

Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

☐ This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Abstract

The simulated data are produced by the included R code. The hippocampus segmentation example uses data from the Alzheimer's Disease Neuroimaging Initiative. The Data Use Agreement prohibits the authors directly sharing the data, but the data are readily available from the ADNI upon approval of the request. (Most requests are granted quickly.) The authors have included detailed pre-processing scripts for producing the inputs to the MCMC code, including web addresses of the data (after approval) and step-by-step processing instructions. The authors are happy to consult with any interested parties for acquiring the ADNI data and processing them accordingly. A .csv file (./Results/All_AD_Brains/Ctrl2Ad_GMI_Summary.csv) is included with summary measures from the aggregated analysis, which can be used to produce Figure 12 in the manuscript.

Availability

- ☒ Data **are** publicly available
☐ Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available data* section, below.

Publicly available data

- ☐ Data are available online at:
☐ Data are available as part of the paper's supplementary material.

☒ Data are publicly available by request, following the process described here:

A user may request access to the data at the ADNI website:

<http://adni.loni.usc.edu/data-samples/access-data/>

☐ Data are or will be made available through some other mechanism, described here:

Non-publicly available data

Discussion of lack of publicly available data:

Description

File format(s)

- ☒ CSV or other plain text: .csv, .txt, .md
- ☒ Software-specific binary format (.Rda, Python pickle, etc.):
- ☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
- ☒ Other (described here):

Preprocessing the ADNI data will produce NIfTI (.nii.gz) files that R can read

Data dictionary

- ☐ Provided by the authors in the following file(s):
- ☐ Data file(s) is (are) self-describing (e.g., netCDF files)
- ☒ Available at the following URL:

<http://adni.loni.usc.edu/data-samples/access-data/>
<http://www.hippocampal-protocol.net/SOPs/index.php>

Additional information (optional)

--

Part 2: Code

Abstract

The provided code includes preprocessing shell scripts and Python / R script files for generating the processed ADNI data after it has been downloaded, as well as R and C++ scripts for producing the results (including figures) in Sections 3 and 4 of the manuscript, after the processed data have been produced. (The processed data are needed for Section 4, but not for Section 3.)

Description

Code format(s)

- ☒ Script files
 - ☒ R ☒ Python ☐ Matlab
 - ☒ Other: C++
- ☐ Package
 - ☐ R ☐ Python ☐ MATLAB toolbox
 - ☐ Other:
- ☐ Reproducible report
 - ☐ R Markdown ☐ Jupyter notebook
 - ☐ Other:
- ☒ Shell script
- ☐ Other (described here):

Supporting software requirements

Version of primary software used

Preprocessing:

R 3.2.5 and Python 3.9.2

Paper Results:

R 4.0.2 and Matlab R2018a

Libraries and dependencies used by the code

Preprocessing:

Dependencies: ANTs 2.3.1, ANTs2 2.2.0-111, MASS 1.1.0, FSL 6.0.3

Python libraries: os, sys, re, nipype

(Non-base) R packages: fslr 2.22.0, neurobase 1.19.1, oro.nifti 0.9.1, extrantsr 2.17, ANTsR 0.4.0, WhiteStripe 2.2.2

Paper Results:

Matlab Toolbox: Image Processing 10.2

(Non-base) R packages: gridExtra 2.3, reshape2 1.4.4, forcats 0.5.1, stringr 1.4.0, dplyr 1.0.5, purrr 0.3.4, readr 1.4.0, tidyr 1.1.3, tibble 3.1.2, ggplot2 3.3.5, tidyverse 1.3.1, R.matlab 3.6.2, RcppArmadillo 0.0.900.1.0, RcppTN 0.2-2, Rcpp 1.0.5, Matrix 1.2-18, coda 0.19-3, neurobase 1.29.0, oro.nifti 0.11.0, msm 1.6.8, mnormt 2.0.2, spatstat 2.1-0, spatstat.linnet 2.2-1, spatstat.core 2.2-0, rpart 2.1-15, nlme 3.1-148, spatstat.geom 2.2-0, spatstat.data 2.1-0, mvtnorm 1.1-1, lattice 0.20-41

Supporting system/hardware requirements (optional)

These are NOT required, but were used in generating the results:

Preprocessing: Scripts are written for accessing a computer cluster running CentOS Linux 7 (Core), 2x Intel Xeon 2.3 GHz CPU, 512 GB RAM

Paper results: Dell Precision T3620, Windows 10 x64 (build 19042), Intel Xeon 4.1 GHz CPU, 64GB RAM

Parallelization used

- ☐ No parallel code used
- ☒ Multi-core parallelization on a single machine/node
 - Number of cores used: 7
- ☐ Multi-machine/multi-node parallelization
 - Number of nodes and cores used:

License

- ☐ MIT License (default)

- ☐ BSD
- ☐ GPL v3.0
- ☐ Creative Commons
- ☐ Other (described here):

Additional information (optional)

Part 3: Reproducibility workflow

Scope

The provided workflow reproduces:

- ☒ Any numbers provided in text in the paper
- ☐ All tables and figures in the paper
- ☒ Selected tables and figures in the paper, as explained and justified here:

Figure 1 will not be reproduced, since it is just a graphical depiction of the model.

All results and figures in Section 3 of the manuscript will be reproduced.

After applying for and downloading the ADNI data (see above), all of the results and figures in Section 4 can be reproduced.

Workflow details

Format(s)

- ☐ Single master code file
- ☒ Wrapper (shell) script(s)
- ☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach
- ☒ Text file (e.g., a readme-style file) that documents workflow
- ☐ Makefile
- ☒ Other (more detail in 'Instructions' below)

Instructions

Section 3 Results:

Run SimulationStudyWrapper.R in the main directory. This will reproduce all of the results and figures in Section 3.

Section 4 Results:

The preprocessing code is in its own folder (`~/Preprocessing/code_for_revision3`). The main wrapper file for this step is `_preprocess_wrapper.sh`, which is written as batch jobs submitted to a computer cluster. This will produce all of the processed NIfTI files. For subject ID 0592, the associated segmentations, T1 images, and similarity images should be copied (cut) and pasted into `ADNI_data_3D/0592_3D/`. For the remaining subject IDs, the associated segmentations, T1 images, and similarity images should be copied/cut and pasted into `ADNI_data_3D/all3D/<subject id>/`. Once all of these input files are in place, one needs to only

run ADNI_Application_Wrapper.R in the main directory. This will reproduce all of the results and figures in Section 4.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

- ☐ <1 minute
- ☐ 1-10 minutes
- ☐ 10-60 minutes
- ☐ 1-8 hours
- ☒ >8 hours
- ☒ Not feasible to run on a desktop machine, as described here:

The most computationally expensive steps are for reproducing Section 4.

Preprocessing: On the cluster above with the specified hardware, the MASS skull-stripping step takes ~2 days for each target image. The rest of the preprocessing steps can be done in < 1 hour. The joint label fusion segmentations take ~32 hours to produce.

Paper Results: The setup steps are very fast (only a couple of minutes). Each of the MCMC runs in Subsection 3.1 and 3.2 take about 30-45 minutes each. The summary results / figures in Section 3 are very quick. For Section 4, the setup / data preparation step takes < 5 minutes. Each MCMC run in Section 4.1 takes approximately 17 hours. The repeated MCMC runs over the remaining atlases in Section 4.2 take ~4 days. The summary results and figures take < 5 minutes.

Additional documentation (optional)

Notes (optional)

If requested to verify reproducibility, we are willing to directly share the ADNI NIfTI files with the editorial board and review team. However, we would not be comfortable sharing these data with the broader readership if the paper is accepted and will refer them instead to the included preprocessing code.

If the paper is accepted, all of the code will be made available on the lead author's Github page. (URL withheld for blinded review).