine the initialization of the model and the data loading sequence. However, despite having control over some GPU seeds, certain hardware-related factors such as architecture, memory configuration, clock speed, and calculation variations may still be beyond control, impacting reproducibility across different GPUs.

Rigorous validation. Another development axis has been to generalize experiment preparation to any neuroimaging dataset. Indeed, ClinicaDL initially resulted from work on the reproducibility of Alzheimer’s disease classification¹⁰ and thus some of its features were not generic enough. We have enhanced tools for manipulating TSV files to make them more generic and to handle both cross-sectional and longitudinal studies. Quality control (QC) of both raw and processed data is important to mitigate sources of bias and short-cut learning. QC algorithms that were already included in ClinicaDL have been updated to their latest version²⁴ and we added a new pipeline to check the registration of positron emission tomography (PET) images with a template. We further added a new method to interpret pretrained models.

Usability and performance. Making the software user-friendly has always been a goal for the development team by keeping the documentation up to date, writing tutorials or providing options such as the use of configuration files to simplify the command line. To go further, many new features have been added. Moreover, ClinicaDL now supports tracking of experiments via MLflow[§] (an open-source platform) and Weights and Biases[¶]

[‡]<https://dvc.org>

[§]<https://mlflow.org>

[¶]<https://wandb.ai/site>

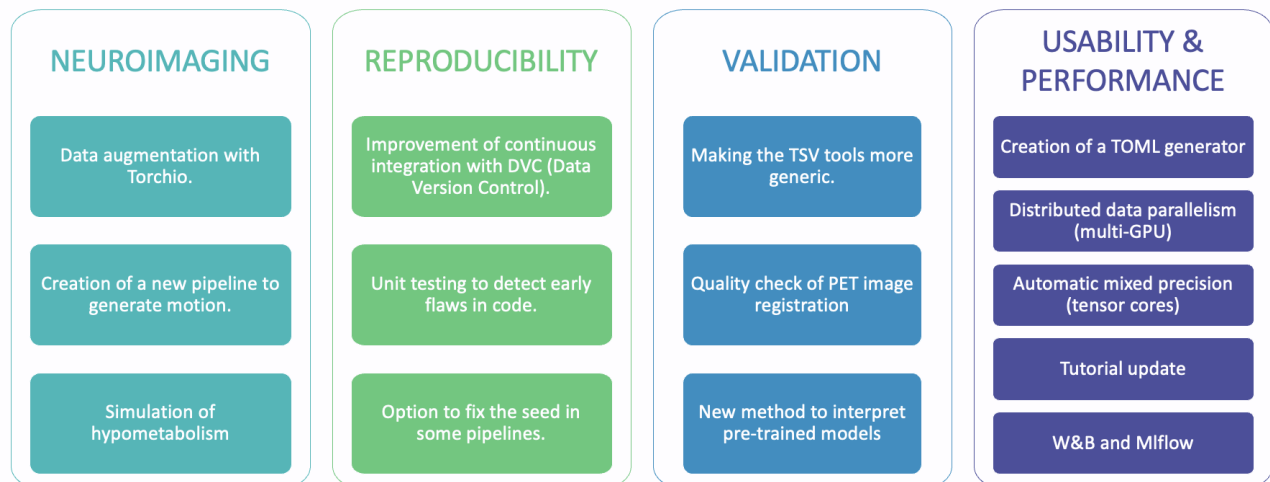


Figure 2. New features of the ClinicaDL software platform.

(a Python based platform),²⁵ which are widely used in the machine learning community. The command line could sometimes be tedious to use, especially when changing many parameters of the experiments. This is why we created a TOML generator[†], a web application that helps to configure experiments through a graphical user interface. Furthermore, new tutorials^{**} are available online: the aim is to show and explain how to use Clinica and ClinicaDL as well to give guidelines and spread good practices to the community

We also added new features to enhance the performance of model training with ClinicaDL. First, we integrated the PyTorch profiler that helps users to track GPU usage. Then, we performed developments to allow people to harness the power of high-performance computing (HPC) clusters (with multiple GPUs) and of state-of-the-art GPUs (Nvidia Tesla V100 and even A100), in particular those including tensor cores. We implemented multi-GPU training through distributed data parallelism. We added the use of automatic mixed precision for optimal use of GPU cards with tensor cores. This is even more crucial in medical imaging as the size of data can saturate the memory of the GPUs. Thus, it gives the possibility to use larger or more complex models, use full resolution high dimensional images or increase the batch size.

4. DISCUSSION

In this paper, we presented the ClinicaDL software platform. It can facilitate and improve the trustworthiness of research in deep learning for neuroimaging.

Good practices are essential in research to provide strong foundations to those whose work is built on the findings of others. Working with neuroimaging data can be complex, making it intricate to reproduce experiments. The same applies to the field of deep learning, the important number of parameters to choose from can make difficult usability, reproducibility or validation. In this way, a versioned and open-source software like ClinicaDL is a first step towards reproducibility. ClinicaDL is built in a way that makes it easy to get started. Researchers can use it “as is” or as a starting point to further develop tools for their research.

There are many well-known software platforms for deep learning in medical imaging, such as MONAI²⁶ and TorchIO.²¹ ClinicaDL is complementary to these platforms, in particular through its focus on: i) the specificities of neuroimaging; ii) providing an end-to-end framework for reproducible experiments and unbiased validation.

To conclude, ClinicaDL can contribute towards addressing issues of reproducibility and biased validation that plague the field. More generally, it can serve as inspiration for improving the implementation of deep learning research in other areas of medical imaging.

[†]<https://clinicadl-toml-generator.streamlit.app>

^{**}<https://aramislab.paris.inria.fr/clinicadl/tuto/2023/html/index.html>

ACKNOWLEDGMENTS

The research leading to these results has received funding from the French government under the management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6). This work was granted access to the HPC resources of IDRIS under the allocation 2019-100963 made by GENCI (Grand Équipement National de Calcul Intensif) in the context of the Jean Zay "Grands Challenges" (2019). Support has been received through the CNRS' PNRIA ("Programme national de recherche en intelligence artificielle").

REFERENCES

- [1] Wang, L., Wang, H., Huang, Y., Yan, B., Chang, Z., Liu, Z., Zhao, M., Cui, L., Song, J., and Li, F., "Trends in the application of deep learning networks in medical image analysis: Evolution between 2012 and 2020," *European journal of radiology* **146**, 110069 (2022).
- [2] Zeng, D., Noteboom, C., Sutrave, K., and Godasu, R., "A meta-analysis of evolution of deep learning research in medical image analysis," *Faculty Research and Publications* **273** (2020).
- [3] Colliot, O., Thibeau-Sutre, E., and Burgos, N., "Reproducibility in machine learning for medical imaging," in *[Machine Learning for Brain Disorders]*, Colliot, O., ed., 631–653, Springer US, New York, NY (2023).
- [4] Colliot, O., Thibeau-Sutre, E., Brianceau, C., and Burgos, N., "Reproducibility in medical image computing: What is it and how is it assessed?," (2024). <https://openreview.net/forum?id=3fIXW9mFfn>.
- [5] Hutson, M., "Artificial intelligence faces reproducibility crisis," *Science* **359** (2018).
- [6] Collaboration, O. S., "Estimating the reproducibility of psychological science," *Science* **349**(6251) (2015).
- [7] Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Waldron, L., Wang, B., McIntosh, C., Goldenberg, A., Kundaje, A., Greene, C. S., et al., "Transparency and reproducibility in artificial intelligence," *Nature* **586**(7829), E14–E16 (2020).
- [8] Simkó, A., Garpebring, A., Jonsson, J., Nyholm, T., and Löfstedt, T., "Reproducibility of the methods in medical imaging with deep learning," in *[Medical Imaging with Deep Learning]*, (2023).
- [9] Varoquaux, G. and Cheplygina, V., "Machine learning for medical imaging: methodological failures and recommendations for the future," *NPJ digital medicine* **5**(1), 48 (2022).
- [10] Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., and Colliot, O., "Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation," *Medical image analysis* **63**, 101694 (2020).
- [11] Thibeau-Sutre, E., Díaz, M., Hassanaly, R., Routier, A., Dormont, D., Colliot, O., and Burgos, N., "ClinicaDL: An open-source deep learning software for reproducible neuroimaging processing," *Computer Methods and Programs in Biomedicine* **220**, 106818 (2022).
- [12] Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., Hassanaly, R., Jacquemont, T., Lu, P., Marcoux, A., Moreau, T., Samper-González, J., Teichmann, M., Thibeau-Sutre, E., Vaillant, G., Wen, J., Wild, A., Habert, M.-O., Durrleman, S., and Colliot, O., "Clinica: An Open-Source Software Platform for Reproducible Clinical Neuroscience Studies," *Frontiers in Neuroinformatics* **15**, 39 (2021).
- [13] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems* **32** (2019).
- [14] Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E., *[Statistical parametric mapping: the analysis of functional brain images]*, Elsevier (2011).
- [15] Fischl, B., "FreeSurfer," *Neuroimage* **62**(2), 774–781 (2012).
- [16] Avants, B. B., Tustison, N., Song, G., et al., "Advanced normalization tools (ANTS)," *Insight Journal* **2**(365), 1–35 (2009).
- [17] Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W., and Beckett, L., "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics of North America* **15**(4), 869–877 (2005).

- [18] Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., Miller, K. L., and Smith, S. M., “Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank,” *NeuroImage* **166**, 400–424 (2018).
- [19] 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., et al., “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments,” *Scientific data* **3**(1), 1–9 (2016).
- [20] Thibeau-Sutre, E., Díaz, M., Hassanaly, R., Colliot, O., and Burgos, N., “A glimpse of ClinicaDL, an open-source software for reproducible deep learning in neuroimaging,” in [*Medical Imaging with Deep Learning*], (2022).
- [21] Pérez-García, F., Sparks, R., and Ourselin, S., “TorchIO: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning,” *Computer Methods and Programs in Biomedicine* **208**, 106236 (2021).
- [22] Loizillon, S., Bottani, S., Maire, A., Ströer, S., Dormont, D., Colliot, O., and Burgos, N., “Transfer learning from synthetic to routine clinical data for motion artefact detection in brain T1-weighted MRI,” in [*SPIE Medical Imaging 2023: Image Processing*], (2023).
- [23] Hassanaly, R., Bottani, S., Sauty, B., Colliot, O., and Burgos, N., “Simulation based evaluation framework for deep learning unsupervised anomaly detection on brain FDG-PET,” in [*SPIE Medical Imaging 2023: Image Processing*], **12464**, 511–518, SPIE (2023).
- [24] Fonov, V., Dadar, M., and Collins, D., “DARQ: Deep learning of quality control for stereotaxic registration of human brain mri to the T1w MNI-ICBM 152 template,” *Neuroimage* **257**, 119266 (2022).
- [25] Biewald, L. et al., “Experiment tracking with weights and biases,” *Software available from wandb.com* **2**, 233 (2020).
- [26] Cardoso, M. J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al., “MONAI: An open-source framework for deep learning in healthcare,” *arXiv preprint arXiv:2211.02701* (2022).