HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2023 August 18.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2023 February ; 12464: . doi:10.1117/12.2653132.

Automatic Preprocessing Pipeline for White Matter Functional Analyses of Large-Scale Databases

Yurui Gao[*,a,b], **Richard D Lawless**[a,c], **Muwei Li**[a,d], **Yu Zhao**[a,d], **Kurt G Schilling**[a,d], **Lyuan Xu**[a,c], **Andrea T Shafer**[e], **Lori L Beason-Held**[e], **Susan M Resnick**[e], **Baxter P Rogers**[a,d], **Zhaohua Ding**[a,c], **Adam W Anderson**[a,b], **Bennett A Landman**[a,b,c,d], **John C Gore**[a,b,d], **Alzheimer's Disease Neuroimaging Initiative**

[a]Vanderbilt University Institute of Imaging Science, Vanderbilt University Medical Center, Nashville, TN, USA;

[b]Biomedical Engineering, Vanderbilt University, Nashville, TN, USA;

[c]Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA;

[d]Radiology and Radiological Sciences, Vanderbilt University Medical Center, Nashville, TN, USA;

[e]Laboratory of Behavioral Neuroscience, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

Abstract

Recently, increasing evidence suggests that fMRI signals in white matter (WM), conventionally ignored as nuisance, are robustly detectable using appropriate processing methods and are related to neural activity, while changes in WM with aging and degeneration are also well documented. These findings suggest variations in patterns of BOLD signals in WM should be investigated. However, existing fMRI analysis tools, which were designed for processing gray matter signals, are not well suited for large-scale processing of WM signals in fMRI data. We developed an automatic pipeline for high-performance preprocessing of fMRI images with emphasis on quantifying changes in BOLD signals in WM in an aging population. At the image processing level, the pipeline integrated existing software modules with fine parameter tunings and modifications to better extract weaker WM signals. The preprocessing results primarily included whole-brain time-courses, functional connectivity, maps and tissue masks in a common space. At the job execution level, this pipeline exploited a local XNAT to store datasets and results, while using DAX tool to automatic distribute batch jobs that run on high-performance computing clusters. Through the pipeline, 5,034 fMRI/T1 scans were preprocessed. The intraclass correlation coefficient (ICC) of test-retest experiment based on the preprocessed data is 0.52 – 0.86 (N=1000), indicating a high reliability of our pipeline, comparable to previously reported ICC in gray matter experiments. This preprocessing pipeline highly facilitates our future analyses on WM functional alterations in aging and may be of benefit to a larger community interested in WM fMRI studies.

[*] yurui.gao@vanderbilt.edu; phone 1 (615) 322-8359.

Keywords

white matter function; resting state fMRI; aging; Alzheimer's disease; XNAT; automatic preprocessing pipeline

1. INTRODUCTION

The majority of conventional functional MRI (fMRI) studies have focused on neural activity in gray matter (GM) while ignoring the blood oxygenation level dependent (BOLD) signal in white matter (WM), at times even considering the WM signal as a nuisance regressor. Recently, a growing body of evidence has demonstrated that BOLD signals in WM, although weaker, are reliably detectable and related to neural activity; moreover, local signal changes and inter-regional correlations within WM can be robustly measured with appropriate methods[1–6]. These findings stimulate the demand for more elaborate studies of BOLD signals in WM in different clinical and developmental populations[7,8].

Structural deteriorations in WM have been consistently found in aging and Alzheimer's disease (AD) populations[9,10]; however, the age- or AD- related alterations in WM BOLD signals measured with fMRI are barely reported[7], especially in large-scale analyses. Fortunately, increasingly abundant databases for both longitudinal and cross-sectional analyses on aging and AD are available, including, but not limited to, the Alzheimer's Disease Neuroimaging Initiative (ADNI), Baltimore Longitudinal Study of Aging (BLSA), and Open Access Series of Imaging Studies (OASIS), which provide the data basis for investigations of how WM function alters with aging of human brain and progression of AD.

We developed an automatic pipeline to preprocess fMRI images of whole brains with an emphasis on WM in aging populations, aiming to prepare for succeeding large-scale WM-related fMRI analyses including detection of BOLD signals, feature extraction, classification and statistical analyses. We executed the pipeline to complete pre-processing of over 5,000 fMRI scans acquired from multiple aging and AD databases. Based on a subset of the pre-processed data, we conducted a test-retest experiment to evaluate the reproducibility of the proposed pipeline.

2. METHODS

2.1 Data

T1-weighted (T1w) images and resting state fMRI (rsfMRI) images were downloaded from existing databases: ADNI (stages 2 and 3, https://adni.loni.usc.edu), BLSA (https://www.blsa.nih.gov) and OASIS (stage 3, https://www.oasis-brains.org) whose images were acquired with relatively consistent imaging protocols. Those downloaded images in DICOM format were converted into Nifti files using dcm2nii. All the Nifti images and the corresponding JSON files were curated, organized and stored on the Vanderbilt University Institute of Imaging Science - eXtensible Neuroimaging Archive Toolkit (VUIIS-XNAT[11], https://xnat4.vandyxnat.org).

2.2 Core Routine for Preprocessing

The core routine for the rsfMRI/T1 preprocessing, coded in MATLAB language, is graphically shown in Fig. 1. Briefly, the rsfMRI images were corrected for slice timing and head motion, followed by regressing out 24 motion-related parameters[12] and the mean cerebrospinal fluid (CSF) signal. The resulting rsfMRI data were detrended and temporally filtered with passband frequency spanning from 0.01 to 0.1Hz. Furthermore, whole-brain maps of amplitude of low-frequency fluctuation (ALFF[13]), fractional ALFF (fALFF[14]), and regional homogeneity (ReHo[15]) were calculated in individual spaces. All these steps were implemented using modules in Data Processing Assistant for Resting-State fMRI (DPARSF[16]), an SPM-based toolbox. Regarding corresponding T1 images, tissue probability maps (TPM) of GM, WM and CSF were segmented using the Computational Anatomy Toolbox (CAT12[17]). To facilitate future cross-sectional multi-purpose analyses, the detrended rsfMRI, filtered rsfMRI, whole-brain maps and T1-based TPMs were all spatially normalized into MNI space using co-registration and normalization functions in CAT12. Lastly, the functional connectivity (FC) matrix between predefined WM bundles and GM parcels was computed based on the filtered signals in MNI space using homemade MATLAB functions[7]. Several abovementioned steps in the routine are detailed as below.

To correct slice-dependent delays of BOLD signals that stem from 2D fMRI acquisitions, slice-timing information of fMRI images was hard-coded in the routine (for BLSA) or automatically acquired from JSON files (for ADNI2&3 and OASIS-3) and a summary spreadsheet (for ADNI2&3) provided by the databases. For the fMRI images acquired with multi-band sequences, the slice timing corrections were skipped since the short TRs (0.6–0.7s) in the multi-band acquisition leads to tolerable delays of signals.

In this research, both GM and WM are the compartments of interest, so signals in only CSF are regarded as nuisance regressors. Using a common approach, the CSF region was segmented at the first place and the mean time-course within the CSF mask was extracted and regressed out in a general linear model. Notably, partial volume effects are potentially severe at both the peripheral CSF boundary (adjacent to GM) and on the border of ventricular CSF (adjacent to WM), especially considering that WM signals are more vulnerable to the influence of regression than GM signals due to their lower amplitude, so it is important to conservatively choose the CSF voxels for the regression. To this end, the CSF TMP was thresholded at a high value 0.99 and the resulting mask was eroded iteratively until the volume of ventricular CSF was less than $7cm^3$.

It has been demonstrated that aging and AD populations tend to have ventricular enlargements due to brain tissue atrophy[18], which compromises the accuracy of spatial normalizations especially for the WM around the ventricles. We integrated normalization module in CAT12 into our pipeline to pursue acceptable accuracy and compatibility with our routine.

2.3 Automation for Large-scale Batch Processing

The core preprocessing routine was encapsulated as a singularity container (called a “spider”) and executed within the Distributed Automation for XNAT (DAX[19,20], https://

github.com/VUIIS/dax) service that was triggered by REDCap[21] and automated distribution of computation from the VUIIS-XNAT to the high-performance computing clusters provided by the Vanderbilt Advanced Computing Center for Research and Education (ACCRE). This automation workflow is summarized in Fig. 2. Each job runs the preprocessing routine for one dataset which includes all the rsfMRI images (1–6 runs) acquired for one subject in one visit. The multiple rsfMRI images in the same dataset were processed one by one in the same job.

2.4 Quality Control (QC)

All the results were downloaded from XNAT to a local workstation for quality control (QC). Whether the preprocessing results could be accepted for further analyses depended on the following criteria. First, the core preprocessing routine was fully executed, that is, all the preprocessed results were successfully generated. Second, the translations and rotations of head motion during fMRI scan must be less than 2 mm and 2 degrees, respectively. Third, the spatial normalization was acceptable by an expert's visual inspection.

2.5 Test-Retest Reliability

The datasets with two fMRI scans acquired in the same visit from OASIS-3 were selected for the test-retest experiment (N=1000). The first scan and second scan were preprocessed separately and assigned to the test and retest groups, respectively. We calculated the FC matrix based on the preprocessed rsfMRI data. The FC values were the Pearson's correlation coefficients (CC) between mean time-courses of 48 WM bundles and 82 GM parcels defined by JHU's white matter atlas[22] and PickAtlas[23]. To measure the test-retest reliability, the Pearson's CC and interclass correlation coefficient (ICC)[24] between FCs of test and retest groups were calculated for each WM-GM pair (or WM-WM) pair.

3. RESULTS

Through the automatic pipeline, 5,034 fMRI/T1 scans archived on VUIIS-XNAT were preprocessed. The time of job execution for one rsfMRI scan increases with the number of voxels in a 4D rsfMRI image (Fig. 3).

Figure 4 shows a comparison of the mean WM-GM FC matrices (N=30 from all the three databases) computed based on preprocessed time-courses using pipelines without and with iterative erosions on CSF extraction for nuisance regression-out step. Without erosion, the CSF mask (Fig. 4A) included large areas around the brain where the signals tended to be contaminated by GM signals, which led to over-controlling for the signal of interests, as shown by lower WM-GM FC values (Fig. 4C). The eroded CSF mask contained voxels only totally inside ventricles (Fig. 4B), which gave rise to the correction of the FC matrix (Fig. 4D).

Figure 5 presents an example of spatial normalization accomplished in our pipeline. The significantly enlarged ventricle in the individual space (Fig.5AC) was transformed into MNI space with a large shrinkage (Fig. 5BD).

Based on the QC criteria illustrated in the methods section, approximately 85% preprocessed results were accepted for future WM-related analysis. 0.8% were rejected duo to job failure (no TR, incomplete image, etc.), 14% due to large head motion, and 0.2% due to questionable normalization, wherein only job execution and normalization were directly related to the performance of the proposed pipeline.

Figure 6 exhibits the test-retest reliability. The mean WM-GM (or WM-WM) FC matrix of the test group (Fig. 6A or 6C) appears highly similar to the mean matrix of the retest group (Fig. 6B or 6D). In particular, the Pearson's CC of all the FC values between the two mean matrices is higher than 0.99 ($p << 0.01$). Regarding the rest-retest reliability, Pearson's CC across 1000 datasets ranged from 0.35 to 0.78 (Fig. 6E and 6G) and the ICC ranged from 0.52 to 0.88 (Fig. 6F and 6H).

4. CONCLUSION AND DISCUSSION

This paper describes an automatic pipeline to preprocess rsfMRI images with emphasis on WM BOLD signals in aging populations. The pipeline was executed on 5,034 fMRI scans with high QC acceptance. The test-retest ICC of WM-GM and WM-WM FC is comparable with previously reported ICC of GM-GM FC[25], suggesting our pipeline is highly reliable.

Acknowledgments:

The project is supported by NIH grant 1RF1MH123201 (Gore and Landman), NIH grant R01NS113832 (Gore), Vanderbilt Discovery Grant FF600670 (Gao) and by Intramural Research Program of the National Institute on Aging of the NIH. ADNI was primarily funded by NIH grant U01 AG024904 and DOD award W81XWH-12-2-0012. We also thank the Vanderbilt Advanced Computing Center for Research and Education (ACCRE) and VUIIS-XNAT and DAX team.

REFERENCES

[1]. Wu T-L, Wang F, Li M, Schilling KG, Gao Y, Anderson AW, Chen LM, Ding Z and Gore JC, "Resting-state white matter-cortical connectivity in non-human primate brain," Neuroimage 184 (2019).

[2]. Huang Y, Bailey SK, Wang P, Cutting LE, Gore JC and Ding Z, "Voxel-wise detection of functional networks in white matter," Neuroimage 183, 2018/08/23, 544–552 (2018). [PubMed: 30144573]

[3]. Wu TL, Wang F, Anderson AW, Chen LM, Ding Z and Gore JC, "Effects of anesthesia on resting state BOLD signals in white matter of non-human primates," Magn Reson Imaging 34(9) (2016).

[4]. Zhao Y, Gao Y, Li M, Anderson AW, Ding Z and Gore JC, "Functional parcellation of human brain using localized topo-connectivity mapping," IEEE Trans Med Imaging, 1 (2022).

[5]. Zhao Y, Gao Y, Zu Z, Li M, Schilling KG, Anderson AW, Ding Z and Gore JC, "Detection of functional activity in brain white matter using fiber architecture informed synchrony mapping," Neuroimage 258, 119399 (2022). [PubMed: 35724855]

[6]. Gore JC, Li M, Gao Y, Wu TL, Schilling KG, Huang Y, Mishra A, Newton AT, Rogers BP, Chen LM, Anderson AW and Ding Z, "Functional MRI and resting state connectivity in white matter - a mini-review," Magn Reson Imaging 63 (2019).

[7]. Gao Y, Sengupta A, Li M, Zu Z, Rogers BP, Anderson AW, Ding Z and Gore JC, "Functional connectivity of white matter as a biomarker of cognitive decline in Alzheimer's disease," PLoS One 15(10 October 2020), e0240513 (2020). [PubMed: 33064765]

[8]. Gao Y, Li M, Huang AS, Anderson AW, Ding Z, Heckers SH, Woodward ND and Gore JC, "Lower functional connectivity of white matter during rest and working memory tasks is

associated with cognitive impairments in schizophrenia.," Schizophr Res 233, 101–110 (2021). [PubMed: 34215467]

[9]. Guttmann CRG, Jolesz FA, Kikinis R, Killiany RJ, Moss MB, Sandor T and Albert MS, "White matter changes with normal aging," Neurology 50(4), 972 (1998). [PubMed: 9566381]

[10]. Salat DH, Kaye JA and Janowsky JS, "Prefrontal Gray and White Matter Volumes in Healthy Aging and Alzheimer Disease," Arch Neurol 56(3), 338–344 (1999). [PubMed: 10190825]

[11]. Harrigan RL, Yvernault BC, Boyd BD, Damon SM, Gibney KD, Conrad BN, Phillips NS, Rogers BP, Gao Y and Landman BA, "Vanderbilt University Institute of Imaging Science Center for Computational Imaging XNAT: A multimodal data archive and processing environment," Neuroimage 124 (2016).

[12]. Friston KJ, Williams S, Howard R, Frackowiak RSJ and Turner R, "Movement-related effects in fMRI time-series," Magn Reson Med 35(3) (1996).

[13]. Yu-Feng Z, Yong H, Chao-Zhe Z, Qing-Jiu C, Man-Qiu S, Meng L, Li-Xia T, Tian-Zi J and Yu-Feng W, "Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI," Brain Dev 29(2), 83–91 (2007). [PubMed: 16919409]

[14]. Zou Q-H, Zhu C-Z, Yang Y, Zuo X-N, Long X-Y, Cao Q-J, Wang Y-F and Zang Y-F, "An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF.," J Neurosci Methods 172(1), 137–141 (2008). [PubMed: 18501969]

[15]. Zang Y, Jiang T, Lu Y, He Y and Tian L, "Regional homogeneity approach to fMRI data analysis," Neuroimage 22(1), 394–400 (2004). [PubMed: 15110032]

[16]. Yan C-G, Wang X-D, Zuo X-N and Zang Y-F, "DPABI: Data Processing & Analysis for (Resting-State) Brain Imaging," Neuroinformatics 14(3), 339–351 (2016). [PubMed: 27075850]

[17]. Gaser C and Dahnke R, "CAT – A Computational Anatomy Toolbox for the Analysis of Structural MRI Data," bioRxiv (2022).

[18]. Nestor SM, Rupsingh R, Borrie M, Smith M, Accomazzi V, Wells JL, Fogarty J, Bartha R and Initiative, the A. D. N., "Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database," Brain 131(9), 2443–2454 (2008). [PubMed: 18669512]

[19]. Damon SM, Boyd BD, Plassard AJ, Taylor WD and Landman BA, "DAX - the next generation: towards one million processes on commodity hardware," Medical Imaging (2017).

[20]. Huo Y, Blaber J, Damon SM, Boyd BD, Bao S, Parvathaneni P, Noguera CB, Chaganti S, Nath V, Greer JM, Lyu I, French WR, Newton AT, Rogers BP and Landman BA, "Towards Portable Large-Scale Image Processing with High-Performance Computing," J Digit Imaging 31(3), 304–314 (2018). [PubMed: 29725960]

[21]. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N and Conde JG, "Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support," J Biomed Inform 42(2), 377–381 (2009). [PubMed: 18929686]

[22]. Mori S, Oishi K, Jiang H, Jiang L, Li X, Akhter K, Hua K, Faria A. v, Mahmood A, Woods R, Toga AW, Pike GB, Neto PR, Evans A, Zhang J, Huang H, Miller MI, van Zijl P and Mazziotta J, "Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template," Neuroimage 40(2), 2008/01/03, 570–582 (2008). [PubMed: 18255316]

[23]. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov P. v., Nickerson D, Mikiten SA and Fox PT, "Automated Talairach Atlas labels for functional brain mapping," Hum Brain Mapp 10(3) (2000).

[24]. Noble S, Scheinost D and Constable RT, "A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis," Neuroimage 203, 116157 (2019). [PubMed: 31494250]

[25]. Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME and Prabhakaran V, "The effect of scan length on the reliability of resting-state fMRI connectivity estimates," Neuroimage 83, 550–558 (2013). [PubMed: 23747458]

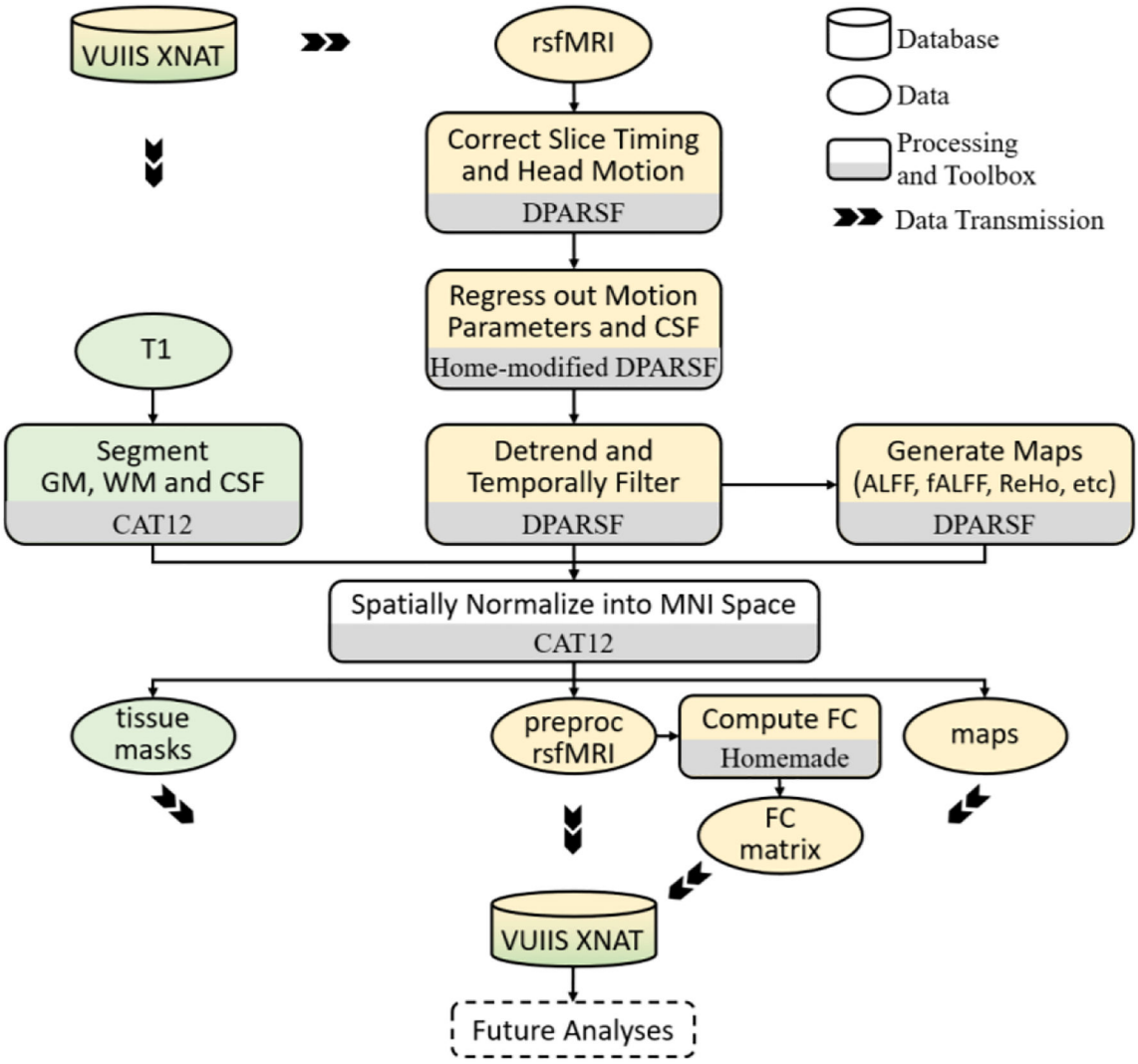

Fig. 1.
Workflow of rsfMRI/T1 image preprocessing routine.

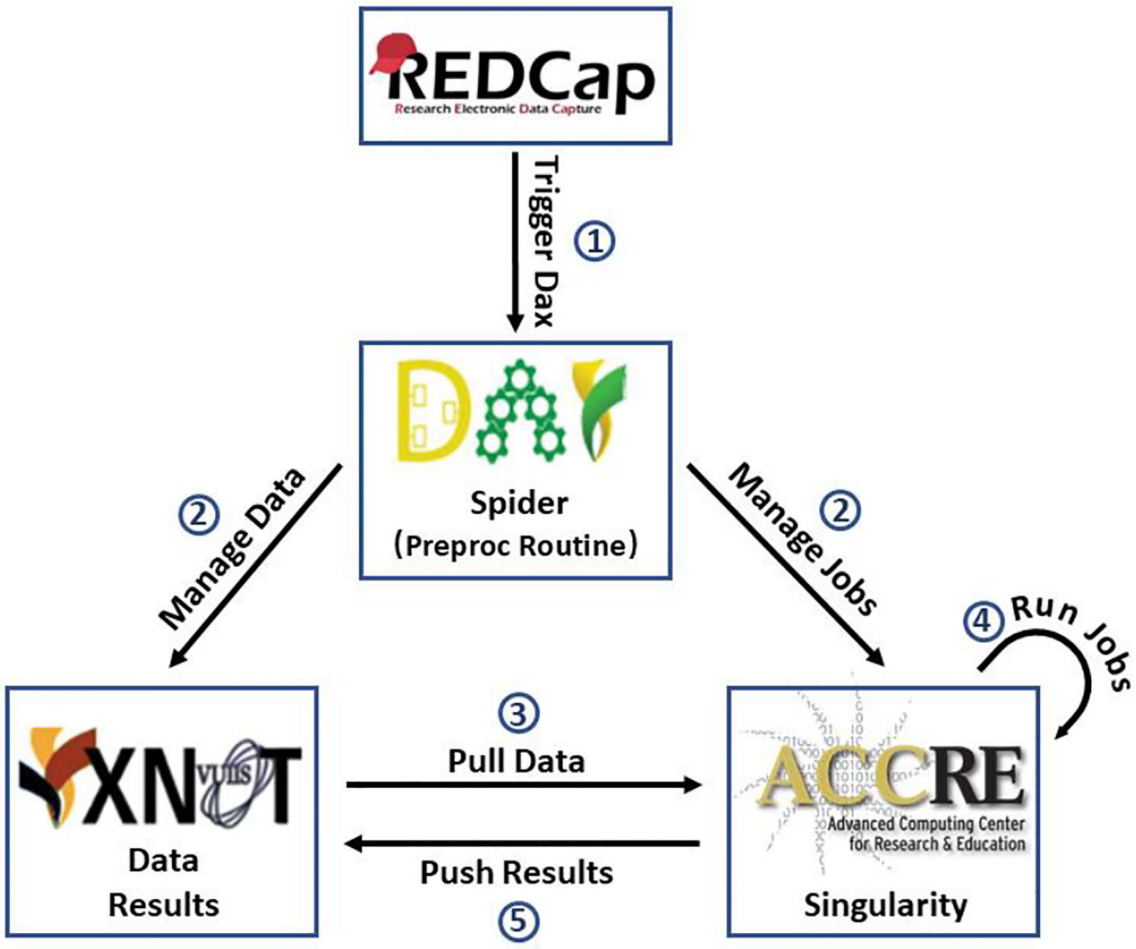

Fig. 2.

Workflow of automatic pipeline for batch preprocessing images. The circled numbers in the workflow chart indicate the temporal order of the operations.

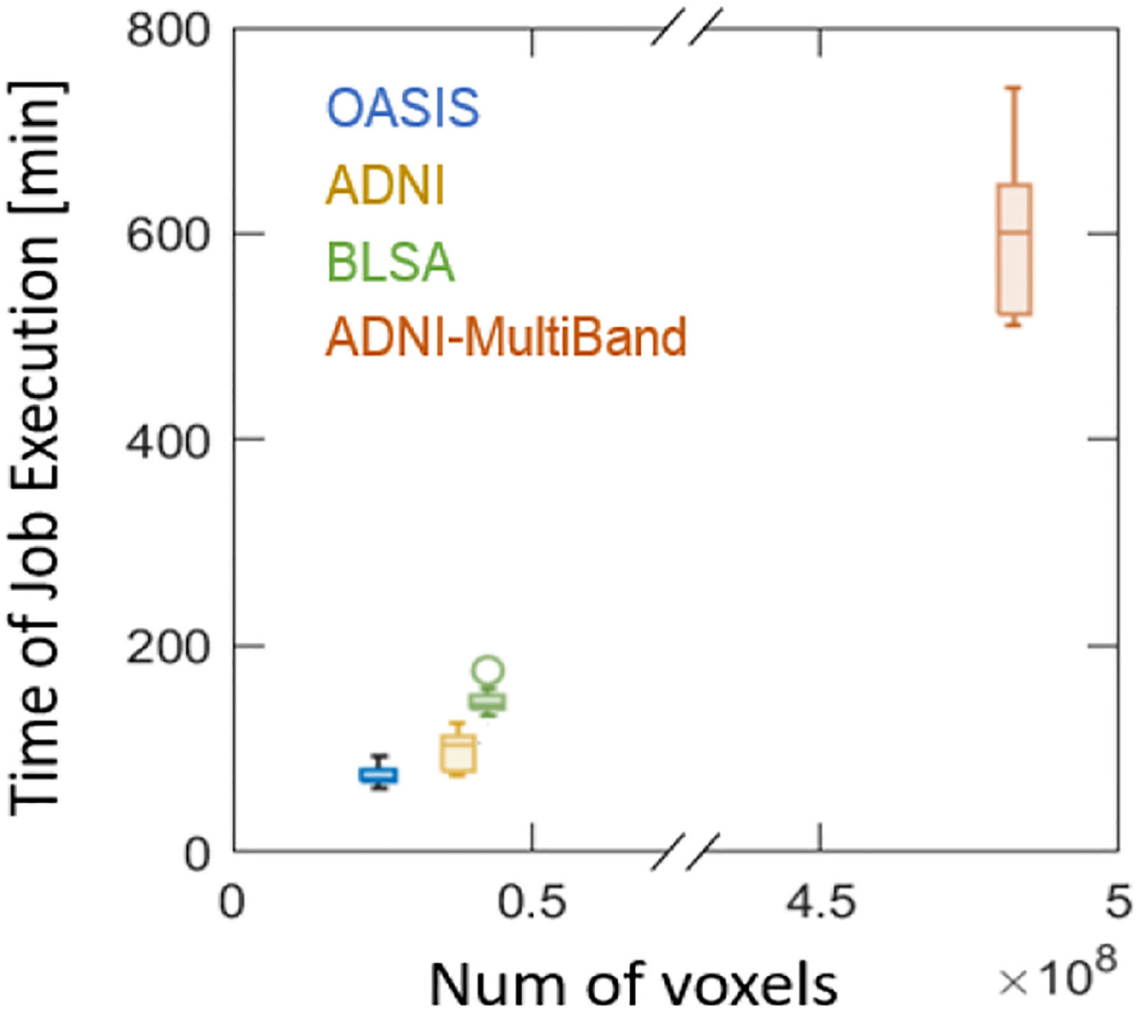

Fig. 3.
Time of job execution vs. numbers of voxels in fMRI image (N=40). Each box included 10 datasets that were randomly selected from one of the 4 groups.

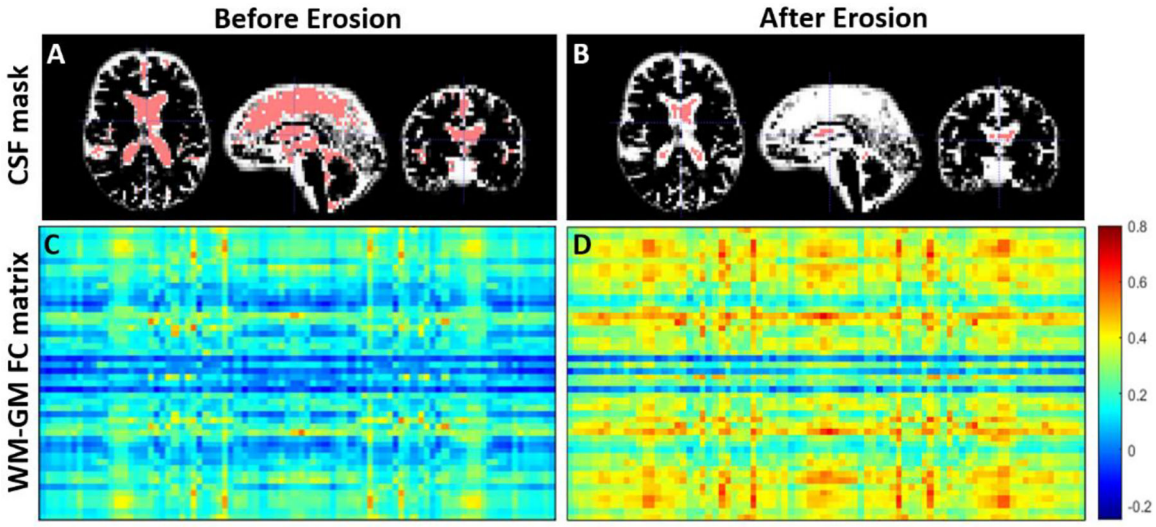

Fig. 4.

CSF masks (in red; overlaid on CSF tissue possibility maps) obtained before (A) and after (B) iterative erosions for one subject in individual space and the mean WM-GM functional connectivity matrices (N=30) calculated based on the preprocessed time-courses using pre- and post-eroded CSF masks in nuisance regression step (C, D).

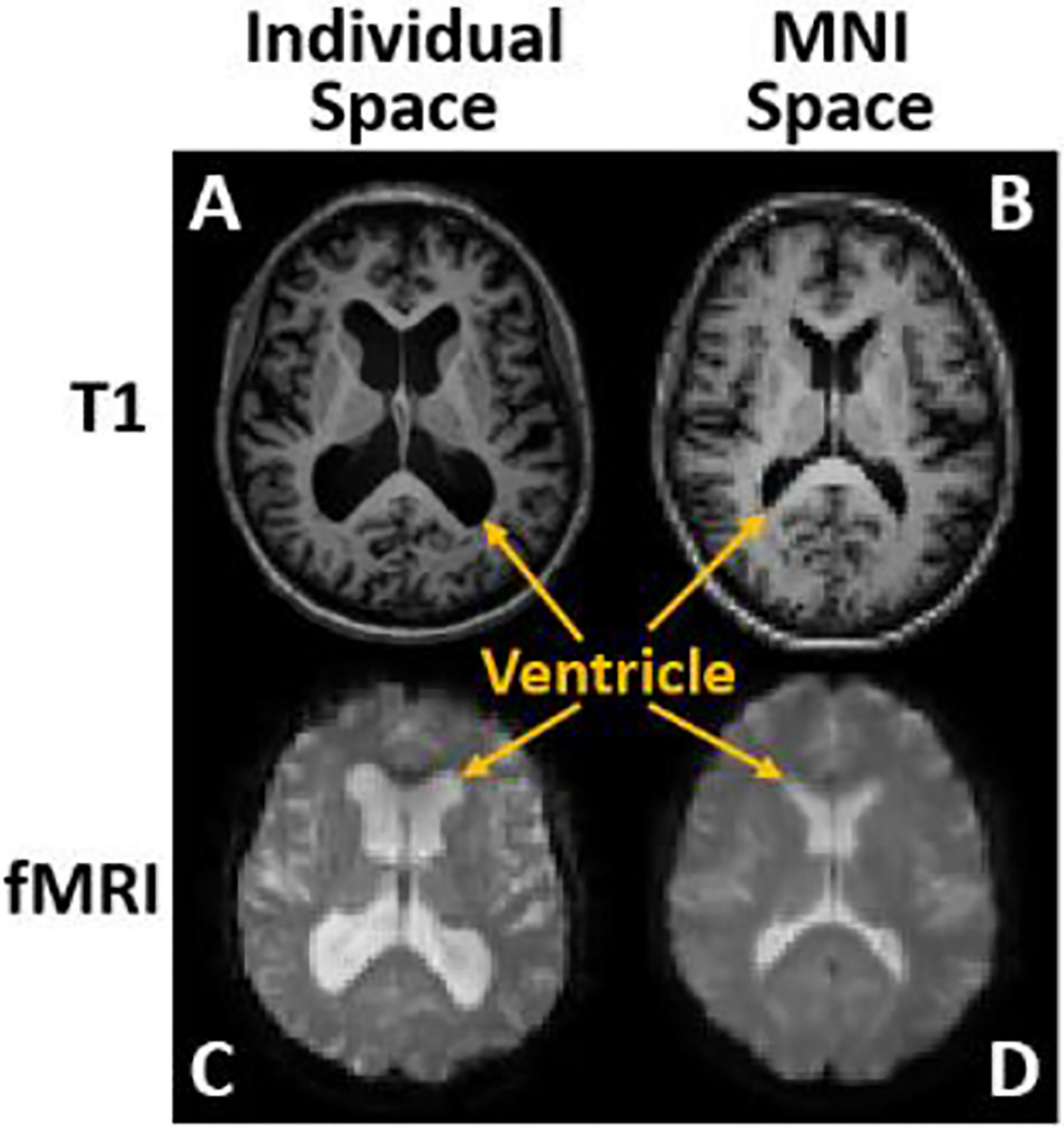

Fig. 5.

One example of spatial normalization result. (A, C) T1 and fMRI image of an elderly subject in the individual space; (B, D) Normalized T1 and fMRI in MNI space. The originally enlarged ventricle shrank properly during spatial normalization using CAT12.

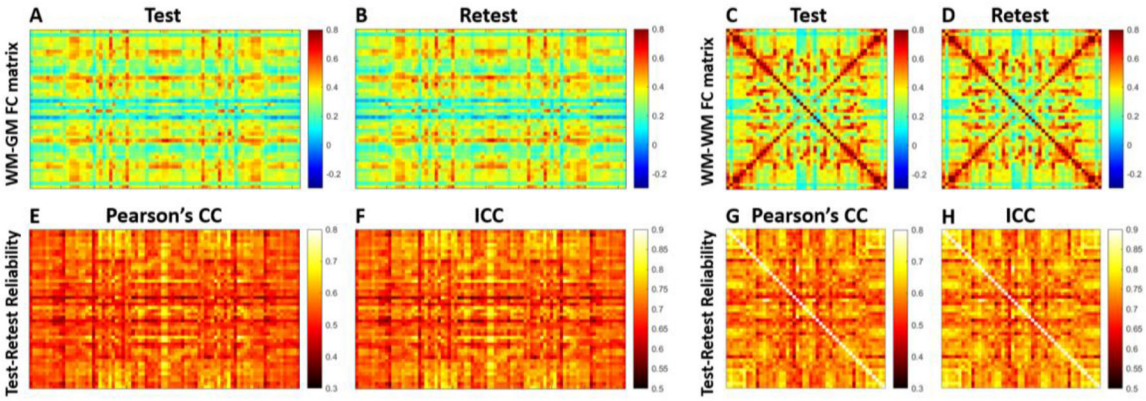

Fig. 6.

Mean WM-GM and WM-WM functional connectivity matrices of test and retest groups (A-D), and the test-retest reliability measures: Pearson's correlation coefficient (E, G) and intraclass correlation coefficient (F, H) (N=1000).