



HAL
open science

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

Alexandre Routier, Ninon Burgos, Jérémy Guillon, Jorge Samper-González,
Junhao Wen, Simona Bottani, Arnaud Marcoux, Michael Bacci, Sabrina
Fontanella, Thomas Jacquemont, et al.

► To cite this version:

Alexandre Routier, Ninon Burgos, Jérémy Guillon, Jorge Samper-González, Junhao Wen, et al.. Clinica: an open source software platform for reproducible clinical neuroscience studies. 2019. hal-02308126v1

HAL Id: hal-02308126

<https://inria.hal.science/hal-02308126v1>

Preprint submitted on 8 Oct 2019 (v1), last revised 16 Aug 2021 (v4)

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.

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

Alexandre Routier^{1,2,3,4,5}, Ninon Burgos^{2,3,4,5,1}, Jérémy Guillon^{2,3,4,5,1}, Jorge Samper-González^{2,3,4,5,1}, Junhao Wen^{2,3,4,5,1}, Simona Bottani^{1,2,3,4,5}, Arnaud Marcoux^{2,3,4,5,1}, Michael Bacci^{1,2,3,4,5}, Sabrina Fontanella^{2,3,4,5,1}, Thomas Jacquemont^{2,3,4,5,1}, Adam Wild^{2,3,4,5,1}, Pietro Gori^{1,2,3,4,5}, Alexis Guyot^{2,3,4,5,1}, Pascal Lu^{2,3,4,5,1}, Mauricio Diaz⁷, Elina Thibeau--Sutre^{2,3,4,5,1}, Tristan Moreau^{2,3,4,5,7}, Marc Teichmann^{3,5,7,8}, Marie-Odile Habert^{9,10,11}, Stanley Durrleman^{1,2,3,4,5}, Olivier Colliot^{2,3,4,5,1,12}

¹ Inria, Aramis project-team, F-75013, Paris, France

² Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France

³ Inserm, U 1127, F-75013, Paris, France

⁴ CNRS, UMR 7225, F-75013, Paris, France

⁵ Sorbonne Université, F-75013, Paris, France

⁶ Inria Paris, SED, 75012 Paris, France

⁷ Institut du Cerveau et de la Moelle épinière, ICM, FrontLab, F-75013, Paris, France

⁸ Department of Neurology, Institute for Memory and Alzheimer's Disease, Pitié-Salpêtrière Hospital, AP-HP, Paris, France

⁹ Sorbonne Université, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale, LIB, F-75006, Paris, France

¹⁰ AP-HP, Hôpital Pitié-Salpêtrière, Médecine Nucléaire, F-75013, Paris, France

¹¹ Centre d'Acquisition et Traitement des Images (CATI), www.cati-neuroimaging.com

¹² AP-HP, Department of Neurology and Neuroradiology, Pitié-Salpêtrière Hospital, Paris, France

* Correspondence:

Olivier Colliot, PhD - olivier.colliot@upmc.fr

ICM – Brain and Spinal Cord Institute

ARAMIS team Pitié-Salpêtrière Hospital

47-83, boulevard de l'Hôpital, 75651 Paris Cedex 13, France

Keywords: Software, Pipeline, Data processing, Machine learning, Reproducibility, BIDS, Neuroimaging, Nipype.

30 **Abstract**

31 We present Clinica (www.clinica.run), an open source software platform designed to make clinical
32 neuroscience studies easier and more reproducible. Clinica aims for researchers to i) spend less time
33 on data management and processing, ii) perform reproducible evaluations of their methods, and iii)
34 easily share data and results within their institution and with external collaborators. Clinica relies on
35 the brain imaging data structure (BIDS) for the organization of raw neuroimaging datasets and on
36 established tools written by the community to build standardized pipelines for the analysis of
37 neuroimaging data. It also provides converters of public neuroimaging datasets to BIDS (currently
38 ADNI, AIBL, OASIS and NIFD), statistical analysis and machine learning algorithms. Currently,
39 Clinica handles T1-weighted MRI, diffusion MRI, functional MRI and PET data. Processed data
40 include image-valued scalar fields (e.g. tissue probability maps), meshes, surface-based scalar fields
41 (e.g. cortical thickness maps) or scalar outputs (e.g. regional averages). These data follow the Clinica
42 Processed Structure (CAPS) format which shares the same philosophy as BIDS. Standardized
43 organization of raw and processed neuroimaging files facilitates the execution of single pipelines and
44 of sequences of pipelines, as well as the integration of processed data into statistics or machine learning
45 frameworks. The target audience of Clinica is neuroscientists or clinicians conducting clinical
46 neuroscience studies involving multimodal imaging, and researchers developing advanced machine
47 learning algorithms.

48 **1 Introduction**

49 Neuroimaging plays an important role in clinical neuroscience studies. While the meaning of clinical
50 neuroscience studies may vary, we use it to refer to studies involving human participants (i.e. patients
51 with neurological and psychiatric diseases, and control subjects) explored with multimodal data
52 (neuroimaging, clinical and cognitive evaluations, genetic data...) and most often involving
53 longitudinal follow-up. Carrying out such studies involves many data analysis steps including image
54 pre-processing, extraction of image-derived measurements and statistical analysis, thus requiring a
55 wide range of expertise. A similar situation is faced by researchers in machine learning for
56 neuroimaging: various steps are needed to extract features that are then fed to advanced learning
57 algorithms.

58 The first issue met when working on clinical studies concerns the organization of neuroimaging
59 datasets within or between institutions. The lack of a standardized structure makes arduous the sharing
60 or reuse of data. This is true for in-house, but also for publicly available neuroimaging datasets.
61 Another consequence of the lack of standard is the difficulty to apply automatic pipelines (e.g.
62 extraction of neuroimaging features, statistical analysis or machine learning) and to perform quality
63 assurance. The second issue faced by researchers processing data from clinical studies is related to the
64 high number of software packages, such as FreeSurfer¹ (Fischl, 2012), FMRIB Software Library
65 (FSL)² (Jenkinson et al., 2012) or Statistical Parametric Mapping³ (SPM) (Friston et al., 2007), that
66 exist in the community. Researchers have to understand the methodology behind each tool (e.g.
67 segmentation, registration, etc.) and master them from a programming perspective before being able to
68 combine them and develop image processing pipelines. Moreover, such “handicraft” approach makes
69 it difficult to transmit tools and knowledge, and to merge and share results of several studies due to the
70 heterogeneous organization of outputs. Finally, the difficulty to access or share both raw and processed
71 neuroimaging data hinders the reproducibility of neuroimaging studies (Poline et al., 2012).

72 Major progress has been made in the last years to ease neuroimaging studies. First, difficulties
73 related to the heterogeneity of image processing tools have been partly handled by the Nipype
74 (Neuroimaging in Python – Pipelines and Interfaces) software package⁴ (Gorgolewski et al., 2011).
75 Nipype is an open-source Python project that provides a uniform environment facilitating interaction
76 between neuroimaging software tools or algorithms, regardless of their programming language, within
77 a single workflow. Later, the issues related to the organization of the clinical and imaging data have
78 been tackled by the Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016), a new standard
79 from the community for the community. The BIDS standard is based on a file hierarchy rather than on
80 a database management system, thus facilitating its deployment in any environment. Thanks to its clear
81 and simple way to describe neuroimaging and behavioral data, the BIDS standard has been easily
82 adopted by the neuroimaging community. Organizing a dataset following the BIDS hierarchy
83 simplifies the execution of neuroimaging software tools, resulting in the development of user-friendly
84 software. For instance, BIDS Apps (Gorgolewski et al., 2017) provides a set of pipelines for the
85 processing of neuroimaging data that follow a BIDS hierarchy. Currently, it mainly wraps
86 neuroimaging software packages from the community into a Docker image and is used via a command

¹ <http://surfer.nmr.mgh.harvard.edu/>

² <https://fsl.fmrib.ox.ac.uk/>

³ <http://www.fil.ion.ucl.ac.uk/spm/>

⁴ <https://nipype.readthedocs.io>

87 line interface. Moreover, the Nilearn⁵ (Abraham et al., 2014) package facilitates the application of
88 advanced machine learning approaches to neuroimaging data. To that purpose, it leverages the scikit-
89 learn library⁶ (Pedregosa et al., 2011) and provides tools for handling and visualizing different types
90 of neuroimaging data and building predictive models.

91 Nevertheless, carrying out a multimodal neuroimaging study remains challenging due to the
92 know-how necessary to grasp each modality and tool involved. While technical implementation has
93 been facilitated by Nipype, the development of a pipeline still requires substantial programming skills
94 and time to master both the neuroimaging software tools and Nipype. While the BIDS standard is being
95 adopted by the scientific community, not all public neuroimaging datasets provide a BIDS version of
96 their data, and tools to curate and convert these datasets into BIDS are lacking. Besides, performing a
97 single or multimodal neuroimaging study will also require methodological expertise. For instance, a
98 classification study of healthy subjects and patients with a neurodegenerative disease using ¹⁸F-
99 fluorodeoxyglucose positron emission tomography (FDG PET) could involve notions of multimodal
100 registration between FDG PET and T1-weighted magnetic resonance imaging (MRI), tissue
101 segmentation of T1-weighted (T1w) MRI, PET partial volume correction, normalization into a
102 standard space, and machine learning-based classification, as well as know-how of the tools used to
103 perform these steps. Moreover, the image processing steps need to be chained from one to the other
104 and the absence of data organization for processed neuroimaging data makes data analysis more
105 complex. Finally, the neuroimaging features generated by the pipelines need to be correctly connected
106 to statistical or machine learning frameworks.

107 Clinica (www.clinica.run) aims to make clinical research studies easier and pursues the
108 community effort of reproducibility. Clinica relies on tools written by the scientific community and
109 provides converters of public neuroimaging datasets to BIDS, standardized processing pipelines and
110 organization for processed files, statistical analysis, and machine learning algorithms.

111 The target audience is mainly of two types. First, neuroscientists or clinicians conducting
112 clinical neuroscience studies involving multimodal imaging, typically not experts in image processing
113 for all of the involved imaging modalities or in statistical analysis. They will benefit from a unified set
114 of tools covering the complete set of steps involved in a study (from raw data to statistical analysis).
115 Second, researchers developing advanced machine learning algorithms, typically not experts in brain
116 image analysis. They will benefit from tools to convert public datasets into BIDS, standardized and
117 fully automatic feature extraction methods, and baseline classification algorithms to which they could
118 compare their results. Overall, we hope that Clinica will allow users to spend less time on data
119 management and processing, to perform reproducible evaluations of their methods, and to easily share
120 data and results within their institution and with external collaborators.

⁵ <https://nilearn.github.io>

⁶ <https://scikit-learn.org/>

121 **2 Clinica overview**

122 Clinica is an open source software platform for reproducible clinical neuroimaging studies. It can take
123 as inputs different neuroimaging modalities, currently anatomical MRI, diffusion MRI, functional MRI
124 and PET. Clinica provides standardized processing pipelines that involve the combination of different
125 software packages. It is currently relying on FreeSurfer (Fischl, 2012), FSL (Jenkinson et al., 2012),
126 SPM (Frackowiak et al., 1997), Advanced Normalization Tools (ANTs)⁷ (Avants et al., 2014),
127 MRtrix3⁸ (Tournier et al., 2012), and PET Partial Volume Correction (PETPVC) toolbox⁹ (Thomas et
128 al., 2016). Features extracted with the different pipelines can be used as inputs to statistical analysis,
129 which relies on SurfStat¹⁰ (Worsley et al., 2009), or machine learning analysis, which relies on scikit-
130 learn (Pedregosa et al., 2011).

131 Input neuroimaging data are expected to follow the BIDS data structure (Gorgolewski et al.,
132 2016), as explained in section 3.2. Since this new standard has only recently been adopted by the
133 community, not all public neuroimaging datasets are yet proposed in BIDS format. To facilitate the
134 adoption of BIDS, Clinica curates several publicly available neuroimaging datasets and provides tools
135 to convert them into the BIDS format. Processed data are organized following the ClinicA Processed
136 Structure (CAPS) format, detailed in section 3.3, which shares the same philosophy as BIDS. Finally,
137 a set of tools is provided to handle input and output data generated by Clinica, thus facilitating data
138 management or connection to statistical or machine learning analysis.

139 The main functionalities of Clinica are described in the paper, but for further details the reader
140 can refer to the documentation available on the website¹¹. For each pipeline, the reader will find a
141 description of its functionalities, a list of the tools on which it relies, an example showing how to run
142 the pipeline, and a description of the outputs generated. The documentation of a pipeline can have
143 several levels of reading, which are respectively targeting people new or familiar with neuroimaging
144 and scientists working on pattern recognition and machine learning.

145

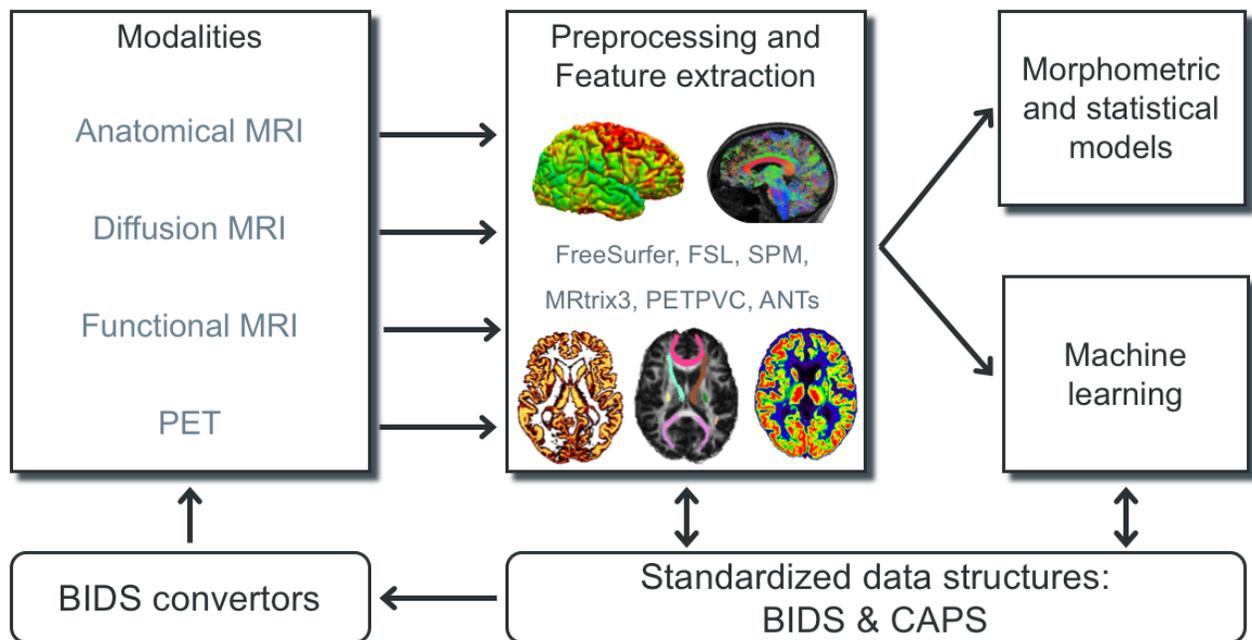
⁷ <https://stnava.github.io/ANTs/>

⁸ <http://mrtrix.org>

⁹ <https://github.com/UCL/PETPVC>

¹⁰ <http://www.math.mcgill.ca/keith/surfstat/>

¹¹ <http://www.clinica.run/doc>



146

147 **Figure 1.** Overview of Clinica's functionalities. Clinica provides standardized processing pipelines for
 148 MRI and PET images that involve the combination of different software packages, and whose outputs
 149 can be used for statistical or machine learning analysis. Clinica expects data to follow the Brain
 150 Imaging Data Structure (BIDS). Output data are stored using the ClinicA Processed Structure (CAPS).
 151 Finally, Clinica provides tools to convert public neuroimaging datasets into the BIDS format.

152 **3 Clinica environment**153 **3.1 Software architecture of Clinica**

154 The core of Clinica is written in Python and mainly relies on the neuroimaging Nipype framework
 155 (Gorgolewski et al., 2011) to create pipelines. Python dependencies also include NumPy (van der Walt
 156 et al., 2011), NiBabel (Brett et al., 2019), Pandas (McKinney, 2010), DIPY (Garyfallidis et al., 2014),
 157 NIPY, SciPi (Jones et al., 2001), scikit-learn (Pedregosa et al., 2011), scikit-image (van der Walt et al.,
 158 2011) and PyBIDS.

159 Clinica is provided to the end user in form of a Python package distributed through the classical
 160 channels: the Python Package Index (PyPI) or the Conda repository. However, we strongly recommend
 161 the use of virtual environments (Conda env or virtualenv) to facilitate the cohabitation of several
 162 versions of Python and associated libraries. For example, the installation of Clinica can simply be done
 163 by typing `conda install -c aramislab clinica` through the terminal, inside a virtual
 164 environment.

165 The main usage of Clinica is through the command line, which is facilitated by the support of
 166 autocompletion. The commands are gathered into four main categories. The first category of command
 167 line (`clinica run`) allows the user to run the different pipelines on neuroimaging datasets following
 168 a BIDS or CAPS hierarchy. The `clinica convert` category allows the conversion of publicly
 169 available neuroimaging datasets into a BIDS hierarchy. To help with data management, the `clinica`
 170 `iotools` category comprises a set of tools that allows the user to handle BIDS and CAPS datasets,
 171 including generating lists of subjects or merging all tabular data into a single TSV file for analysis with
 172 external statistical software packages. Finally, the last category (`clinica generate`) is dedicated to
 173 developers and currently generates the skeleton for a new pipeline. Examples of command line can be
 174 found in Table 1.

175 **Table 1.** Categories of command line

Category of command line	Description	Example usage
run	Run pipelines on BIDS or CAPS datasets. The full list of pipelines is available in section 4 .	<code>clinica run pet-surface bids_directory caps_directory --pet_tracer fdg</code>
convert	Convert public neuroimaging datasets into BIDS. The list of the public datasets can be found in section 6.1 .	<code>clinica convert adni-to-bids dataset_directory clinical_data_directory bids_directory</code>

iotools	Set of tools to handle BIDS and CAPS datasets. The list of the I/O tools can be found in section 6.2 .	<code>clinica iotools merge-tsv bids_directory --caps_directory caps_directory my_population.tsv</code>
generate	(For developers) Generate the skeleton source code for a new pipeline.	<code>clinica generate template "Modality Feature Extracted" pipeline_folder</code>

176 3.2 Input data with the BIDS standard

177 When dealing with multiple datasets, it is difficult to automate the execution of neuroimaging pipelines
 178 since their organization may vary from each other or even within each individual dataset. If we consider
 179 neuroimaging datasets involving many participants, the lack of a clear structure will necessitate a large
 180 amount of time to curate these databases and make them easily usable. Besides, large databases are
 181 often associated with database management systems, which involve additional technical and financial
 182 resources to be maintained.

183 BIDS (Gorgolewski et al., 2016) is a community standard enabling the storage of multiple
 184 neuroimaging modalities and behavioral data. The BIDS standard provides a unified structure and
 185 makes easier the development and distribution of code that uses neuroimaging datasets. Moreover, the
 186 BIDS format is based on a file hierarchy rather than on a database management system, thus avoiding
 187 the installation and maintenance of additional software. As a result, BIDS can be easily deployed in
 188 any environment. The specification is intentionally based on simple file formats and folder structures
 189 to reflect current laboratory practices, which makes it accessible to a wide range of scientists coming
 190 from different backgrounds. People unfamiliar with the BIDS format can see an example of a BIDS
 191 folder in Figure 3.

192 For these reasons, we also adopted this standard and Clinica expects that the input data are
 193 BIDS-compliant for the execution of pipelines. Note that if a cross-sectional dataset (i.e. with no
 194 `session` folder) is provided, Clinica will interactively propose to convert the cross-sectional dataset
 195 into a longitudinal dataset with a unique session.

196

197 3.3 Input/Output data with the CAPS structure

198 Clinica has its own specifications for storing processed data, called CAPS (Clinica Processed
 199 Structure). Processed data include: image-valued scalar fields (e.g. segmentation labels, tissue maps),
 200 meshes, mesh-valued scalar fields (e.g. cortical thickness maps), deformation fields, scalar outputs
 201 (e.g. volumes, regional averages), etc. Carrying out a neuroimaging study often involves the
 202 combination of different pipelines or the chaining of a pipeline to another one. This is the case for
 203 multimodal studies where processed outputs from a modality will be inputs for another pipeline, but it
 204 is also true for studies involving a single modality: features extracted from one or several pipelines are
 205 usually connected to statistical or machine learning frameworks. Finally, a standardized organization

206 for processed data will ease the access and sharing of data, thus improving the reproducibility of
207 neuroimaging studies.

208 The CAPS format defines a hierarchy for the Clinica processed data. The idea is to include in
209 a single folder all the results generated by the different pipelines and to organize the data following the
210 main patterns of the BIDS specification. CAPS folders are kept separate from the raw data. Indeed,
211 when processing data, it is very common to have the raw dataset located on a separated storage or read-
212 only storage, while ongoing processed data are located on a separate location or on a faster data storage.

213 Another notion we often meet in neuroimaging studies is the notion of group, e.g. template
214 creation from a set of subjects or statistical analysis of a population. To handle these situations, we
215 simply add a level to the CAPS folder hierarchy. While pipeline outputs for individuals are stored in
216 the `subjects` folder, results of group studies are stored in the `groups` folder together with the set of
217 participants involved. For instance, an AD group label could be used when a template is created for a
218 group of Alzheimer's disease patients. Any time this AD template is used, the `group_label` is
219 provided to identify the pipeline outputs obtained for this group. The `group_HCvsAD` could be used as
220 `group_label` for a statistical group comparison between healthy controls (HC) and Alzheimer's
221 disease patients. An illustration showing the chaining of pipelines and the creation of a group ID can
222 be found in section 7.

223 Of note, there exists an ongoing initiative called BIDS-derivatives that aims to provide a BIDS
224 standard for processed data. However, these specifications are not yet finalized and, in their current
225 state, several outputs needed by Clinica are not covered or well adapted. Of note, the notion of group
226 does not exist yet. Nonetheless, we made humble contributions to BIDS-derivatives and we aim to
227 increasingly contribute. Ultimately, the two specifications will probably converge.

228

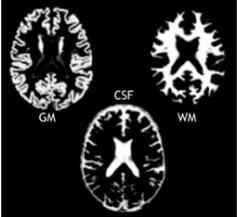
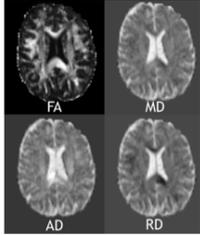
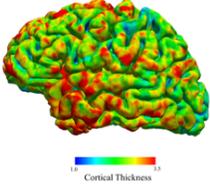
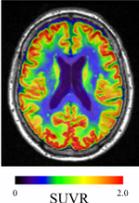
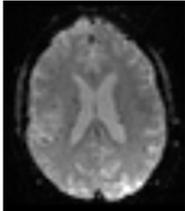
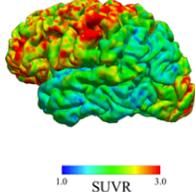
229 **4 Image processing pipelines (clinica run)**

230 This section gives a brief description of the different pipelines currently provided by Clinica as well as
231 the types of features Clinica can produce. An illustrative summary of the pipelines can be found in
232 Figure 2.

233 For technical details, we shall refer to the wiki available on the Clinica website where a longer
234 description of each pipeline is provided.

235

236

Voxel-based pipelines		Surface-based pipelines
<p>t1-volume-* Tissue segmentation (GM, WM, CSF), inter-subject registration using Dartel, spatial normalization to standard space (MNI) Dependencies: SPM, CAT12</p> <ul style="list-style-type: none"> • Voxel-based features (GM, WM, CSF) • Regional features (average GM) using atlases (AAL2, AICHA, Hammers, LPBA40, Neuromorphometrics) 	<p>dwi-preprocessing-*; dwi-dti Correction of raw DWI data, extraction of DTI-based measures, normalization to standard space (MNI) Dependencies: FSL, ANTs, MRtrix3</p> <ul style="list-style-type: none"> • Voxel-based features (FA, MD, AD, RD) • Regional features (average FA, MD, AD, RD) using atlases (JHU DTI81, JHUTracts) 	<p>t1-freesurfer Cortical surface extraction, segmentation of subcortical structures, cortical thickness estimation, spatial normalization to standard space (FsAverage) Dependencies: FreeSurfer</p> <ul style="list-style-type: none"> • Surface-based features (cortical thickness) • Regional features (average cortical thickness) using atlases (Desikan, Destrieux) 
<p>pet-volume Registration to T1 MRI, partial volume correction, spatial normalization to standard space (MNI), intensity normalization Dependencies: SPM, PETPVC, CAT12</p> <ul style="list-style-type: none"> • Voxel-based features (FDG uptake, amyloid uptake) • Regional features (average FDG, amyloid uptake) using atlases (AAL2, AICHA, Hammers, LPBA40, Neuromorphometrics) 	<p>fmri-preprocessing Slice-timing and motion correction, brain extraction, registration to T1 MRI, spatial normalization to standard space (MNI) Dependencies: SPM, FSL (optional)</p> <ul style="list-style-type: none"> • Voxel-based features (BOLD signal) 	<p>pet-surface Registration to T1 MRI, intensity normalization, partial volume correction, projection to cortical surface, spatial normalization to standard space (FsAverage) Dependencies: FreeSurfer, FSL, SPM, PETPVC</p> <ul style="list-style-type: none"> • Surface-based features (FDG uptake, amyloid uptake) • Regional features (average FDG, amyloid uptake) using atlases (Desikan, Destrieux) 

237

238

239

240

Figure 2 List of the pipelines currently available in Clinica with their dependencies and outputs. Explanations regarding the atlases can be found in section 4.5. GM: grey matter; CSF: cerebrospinal fluid; WM: white matter; FA: fractional anisotropy; MD: mean diffusivity; AD: axial diffusivity; RD: radial diffusivity.

241 4.1 Anatomical MRI

242 4.1.1 Processing of T1-weighted MR images for volume analyses using SPM (t1-volume*)

243 The `t1-volume*` pipelines extract voxel-based anatomical features from T1w MR images.
244 Specifically, they perform segmentation of tissues (gray matter (GM), white matter (WM),
245 cerebrospinal fluid (CSF)), normalization to MNI space and computation of regional measures using
246 atlases. Their main outputs are voxel-based maps of tissue density and average measures within cortical
247 regions stored as TSV files.

248 To that purpose, the pipeline wraps the Segmentation, Run Dartel and Normalise to MNI Space
249 routines implemented in SPM (Ashburner, 2012). First, the Unified Segmentation procedure
250 (Ashburner and Friston, 2005) is used to simultaneously perform tissue segmentation, bias field
251 correction and spatial normalization of the input image. Next, a group template is created using
252 DARTEL, an algorithm for diffeomorphic image registration (Ashburner, 2007), from the subjects'
253 tissue probability maps on the native space, usually GM, WM and CSF, obtained at the previous step.
254 The DARTEL to MNI method (Ashburner, 2007) is then applied, providing a registration of the native
255 space images into the MNI space.

256

257 4.1.2 Processing of T1-weighted MR images for surface analyses using FreeSurfer (t1-freesurfer)

258 The `t1-freesurfer` pipeline is mainly a wrapper of the `recon-all` tool of FreeSurfer (Fischl,
259 2012). It performs segmentation of subcortical structures (Fischl et al., 2002, 2004a), extraction of
260 cortical surfaces, cortical thickness estimation (Fischl and Dale, 2000), spatial normalization onto the
261 FreeSurfer surface template (FsAverage) (Fischl et al., 1999), and parcellation of cortical regions using
262 the Desikan and Destrieux atlases (Fischl et al., 2004b). Its main outputs are surface-based cortical
263 thickness features and regional statistics (e.g. regional volume, mean cortical thickness).

264

265 4.2 Diffusion MRI (DWI)

266 4.2.1 DWI pre-processing (dwi-preprocessing-*)

267 The `dwi-preprocessing-*` pipelines correct DWI datasets for motion, eddy current, magnetic
268 susceptibility and bias field distortions, assuming that the data have been acquired using an echo-planar
269 imaging (EPI) sequence.

270 Due to the heterogeneity in acquisitions of fieldmaps and techniques to correct magnetic
271 susceptibility distortions, several pipelines are proposed. Currently, Clinica can handle DWI datasets
272 with fieldmap data containing a phase difference image and at least one magnitude image (case 1 in
273 the BIDS specifications¹², section 8.3.5) (`dwi-preprocessing-using-fieldmap`) and DWI

¹² <http://bids.neuroimaging.io/>

274 datasets with no extra data (`dwi-preprocessing-using-t1`) which is the case of the public
275 Alzheimer's Disease Neuroimaging Initiative (ADNI)¹³ dataset for instance.

276 In all cases, motion and eddy motion corrections are performed with the FSL software
277 (Jenkinson et al., 2012) while bias field is corrected with the ANTs N4 bias correction (Tustison et al.,
278 2010). Regarding susceptibility correction, the FSL prelude/fugue tools were used for the `dwi-`
279 `preprocessing-using-fieldmap` pipeline and the ANTs SyN registration algorithm (Avants et
280 al., 2008; Leow et al., 2007) for the `dwi-preprocessing-using-t1` pipeline. The outputs of the
281 pipelines are the corrected DWI datasets and a brain mask of the `b=0` image.

282 These pipelines are prerequisites for the `dwi-dti` and `dwi-connectome` pipelines.

283

284 **4.2.2 Computation of DTI, DTI-scalar maps and ROI analysis (`dwi-dti`)**

285 The `dwi-dti` pipeline extracts voxel-based features from diffusion tensor imaging (DTI), namely the
286 fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD)
287 using MRtrix3 (Tournier et al., 2012). Then, the DTI-derived scalar maps (FA, MD, AD, RD) are
288 normalized with ANTs (Avants et al., 2008) onto an FA-atlas with labelled tracts. Its main outputs are
289 voxel-based maps from DTI and average measures within tracts stored as TSV files.

290

291 **4.2.3 Computation of fiber orientation distributions, tractogram and structural connectome** 292 **(`dwi-connectome`)**

293 The `dwi-connectome` pipeline computes a weighted graph encoding anatomical connections between
294 a set of brain regions from corrected DWI datasets. To that aim, it relies on the MRtrix3 (Tournier et
295 al., 2012) software to compute the constrained spherical deconvolution diffusion model (Tournier et
296 al., 2007), perform probabilistic tractography (Tournier et al., 2010) and computes a connectome using
297 the Desikan & Destrieux atlases from FreeSurfer (Fischl, 2012). Its main outputs are the diffusion
298 model, the whole-brain tractography and the connectivity matrices.

299

300 **4.3 Functional MRI (fMRI)**

301 **4.3.1 fMRI pre-processing (`fmri-preprocessing`)**

302 The `fmri-preprocessing` pipeline performs the pre-processing of functional MR images. It is
303 almost fully based on SPM (Ashburner, 2012) tools and includes the following processing steps:
304 correction of shifts in the time-series due to the time required to acquire each 2D slice of the volumes,
305 called slice-timing correction (Henson et al., 1999), correction of the subject's movement artifacts (and
306 an optional unwarping of the associated magnetic susceptibility) (Friston et al., 1995), brain extraction
307 using SPM's "Segment" tool followed by FSL's (Jenkinson et al., 2012) dilation, erosion, thresholding
308 and filling operations (alternatively, the user can use the brain extracted image from `t1-freesurfer`
309 pipeline), coregistration with the subject's T1w MR image in native space, spatial normalization into

¹³ <http://adni.loni.usc.edu/>

310 MNI space (Ashburner and Friston, 2005), and spatial smoothing using a Gaussian filter. The main
311 outputs of the pipeline are the corrected fMRI dataset and its brain mask.

312

313 **4.4 Positron emission tomography (PET)**

314 **4.4.1 Processing of PET images for volume analyses (pet-volume)**

315 The `pet-volume` pipeline extracts voxel-based features from PET data. Currently, Clinica is
316 supporting amyloid and FDG PET data but other tracers will be added in the future. Specifically, it
317 performs intra-subject registration of the PET image into the space of the subject's T1w MR image
318 using SPM (Ashburner, 2012) Optionally, partial volume correction (PVC) can be applied thanks to
319 the PETPVC toolbox (Thomas et al., 2016). Then, inter-subject spatial normalization of the PET image
320 into MNI space is performed based on the DARTEL deformation model of SPM (Ashburner, 2007)
321 and intensity normalization is done using the average PET uptake in a reference region resulting in a
322 standardized uptake value ratio (SUVR) map. Its main outputs are voxel-based maps of SUVR map
323 and average measures within cortical regions.

324

325 **4.4.2 Processing of PET images for surface analyses (pet-surface)**

326 The `pet-surface` pipeline extracts the PET signal and projects it onto the cortical surface. More
327 precisely, it performs co-registration of PET and T1w MRI, intensity normalization, PVC with the
328 PETPVC toolbox (Thomas et al., 2016), robust projection of the PET signal onto the subject's cortical
329 surface, parcellation of the cortical regions using the Desikan and Destrieux atlases and spatial
330 normalization onto the FreeSurfer (Fischl, 2012) surface template (FsAverage). Its main outputs are
331 surface-based PET uptake and regional statistics (mean PET uptake) stored as TSV files.

332

333 **4.5 List of atlases available in Clinica**

334 Depending on the modality studied and the type of analysis (voxel-based or surface-based), different
335 atlases can be used to generate regional features. These atlases are briefly listed below and the reader
336 can refer to the documentation available on the website for further details.

337 When performing volumetric processing of T1w MRI and PET images, as done in the `t1-`
338 `volume*` and `pet-volume` pipelines, atlases defined in MNI space containing regions covering the
339 whole cortex and the main subcortical structures available are used (Samper-González et al., 2018),
340 currently AAL2 (Tzourio-Mazoyer et al., 2002), AICHA (Joliot et al., 2015), Hammers (Gousias et al.,
341 2008; Hammers et al., 2003), LPBA40 (Shattuck et al., 2008), and Neuromorphometrics¹⁴. Of note,
342 the user has to download the CAT12¹⁵ toolbox in order to use the Hammers, LPBA40 and
343 Neuromorphometrics atlases.

¹⁴ www.neuromorphometrics.com

¹⁵ <http://www.neuro.uni-jena.de/cat/>

344 When running the `dwi-dti` pipeline, the JHU DTI81 (Hua et al., 2008; Wakana et al., 2007)
 345 and JHUTracts[0|25|50] (Mori et al., 2005) atlases¹⁶, included in FSL (Jenkinson et al., 2012), defined
 346 in MNI space, are used. JHU DTI81 contains 48 white matter tract labels and JHUTracts[0|25|50]
 347 contains 20 white matter probabilistic tract labels thresholded at 0%, 25% and 50%.

348 Finally, surface atlases are used when processing T1w MRI (respectively PET images) with the
 349 `t1-freesurfer` (respectively `pet-surface`) pipelines. Currently, Clinica provides the Desikan-
 350 Killiany (Desikan et al., 2006) atlas, which divides the cerebral cortex into gyri and contains 34 regions
 351 per hemisphere, and the Destrieux (Destrieux et al., 2010) atlas, which divides the cerebral cortex into
 352 gyri and sulci and contains 74 regions per hemisphere.

353

354 4.6 Clinica command line arguments

355 For each pipeline, the command line interface will require a set of arguments which can be compulsory
 356 or optional. The number of mandatory arguments is kept as small as possible, to ease its use. This set
 357 of arguments is gathered into four categories.

358 First, the user will be asked to provide the Clinica mandatory arguments. These arguments are
 359 in general the BIDS directory, the CAPS directory and/or the Group ID, which were described in the
 360 Clinica environment section (sections 3.2 and 3.3).

361 Then, several options are common to every pipeline: the Clinica standard options. For instance,
 362 we can run a pipeline on a subset of participants and sessions by specifying a TSV file. Moreover, it is
 363 possible to specify the number of cores of your machine used to run pipelines in parallel thanks to the
 364 Nipype engine (Gorgolewski et al., 2011). A working directory can be specified for each pipeline. This
 365 directory gathers all the inputs and outputs of the different steps of the pipeline, which is very useful
 366 for debugging. It is especially useful in case a pipeline execution crashes to relaunch it with the exact
 367 same parameters, allowing the execution to continue from the last successfully executed node.

368 Other parameters, specific to each pipeline, are gathered in the category “Optional parameters”.
 369 For instance, when applying a smoothing with a specific full width at half maximum (FWHM), this
 370 parameter can be specified.

371 Finally, advanced parameters for users with good knowledge of the pipeline itself or of the
 372 software behind the pipeline will be gathered in the category “Advanced pipeline options”.

373

374 5 Statistics and machine learning

375 5.1 Statistics

376 5.1.1 Surface-based mass-univariate analysis with SurfStat (statistics-surface)

377 The `statistics-surface` command performs statistical analysis on surface-based features using
 378 the general linear model (GLM). To that aim, the pipeline relies on the Matlab toolbox SurfStat
 379 designed for statistical analyses of univariate and multivariate surface and volumetric data using the

¹⁶ <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases>

380 GLM (Worsley et al., 2009). Surface-based measurements are analyzed on the FsAverage surface
381 template from FreeSurfer and the pipeline can handle cortical thickness from the `t1-freesurfer`
382 pipeline or PET measurements from the `pet-surface` pipeline.

383 Currently, statistical analysis includes group comparison and correlation.

384

385 **5.2 Machine Learning**

386 **5.2.1 Classification based on machine learning**

387 Clinica provides a modular way to perform classification based on machine learning. To build their
388 own classification pipeline, the user can combine three modules based on the scikit-learn library
389 (Pedregosa et al., 2011):

- 390 - Input module. The user can select the inputs from the features available in the CAPS directory,
391 such as gray matter maps obtained from T1w MR images, or SUVR maps obtained from FDG
392 PET images.
- 393 - Algorithm module. The user can choose between different classifiers, currently support vector
394 machine, logistic regression and random forest.
- 395 - Validation module. Several cross-validation (CV) methods are available: k-fold CV, repeated
396 k-fold CV and repeated hold-out CV.

397 Note that no command line interface is available for these specific tools.

398 More details regarding the different modules and a description of the way they can be used to
399 perform reproducible evaluation of classification methods in Alzheimer's disease can be found in
400 (Samper-González et al., 2018) and its dedicated repository¹⁷.

401

402 **5.2.2 Spatially-regularized support vector machine (SVM) (`machinelearning-prepare-spatial-` 403 `svm`)**

404 The `machinelearning-prepare-spatial-svm` pipeline allows the preparation of T1w MRI and
405 PET data to perform classification with an SVM with spatial and anatomical regularization (Cuingnet
406 et al., 2013). In this approach, the standard regularization of the SVM is replaced with a regularization
407 that accounts for the spatial and anatomical structure of neuroimaging data. More specifically, it is
408 regularized with respect to the tissue maps (GM, WM, CSF). As a result, the decision function learned
409 by the algorithm will be more regular and anatomically interpretable. Because the SVM is a kernel
410 method, the spatial/anatomical regularization is done as a pre-processing on the feature maps and the
411 result can then be fed to a standard linear SVM.

¹⁷ <https://github.com/aramis-lab/AD-ML>

412 6 Clinica utilities

413 6.1 Conversion of neuroimaging datasets into a BIDS hierarchy (clinica convert)

414 Clinica provides tools to curate several publicly available neuroimaging datasets and automatically
415 convert them into the BIDS standardized data structure. This section explains what the user needs to
416 download prior to running the converter and the rationale behind the selection of data when multiple
417 acquisitions or pre-processing steps are available. For all converters, the user only needs to download
418 the dataset. All subsequent conversion steps are performed automatically (no user intervention is
419 required) and use parallelization for fast conversion. For further details, the reader can refer to (Samper-
420 González et al., 2018).

421

422 6.1.1 Conversion of the ADNI dataset to BIDS (adni-to-bids)

423 The ADNI to BIDS converter requires the user to have downloaded all the ADNI study data (tabular
424 data in CSV format) and the imaging data of interest. Note that the downloaded files must be kept
425 exactly as they were downloaded. The imaging modalities currently being converted to BIDS include
426 T1w MRI, FLAIR, DWI, fMRI, FDG PET, PiB PET, Florbetapir (AV45) PET and Flortaucipir
427 (AV1451) PET. Clinical data are also converted to BIDS. They include data that do not change over
428 time, such as the subject's sex, education level or diagnosis at baseline, as well as session-dependent
429 data, such as the clinical scores. The clinical data being converted are defined in a spreadsheet that is
430 available with the code of the converter. The user can easily modify this file if they want to convert
431 additional clinical data.

432

433 6.1.2 Conversion of the AIBL dataset to BIDS (aibl-to-bids)

434 The AIBL to BIDS converter requires the user to have downloaded the AIBL non-imaging data (tabular
435 data in CSV format) and the imaging data of interest. For each AIBL participant, the T1w MRI and the
436 Florbetapir, PiB and Flutemetamol PET images are converted. As for the ADNI converter, clinical data
437 converted to BIDS are defined in a spreadsheet available with the code of the converter, which the user
438 can modify.

439

440 6.1.3 Conversion of the NIFD dataset to BIDS (nifd-to-bids)

441 The NIFD to BIDS converter requires the user to have downloaded the NIFD imaging data alongside
442 the corresponding clinical data in CSV format. For each NIFD participant, the T1w MRI, FLAIR, PiB
443 PET and FDG PET images are converted. The clinical data conversion is as described in the previous
444 section.

445

446 6.1.4 Conversion of the OASIS dataset to BIDS (oasis-to-bids)

447 The OASIS to BIDS converter requires the user to have downloaded the OASIS-1 imaging data and
448 the associated CSV file. For each subject, among the multiple T1w MR images available, we select the
449 average of the motion-corrected co-registered individual images resampled to 1 mm isotropic voxels.
450 The clinical data are converted as described in the previous section.

451

452 **6.1.5 Syntax to run the converters**

453 After having downloaded the clinical and imaging data of one of these studies, the conversion of a
454 dataset into BIDS is performed using the following syntax:

```
455 clinica convert <dataset>-to-bids dataset_directory clinical_data_directory  
456 bids_directory
```

457 where <dataset>-to-bids can be `adni-to-bids`, `aibl-to-bids`, `nifd-to-bids` or `oasis-`
458 `to-bids`.

459

460 **6.2 Data handling tools (clinica iotools)**

461 We also propose a set of tools that allows the user to handle BIDS and CAPS datasets. For the moment,
462 there are three types of utility:

- 463 - `check-missing-modalities`: This command checks missing modalities in a BIDS
464 directory.
- 465 - `create-subjects-visits`: This command generates a list of subjects with their sessions
466 based on a BIDS directory and stores the outputs in a TSV file.
- 467 - `merge-tsv`: This command merges all the tabular data including the clinical data of a BIDS
468 directory and the regional features from a CAPS directory (e.g. mean GM density in AAL2
469 atlas, mean FA in JHU DTI81 atlas, etc.) into a single TSV file. This file can then be easily
470 plugged into machine learning tools via Clinica or other statistical/machine learning software
471 packages.

472

473 **7 Usage example**

474 In this section, we propose to show how Clinica can be used to perform a group comparison of FDG
 475 PET data projected on the cortical surface between patients with Alzheimer’s disease and healthy
 476 controls from the ADNI database. An illustrative summary of this example can be found in Figure 3.

477 To download the ADNI dataset, it is necessary to register to the LONI Image & Data Archive¹⁸,
 478 a secure research data repository, and request access to the ADNI dataset through the submission of an
 479 online application form. Both the imaging and clinical data need to be downloaded, each to a folder
 480 that we will call `imaging_data_dir` and `clinical_data_dir`, respectively. The following
 481 command can be used to convert the T1 and FDG PET data of the ADNI dataset into BIDS:

```
482 clinica convert adni-to-bids imaging_data_dir clinical_data_dir ADNI_BIDS
483 --modalities T1 PET_FDG
```

484 where the `ADNI_BIDS` folder contains the conversion of ADNI into BIDS. We can now start processing
 485 the data. First, we need to extract the cortical surfaces from each anatomical image. To do so, we simply
 486 need to type on the terminal the following command:

```
487 clinica run t1-freesurfer ADNI_BIDS ADNI_CAPS
```

488 where the output data will be stored in the `ADNI_CAPS` folder. After visual inspection of the generated
 489 outputs, the FDG PET data can be projected onto the cortex. The command line will be:

```
490 clinica run pet-surface ADNI_BIDS ADNI_CAPS -pet_tracer fdg
```

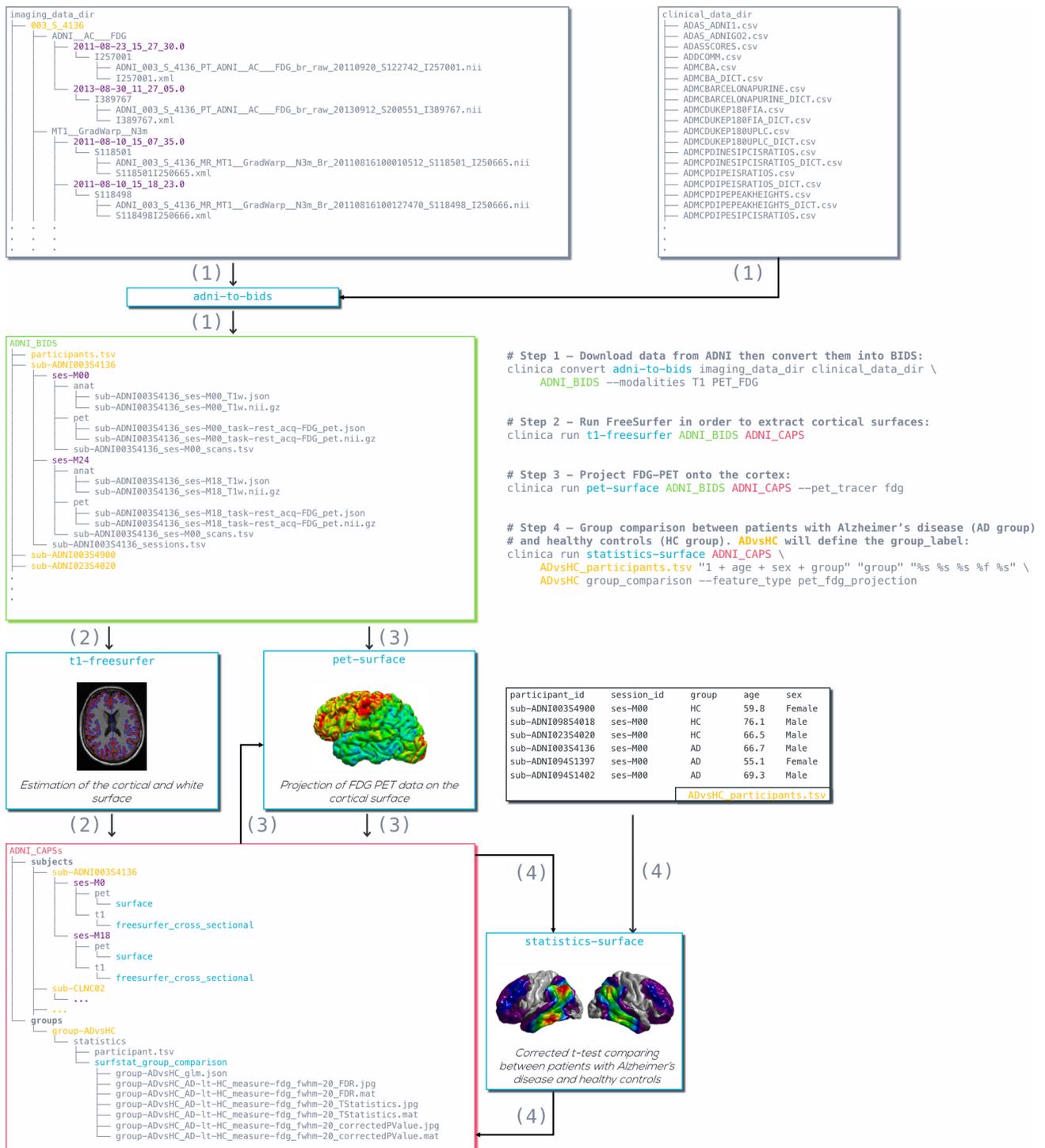
491 Finally, we can perform group comparison of cortical FDG-PET data after having checked the
 492 outputs. The demographic information of the population studied will be stored in a TSV file, looking
 493 as follows:

494	<code>participant_id</code>	<code>session_id</code>	<code>group</code>	<code>age</code>	<code>sex</code>
495	<code>sub-ADNI094S2201</code>	<code>ses-M00</code>	<code>HC</code>	<code>63.7</code>	<code>Female</code>
496	<code>sub-ADNI098S4018</code>	<code>ses-M00</code>	<code>HC</code>	<code>76.1</code>	<code>Male</code>
497	<code>sub-ADNI023S4020</code>	<code>ses-M00</code>	<code>HC</code>	<code>66.5</code>	<code>Male</code>
498	<code>sub-ADNI031S4021</code>	<code>ses-M00</code>	<code>HC</code>	<code>66.5</code>	<code>Male</code>
499	<code>sub-ADNI094S1397</code>	<code>ses-M00</code>	<code>AD</code>	<code>55.1</code>	<code>Female</code>
500	<code>sub-ADNI094S1402</code>	<code>ses-M00</code>	<code>AD</code>	<code>69.3</code>	<code>Male</code>
501	<code>sub-ADNI128S1409</code>	<code>ses-M00</code>	<code>AD</code>	<code>65.9</code>	<code>Male</code>
502	<code>sub-ADNI128S1430</code>	<code>ses-M00</code>	<code>AD</code>	<code>83.4</code>	<code>Female</code>
503	<code>...</code>				

504 where participants with Alzheimer’s disease (respectively healthy controls) have the `AD` label
 505 (respectively `HC` label) in the `group` column. We will call this file `ADvsHC_participants.tsv`.
 506 The first four columns of the TSV file contain strings and the last one contains floats (i.e. numbers),
 507 the format of the columns is therefore `%s %s %s %s %f`. Using age and sex as covariates, the command
 508 line will be:

¹⁸ <https://ida.loni.usc.edu>

```
509 clinica run statistics-surface ADNI_CAPS \  
510     ADvsHC_participants.tsv "1 + age + sex + group" "group" "%s %s %s %s  
511 %f" \  
512     ADvsHC group_comparison -feature_type pet_fdg_projection  
513     The results of the statistical analysis will be stored in the ADNI_CAPS/groups/group-  
514     ADvsHC folder.
```



515

516 **Figure 3** Diagram illustrating the Clinica pipelines involved when performing a group comparison of
 517 FDG PET data projected on the cortical surface between patients with Alzheimer's disease and healthy
 518 controls from the ADNI database. First, clinical and neuroimaging data are downloaded from the ADNI
 519 website and data are converted into BIDS with the `adni-to-bids` tool from Clinica (1). Estimation
 520 of the cortical and white surface is then produced by the `t1-freesurfer` pipeline in a single
 521 command line (2). Afterwards, FDG PET data can be projected on the subject's cortical surface and
 522 normalized to the FsAverage template from FreeSurfer using the `pet-surface` pipeline (3). Finally,
 523 a TSV file with demographic information of the population studied is given to the `statistics-`

524 surface pipeline to generate the results of the group comparison between patients with Alzheimer's
525 disease and healthy controls (4).

526 **8 Discussion**

527 We proposed a software platform that aims at making clinical neuroscience easier and more
528 reproducible. Clinica automates the processing of standardized pipelines involving several
529 neuroimaging modalities (currently, anatomical MRI, diffusion MRI, functional MRI and PET) as well
530 as statistics and machine learning tools. Additionally, Clinica provides tools to convert public
531 neuroimaging datasets focused on dementia (ADNI, AIBL, OASIS and NIFD) into the BIDS standard,
532 and tools to handle raw (BIDS) and processed datasets. The use of the BIDS standard as the only
533 prerequisite on the data and the unified command line interface across the pipelines ease the processing
534 automation. The image analysis automation is also improved by the use of the CAPS hierarchy, which
535 facilitates the chaining of pipelines. The target audience of Clinica is neuroscientists or clinicians
536 conducting clinical neuroscience studies involving multimodal imaging, and researchers developing
537 advanced machine learning algorithms.

538 The last three decades witnessed the development of many software packages for the processing
539 of neuroimaging data. A first category of packages comprises those implementing innovative image
540 processing methodologies (e.g. tissue segmentation, registration). Many tools fall into this category,
541 for instance SPM (Friston et al., 2007), AFNI (Cox, 1996), FreeSurfer (Fischl, 2012), FSL (Jenkinson
542 et al., 2012), PETPVC (Thomas et al., 2016), Camino (A Cook et al., 2005), Dipy (Garyfallidis et al.,
543 2014), DTI-TK¹⁹, MRtrix (Tournier et al., 2012) or ANTs (Avants et al., 2014). Some of these tools
544 cover a variety of modalities while other focus on a specific one (diffusion MRI for Camino, Dipy,
545 DTI-TK and MRtrix, fMRI for AFNI, PET for PETPVC). However, performing a multimodal study
546 can be difficult because one needs to combine tools from different packages. This results in complex
547 pipelines which can be difficult to build, maintain and distribute. Even when analyzing a single
548 modality, one often wants to combine tools from different packages, thus facing similar difficulties.
549 Combination of tools is made even more difficult by the fact that the input and output data are organized
550 differently by each tool.

551 Efforts of the community have alleviated several of these difficulties. The NeuroDebian
552 community²⁰ (Halchenko and Hanke, 2012) aims to provide and ease the installation of a large
553 collection of software packages for the Debian distribution. The Nipype (Gorgolewski et al., 2011)
554 system facilitates the building of complex pipelines through the wrapping of tools in Python. BIDS
555 (Gorgolewski et al., 2017) provides a standard for organizing data. BIDS-Apps provides versions of
556 software packages using BIDS for data organization. More generally, the NIPY community aims to
557 provide a comprehensive set of tools for the analysis of neuroimaging data in a single language, Python.
558 However, many useful tools and packages remain outside of the NIPY scope, being written in different
559 languages. As mentioned above, Nipype allows wrapping these heterogeneous tools. It is a powerful
560 and particularly useful tool for that aim. However, the building of pipelines remains left to the user.

561 There are also software packages that integrate different tools within a single environment. This
562 is for example the case of BCBtoolkit (Foulon et al., 2018), BrainVISA (Cointepas et al., 2001),
563 BrainSuite²¹, fMRIPrep (Esteban et al., 2019), or Pypes (Savio et al., 2017). Clinica falls within this
564 category. It shares some characteristics with these tools but also has important differences. The
565 BCBtoolkit wraps neuroimaging software packages from the community but also highlights new
566 methodological developments to evaluate brain disconnections. BCBookit does not use any pipelining

¹⁹ <http://dti-tk.sourceforge.net>

²⁰ <http://neuro.debian.net/>

²¹ <http://brainsuite.org>

567 system but instead wraps bash scripts that are then made available through a GUI. The BrainVISA
568 platform, even though it also wraps some existing tools, mainly provides innovative tools for the
569 analysis of human or animal brain imaging data. Moreover, it includes its own pipelining system while
570 Clinica relies on the community effort Nipype. BrainSuite does not wrap existing tools but provides a
571 set of innovative tools for the analysis of neuroimaging data. It can be executed using a GUI, a
572 command line, a Nipype interface, or as a BIDS App. fMRIPrep combines software components using
573 Nipype (Gorgolewski et al., 2011) to provide a robust pipeline for pre-processing of fMRI data. It
574 assumes input data to follow the BIDS standard (Gorgolewski et al., 2016) and outputs are organized
575 following the ongoing BIDS-derivatives initiative. Pypes is probably the closest in spirit to Clinica: it
576 also focuses on the integration of existing tools into a set of reusable pipelines built with Nipype. The
577 user needs to specify a configuration file to describe the input data even for BIDS datasets. The output
578 data will then follow the same structure as the input data and can be chained to other pipelines from
579 Pypes. Note that the ability to chain pipelines exists for a limited number of pipelines in the BIDS-App
580 version of BrainSuite and is likely to be more present in BIDS Apps with the advent of BIDS-
581 derivatives. Although Clinica does not provide a GUI, efforts were made to simplify as much as
582 possible its command line interface, which is feasible thanks to the autocompletion, the standard
583 organization of data and the documentation which was designed in order to be readable by a newcomer.

584 Machine learning is now widely used in the neuroimaging community, for cognitive
585 neuroscience or computer-aided diagnosis applications. However, applying such approaches to
586 neuroimaging data can be difficult for newcomers. Nilearn (Abraham et al., 2014) has allowed major
587 progress in that direction by providing in a single environment tools for preprocessing, data
588 manipulation, feature extraction and machine learning wrapping the scikit-learn library (Pedregosa et
589 al., 2011). However, Nilearn currently mostly targets cognitive neuroimaging and mainly deals with
590 functional neuroimaging. On the other hand, Clinica is dedicated to clinical neuroimaging studies, such
591 as biomarker design and clinical decision support systems. As a result, Clinica does not aim to deal
592 with task-based fMRI and is currently mostly focused on the analysis of T1w MRI, diffusion MRI and
593 PET data.

594 A limitation of Clinica is that it is currently operating system dependent, which is not the case
595 of Docker-based frameworks like BIDS Apps. Another limitation is that, even though the core of
596 Clinica is easy to install with Conda, the user has to download and install the third-party software
597 packages prior to running any pipeline. To solve this issue, we aim to provide very soon a container
598 image of Clinica using Docker (Glatard et al., 2017; Gorgolewski et al., 2017). Then, we will integrate
599 the Singularity container technology to import the Docker image and enable the use of Clinica on high
600 performance computers.

601 We aim to add several functionalities to Clinica. Both traceability and quality control are essential
602 to ensure objective and reproducible neuroimaging studies and will be added to Clinica. Clinica
603 currently focuses on cross-sectional analyses while many neuroimaging studies are longitudinal. Thus,
604 new pipelines supporting longitudinal analysis of data will also be developed. As statistical tools are
605 currently limited to the analysis of surface data, we will develop a pipeline for the statistical analysis
606 of volume data.

607

608 9 Conflict of Interest

609 The authors declare that the research was conducted in the absence of any commercial or financial
610 relationships that could be construed as a potential conflict of interest.

611 10 Author Contributions

612 Guarantor of integrity of entire study, AR; study concepts/study design, all authors; manuscript drafting
613 or manuscript revision for important intellectual content, all authors; approval of final version of
614 submitted manuscript, all authors; literature research, AR, NB, OC; and manuscript editing, all authors.

615 11 Funding

616 The research leading to these results has received funding from the program “Investissements d’avenir”
617 ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6)
618 ANR-11-IDEX-004 (Agence Nationale de la Recherche-11- Initiative d’Excellence-004, project
619 LearnPETMR number SU-16-R-EMR-16), from the European Union H2020 program (project
620 EuroPOND, grant number 666992, project HBP SGA1 grant number 720270), from the joint
621 NSF/NIH/ANR program “Collaborative Research in Computational Neuroscience” (project
622 HIPLAY7, grant number ANR-16-NEUC-0001-01), from Agence Nationale de la Recherche (project
623 PREVDEMALS, grant number ANR-14-CE15-0016-07), from the ICM Big Brain Theory Program
624 (project DYNAMO), from the Inria Project Lab Program (project Neuromarkers), from the European
625 Research Council (to Dr Durrleman project LEASP, grant number 678304), and from the “Contrat
626 d’Interface Local” program (to Dr Colliot) from Assistance Publique-Hôpitaux de Paris (AP-HP). N.B.
627 received funding from the People Programme (Marie Curie Actions) of the European Union’s Seventh
628 Framework Programme (FP7/2007-2013) under REA grant agreement no. PCOFUND-GA-2013-
629 609102, through the PRESTIGE programme coordinated by Campus France.

630 12 Acknowledgements

631 This manuscript has been released as a Pre-Print at (Routier, 2018).

632

633 **13 Reference**

- 634 A Cook, P., Bai, Y., Nedjati-Gilani, S., K Seunarine, K., Hall, M., Parker, G., et al. (2005). Camino:
635 Open-Source Diffusion-MRI Reconstruction and Processing. *Proc. Intl. Soc. Magn. Reson.*
636 *Med.* 14.
- 637 Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014).
638 Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* 8.
639 doi:10.3389/fninf.2014.00014.
- 640 Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage* 38, 95–113.
641 doi:10.1016/j.neuroimage.2007.07.007.
- 642 Ashburner, J. (2012). SPM: A history. *Neuroimage* 62–248, 791–800.
643 doi:10.1016/j.neuroimage.2011.10.025.
- 644 Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *NeuroImage* 26, 839–851.
645 doi:10.1016/j.neuroimage.2005.02.018.
- 646 Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image
647 registration with cross-correlation: Evaluating automated labeling of elderly and
648 neurodegenerative brain. *Medical Image Analysis* 12, 26–41. doi:10.1016/j.media.2007.06.004.
- 649 Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., and Gee, J. C. (2014). The Insight
650 ToolKit image registration framework. *Front Neuroinform* 8. doi:10.3389/fninf.2014.00044.
- 651 Brett, M., Hanke, M., Markiewicz, C., Marc-Alexandre Côté, McCarthy, P., Cheng, C., et al. (2019).
652 *nipy/nibabel: 2.3.3*. Zenodo doi:10.5281/zenodo.2541736.
- 653 Cointepas, Y., Mangin, J.-F., Garnero, L., Poline, J.-B., and Benali, H. (2001). BrainVISA: Software
654 platform for visualization and analysis of multi-modality brain data. *NeuroImage* 13, 98.
655 doi:10.1016/S1053-8119(01)91441-7.
- 656 Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance
657 neuroimages. *Comput. Biomed. Res.* 29, 162–173.
- 658 Cuingnet, R., Glaunès, J. A., Chupin, M., Benali, H., Colliot, O., and Alzheimer's Disease
659 Neuroimaging Initiative (2013). Spatial and Anatomical Regularization of SVM: A General
660 Framework for Neuroimaging Data. *IEEE Trans Pattern Anal Mach Intell* 35, 682–696.
661 doi:10.1109/TPAMI.2012.142.
- 662 Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An
663 automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral
664 based regions of interest. *NeuroImage* 31, 968–980. doi:10.1016/j.neuroimage.2006.01.021.
- 665 Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical
666 gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53, 1–15.
667 doi:10.1016/j.neuroimage.2010.06.010.

- 668 Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2019).
 669 fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* 16, 111–116.
 670 doi:10.1038/s41592-018-0235-4.
- 671 Fischl, B. (2012). FreeSurfer. *NeuroImage* 62, 774–781. doi:10.1016/j.neuroimage.2012.01.021.
- 672 Fischl, B., and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from
 673 magnetic resonance images. *PNAS* 97, 11050–11055. doi:10.1073/pnas.200033797.
- 674 Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole Brain
 675 Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain.
 676 *Neuron* 33, 341–355. doi:10.1016/S0896-6273(02)00569-X.
- 677 Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., et al. (2004a).
 678 Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 23, S69–S84.
 679 doi:10.1016/j.neuroimage.2004.07.016.
- 680 Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical Surface-Based Analysis: II: Inflation,
 681 Flattening, and a Surface-Based Coordinate System. *NeuroImage* 9, 195–207.
 682 doi:10.1006/nimg.1998.0396.
- 683 Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., et al. (2004b).
 684 Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex* 14, 11–22.
 685 doi:10.1093/cercor/bhg087.
- 686 Foulon, C., Cerliani, L., Kinkingnéhun, S., Levy, R., Rosso, C., Urbanski, M., et al. (2018). Advanced
 687 lesion symptom mapping analyses and implementation as BCBtoolkit. *GigaScience* 7.
 688 doi:10.1093/gigascience/giy004.
- 689 Frackowiak, R. S. J., Friston, K. J., Frith, C., Dolan, R., and Mazziotta, J. C. eds. (1997). *Human Brain*
 690 *Function*. Academic Press USA Available at:
 691 <http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf1/>.
- 692 Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W. (2007). *Statistical Parametric*
 693 *Mapping*. Elsevier doi:10.1016/B978-0-12-372560-8.X5000-1.
- 694 Friston, K. J., Frith, C. D., Frackowiak, R. S. J., and Turner, R. (1995). Characterizing Dynamic Brain
 695 Responses with fMRI: A Multivariate Approach. *NeuroImage* 2, 166–172.
 696 doi:10.1006/nimg.1995.1019.
- 697 Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., et al. (2014).
 698 Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* 8.
 699 doi:10.3389/fninf.2014.00008.
- 700 Glatard, T., Kiar, G., Aumentado-Armstrong, T., Beck, N., Bellec, P., Bernard, R., et al. (2017).
 701 Boutiques: a flexible framework for automated application integration in computing platforms.
 702 *arXiv:1711.09713 [cs]*. Available at: <http://arxiv.org/abs/1711.09713> [Accessed June 22,
 703 2018].

- 704 Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011).
705 Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in
706 Python. *Front. Neuroinform.* 5. doi:10.3389/fninf.2011.00013.
- 707 Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., et al.
708 (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging
709 data analysis methods. *PLOS Computational Biology* 13, e1005209.
710 doi:10.1371/journal.pcbi.1005209.
- 711 Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The
712 brain imaging data structure, a format for organizing and describing outputs of neuroimaging
713 experiments. *Scientific Data* 3, 160044. doi:10.1038/sdata.2016.44.
- 714 Gousias, I. S., Rueckert, D., Heckemann, R. A., Dyet, L. E., Boardman, J. P., Edwards, A. D., et al.
715 (2008). Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest.
716 *NeuroImage* 40, 672–684. doi:10.1016/j.neuroimage.2007.11.034.
- 717 Halchenko, Y. O., and Hanke, M. (2012). Open is Not Enough. Let's Take the Next Step: An
718 Integrated, Community-Driven Computing Platform for Neuroscience. *Front. Neuroinform.* 6.
719 doi:10.3389/fninf.2012.00022.
- 720 Hammers, A., Allom, R., Koepp, M. J., Free, S. L., Myers, R., Lemieux, L., et al. (2003). Three-
721 dimensional maximum probability atlas of the human brain, with particular reference to the
722 temporal lobe. *Human Brain Mapping* 19, 224–247. doi:10.1002/hbm.10123.
- 723 Henson, R. N. A., Buechel, C., Josephs, O., and Friston, K. J. (1999). The slice-timing problem in
724 event-related fMRI. *NeuroImage* 9, 125.
- 725 Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., et al. (2008). Tract probability maps in
726 stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification.
727 *NeuroImage* 39, 336–347. doi:10.1016/j.neuroimage.2007.07.053.
- 728 Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL.
729 *Neuroimage* 62, 782–790. doi:10.1016/j.neuroimage.2011.09.015.
- 730 Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L., et al. (2015). AICHA: An atlas of
731 intrinsic connectivity of homotopic areas. *Journal of Neuroscience Methods* 254, 46–59.
732 doi:10.1016/j.jneumeth.2015.07.013.
- 733 Jones, E., Oliphant, T., Peterson, P., and others (2001). *SciPy: Open source scientific tools for Python*.
734 Available at: <http://www.scipy.org/>.
- 735 Leow, A. D., Yanovsky, I., Chiang, M. C., Lee, A. D., Klunder, A. D., Lu, A., et al. (2007). Statistical
736 Properties of Jacobian Maps and the Realization of Unbiased Large-Deformation Nonlinear
737 Image Registration. *IEEE Transactions on Medical Imaging* 26, 822–832.
738 doi:10.1109/TMI.2007.892646.
- 739 McKinney, W. (2010). Data Structures for Statistical Computing in Python. in *Proceedings of the 9th*
740 *Python in Science Conference*, eds. S. van der Walt and J. Millman, 51–56.

- 741 Mori, S., Wakana, S., Nagae-Poetscher, L., and van Zijl, P. (2005). *MRI Atlas of Human White Matter*.
742 Amsterdam: Elsevier.
- 743 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-
744 learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- 745 Poline, J.-B., Breeze, J. L., Ghosh, S. S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., et al. (2012).
746 Data sharing in neuroimaging research. *Front. Neuroinform.* 6. doi:10.3389/fninf.2012.00009.
- 747 Routier, A. (2018). Imagerie cérébrale multimodale pour l'étude des aphasies primaires progressives.
748 Available at: <https://tel.archives-ouvertes.fr/tel-01992799>.
- 749 Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., et al. (2018).
750 Reproducible evaluation of classification methods in Alzheimer's disease: Framework and
751 application to MRI and PET data. *NeuroImage*. doi:10.1016/j.neuroimage.2018.08.042.
- 752 Savio, A. M., Schutte, M., Graña, M., and Yakushev, I. (2017). Pypes: Workflows for Processing
753 Multimodal Neuroimaging Data. *Front Neuroinform* 11. doi:10.3389/fninf.2017.00025.
- 754 Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008).
755 Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage* 39, 1064–
756 1080. doi:10.1016/j.neuroimage.2007.09.031.
- 757 Thomas, B. A., Cuplov, V., Bousse, A., Mendes, A., Thielemans, K., Hutton, B. F., et al. (2016).
758 PETPVC: a toolbox for performing partial volume correction techniques in positron emission
759 tomography. *Physics in Medicine and Biology* 61, 7975–7993. doi:10.1088/0031-
760 9155/61/22/7975.
- 761 Tournier, J.-D., Calamante, F., and Connelly, A. (2007). Robust determination of the fibre orientation
762 distribution in diffusion MRI: Non-negativity constrained super-resolved spherical
763 deconvolution. *NeuroImage* 35, 1459–1472. doi:10.1016/j.neuroimage.2007.02.016.
- 764 Tournier, J.-D., Calamante, F., and Connelly, A. (2010). Improved probabilistic streamlines
765 tractography by 2nd order integration over fibre orientation distributions. in *Proceedings of the*
766 *International Society for Magnetic Resonance in Medicine* Available at:
767 https://cds.ismrm.org/protected/10MProceedings/files/1670_4298.pdf [Accessed January 25,
768 2019].
- 769 Tournier, J.-D., Calamante, F., and Connelly, A. (2012). MRtrix: Diffusion tractography in crossing
770 fiber regions. *Int. J. Imaging Syst. Technol.* 22, 53–66. doi:10.1002/ima.22005.
- 771 Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010).
772 N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging* 29, 1310–1320.
773 doi:10.1109/TMI.2010.2046908.
- 774 Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al.
775 (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic
776 Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage* 15, 273–289.
777 doi:10.1006/nimg.2001.0978.

- 778 van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient
779 Numerical Computation. *Comput. Sci. Eng.* 13, 22–30. doi:10.1109/MCSE.2011.37.
- 780 Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., et al. (2007).
781 Reproducibility of quantitative tractography methods applied to cerebral white matter.
782 *NeuroImage* 36, 630–644. doi:10.1016/j.neuroimage.2007.02.049.
- 783 Worsley, K., Taylor, J., Carbonell, F., Chung, M., Duerden, E., Bernhardt, B., et al. (2009). SurfStat:
784 A Matlab toolbox for the statistical analysis of univariate and multivariate surface and
785 volumetric data using linear mixed effects models and random field theory. *NeuroImage* 47,
786 S102. doi:10.1016/S1053-8119(09)70882-1.
- 787
- 788