

# DeepPrep: an accelerated, scalable and robust pipeline for neuroimaging preprocessing empowered by deep learning

Received: 2 April 2024

Accepted: 10 January 2025

Published online: 06 February 2025

 Check for updates

Jianxun Ren <sup>1,11</sup>✉, Ning An <sup>1,11</sup>, Cong Lin<sup>1</sup>, Youjia Zhang<sup>1</sup>, Zhenyu Sun<sup>1</sup>, Wei Zhang<sup>1,2</sup>, Shiyi Li<sup>1</sup>, Ning Guo<sup>1</sup>, Weigang Cui<sup>1</sup>, Qingyu Hu <sup>1</sup>, Weiwei Wang<sup>1</sup>, Xuehai Wu<sup>3</sup>, Yinyan Wang<sup>4,5</sup>, Tao Jiang <sup>4,5</sup>, Theodore D. Satterthwaite <sup>6,7,8</sup>, Danhong Wang<sup>9</sup> & Hesheng Liu <sup>1,10</sup>✉

Neuroimaging has entered the era of big data. However, the advancement of preprocessing pipelines falls behind the rapid expansion of data volume, causing substantial computational challenges. Here we present DeepPrep, a pipeline empowered by deep learning and a workflow manager. Evaluated on over 55,000 scans, DeepPrep demonstrates tenfold acceleration, scalability and robustness compared to the state-of-the-art pipeline, thereby meeting the scalability requirements of neuroimaging.

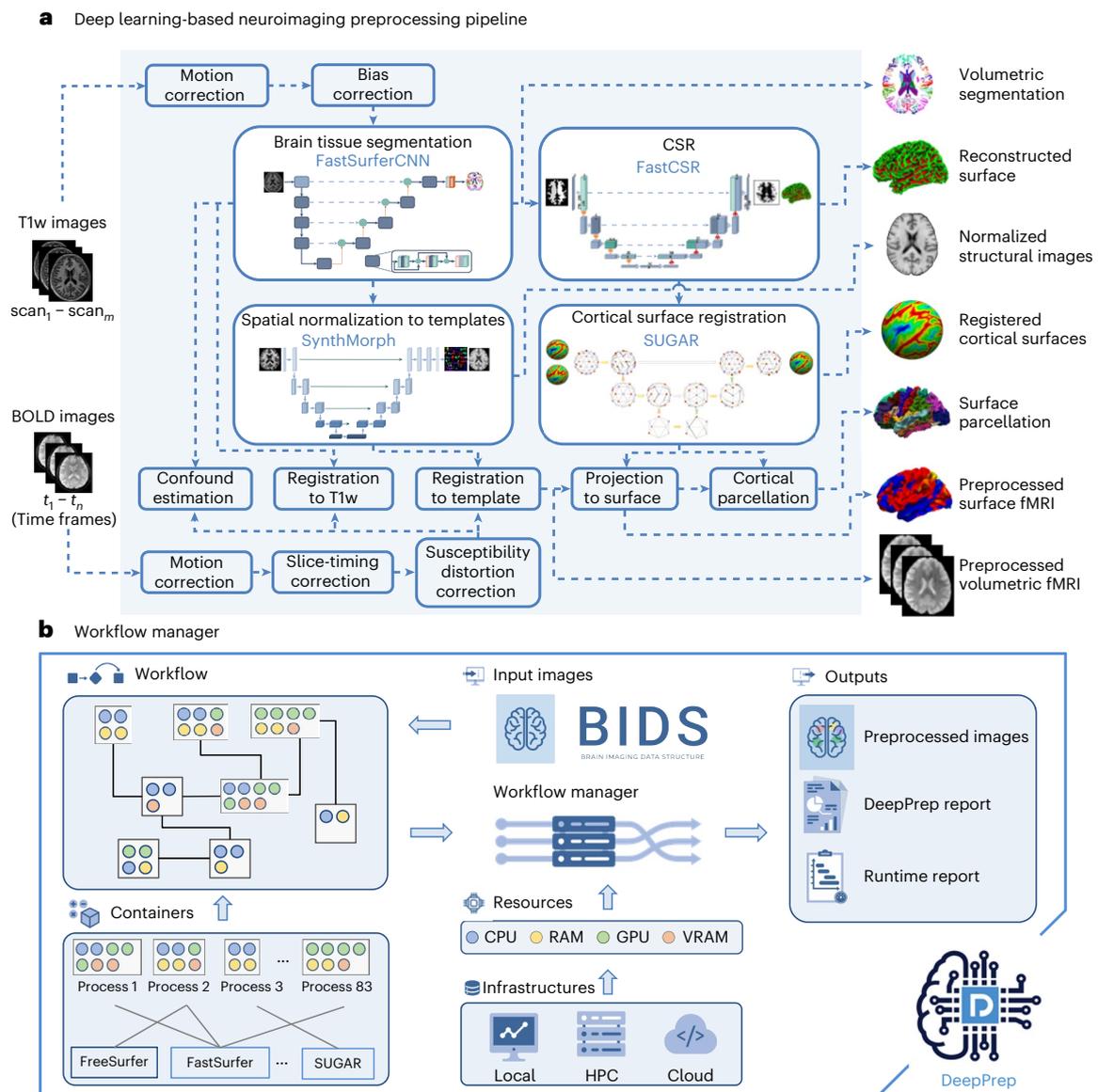
Advances and transparency in neuroimaging are propelled by the rapidly growing amounts of publicly available data from large-scale projects (for example, the UK Biobank (UKBB) with 50,000+ scans<sup>1</sup>), data-sharing initiatives (for example, the OpenNeuro<sup>2</sup>) and international consortia (for example, the ENIGMA<sup>3</sup>) (Extended Data Fig. 1). Neuroimaging data typically require complex and multistage preprocessing pipelines to enable accurate brain tissue segmentation, spatial normalization and other essential preprocessing steps. While prevailing preprocessing pipelines such as FreeSurfer<sup>4</sup>, fMRIPrep<sup>5</sup>, QSIprep<sup>6</sup> and ASLprep<sup>7</sup> have led to numerous important findings, their original design for relatively small sample sizes makes it difficult for them to meet scalability demands in the era of big data. Concurrently, clinical applications of neuroimaging, such as imaging-guided neuromodulation<sup>8</sup>, require fast turnaround time and robustness at the individual level. This becomes particularly challenging when handling patients with brain distortion induced by traumas, gliomas or strokes<sup>9</sup>. To fulfill these emerging requirements, a computationally efficient, scalable and robust preprocessing pipeline is needed.

Hence, we developed DeepPrep, an efficient, scalable and robust preprocessing pipeline for neuroimaging powered by deep learning

algorithms and a workflow manager. Deep learning-based algorithms for neuroimaging have shown potential in improving computational efficiency and robustness<sup>10–13</sup>, but an end-to-end pipeline empowered by these algorithms is still lacking. Workflow managers are widely used to manage complex workflows for big data processing in various fields, such as bioinformatics, due to their scalability, portability and computational resource efficiency<sup>14</sup>. However, this has not been fully exploited in the field of neuroimaging. Here, we applied DeepPrep to structural and functional magnetic resonance imaging (MRI). We integrated multiple learning-based modules, including FastCSR<sup>11</sup>, spherical ultrafast graph attention framework for cortical surface registration (SUGAR)<sup>12</sup>, SynthMorph<sup>13</sup> and FastSurferCNN<sup>10</sup> (see Fig. 1a and Supplementary Table 1 for all dependencies). These deep learning modules facilitate computational efficiency in typically time-consuming operations, such as cortical surface reconstruction (CSR), surface registration, anatomical segmentation and volumetric spatial normalization. All software modules are linked by 83 discrete yet interdependent task processes, which are packaged into a Docker or Singularity container along with all dependencies<sup>15</sup>. Our incorporation of Nextflow, a reproducible, scalable and portable workflow manager<sup>14</sup>, enables our pipeline to

<sup>1</sup>Changping Laboratory, Beijing, China. <sup>2</sup>Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. <sup>3</sup>Department of Neurosurgery, Huashan Hospital, Fudan University, Shanghai, China. <sup>4</sup>Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China. <sup>5</sup>Beijing Neurosurgical Institute, Capital Medical University, Beijing, China. <sup>6</sup>Penn Lifespan Informatics and Neuroimaging Center, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>7</sup>Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>8</sup>Lifespan Brain Institute, Children's Hospital of Philadelphia, Philadelphia, PA, USA. <sup>9</sup>Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA, USA. <sup>10</sup>Biomedical Pioneering Innovation Center, Peking University, Beijing, China. <sup>11</sup>These authors contributed equally: Jianxun Ren, Ning An.

✉e-mail: [jianxun.ren@cpl.ac.cn](mailto:jianxun.ren@cpl.ac.cn); [liuhesheng@cpl.ac.cn](mailto:liuhesheng@cpl.ac.cn)



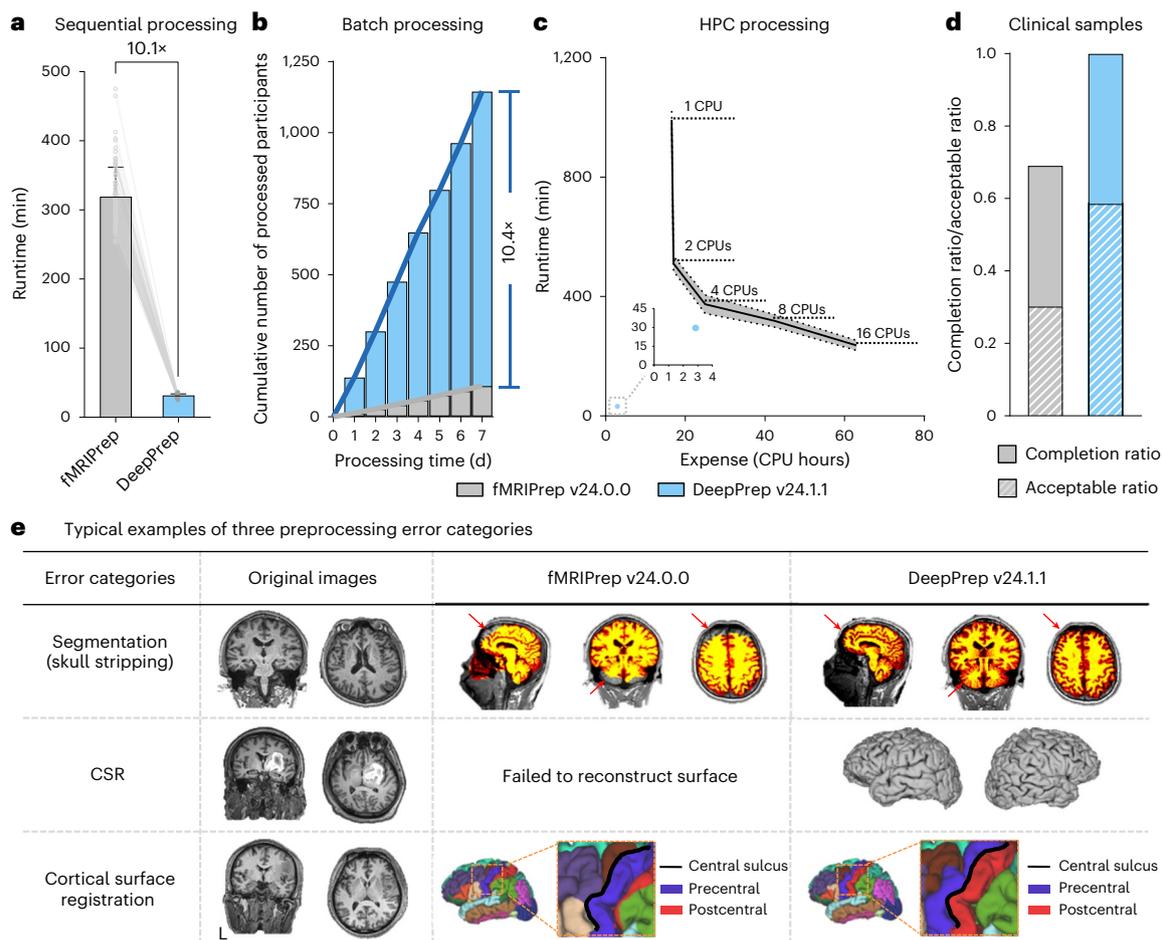
**Fig. 1 | A computationally efficient and scalable neuroimaging pipeline is empowered by deep learning algorithms and the workflow manager. a**, DeepPrep workflow. DeepPrep accepts both anatomical and fMRI data for preprocessing. White boxes highlight deep learning algorithms for brain tissue segmentation, CSR, cortical surface registration and volumetric spatial normalization, which replace the most time-intensive modules in conventional pipelines. T1w,  $T_1$  weighted. BOLD, blood oxygenation level dependent.

**b**, The preprocessing pipeline is organized into multiple relatively independent yet integrated processes. Imaging data adhering to the BIDS format are preprocessed through a structured workflow managed by Nextflow<sup>14</sup>. Nextflow schedules task processes and allocates computational resources across diverse infrastructures, including local computers, HPC clusters and cloud computing environments. The pipeline yields standardized preprocessed imaging and derivative files, quality control reports and runtime reports as its outputs.

maximize computational resource utilization through dynamically scheduling parallelization. The workflow manager also makes it convenient to deploy the pipeline in diverse computing environments, including local computers, high-performance computing (HPC) and cloud computing (Fig. 1b). Importantly, DeepPrep is a brain imaging data structure (BIDS) app that can automatically configure appropriate preprocessing workflows based on the metadata stored in the BIDS layout<sup>16</sup>. In addition to providing preprocessed structural and functional data (see Supplementary Fig. 1 for an example), DeepPrep also generates a visual report for each participant and a summary report for a group of participants to facilitate data quality assessments by adapting from MRIQC<sup>17</sup> (Supplementary Fig. 2) as well as a detailed runtime report (Supplementary Fig. 3). Documentation (<https://deep-prep.readthedocs.io>) and code base (<https://github.com/pBFSLab/DeepPrep>) of DeepPrep is version controlled and actively updated,

enabling users and developers to participate in further development of the pipeline.

To demonstrate the performance of DeepPrep, we applied it to over 55,000 scans from seven datasets with diverse populations, scanners and imaging parameters (Supplementary Note 1 and Supplementary Table 2). These datasets are distinct from those used for training deep learning models (Supplementary Table 3). These evaluation datasets include structural MRI and functional MRI (fMRI) from the UKBB<sup>1</sup> for evaluating computational efficiency and scalability, one manually labeled brain dataset (the Mindboggle-101 dataset<sup>18</sup>) for evaluating brain segmentation, two precision neuroimaging datasets (the Midnight Scanning Club (MSC) dataset<sup>19</sup> and the Consortium for Reliability and Reproducibility Hangzhou Normal University (CoRR-HNU) dataset<sup>20</sup>) for accuracy and reliability assessment and three clinical datasets for robustness assessment. These data were also



**Fig. 2 | Performance of DeepPrep in speed and robustness.** **a**, Runtimes for fMRIPrep and DeepPrep when processing each participant's data sequentially on a local workstation ( $n = 100$  participants from the UKBB dataset). Data are presented as mean  $\pm$  s.d. Error bars represent s.d. **b**, Number of participants that can be processed by fMRIPrep and DeepPrep at the indicated time intervals.

**c**, Trade-off curve between preprocessing time and hardware expense measured by CPU hours. Data are presented as mean  $\pm$  s.d. **d**, Completion and acceptable ratios when preprocessing 53 clinical samples that could not be processed by FreeSurfer. **e**, Preprocessing errors are categorized into three types, with three representative clinical samples shown.

preprocessed using the state-of-the-art processing pipeline fMRIPrep<sup>5</sup> for comparison.

In a local workstation equipped with central processing units (CPUs) and a graphics processing unit (GPU), DeepPrep preprocessed single participants from the UKBB dataset in  $31.6 \pm 2.4$  min (mean  $\pm$  s.d.) which was 10.1 times faster than fMRIPrep (Fig. 2a;  $318.9 \pm 43.2$  min, two-tailed paired  $t$ -test,  $t_{(99)} = 67.0$ ,  $P = 2.6 \times 10^{-84}$ ). DeepPrep demonstrated efficient batch-processing capability, preprocessing 1,146 participants per week (average processing time of 8.8 min per participant), which was 10.4 times more efficient than fMRIPrep, which processed 110 participants in the same time frame (Fig. 2b). Moreover, when we performed separate preprocessing for anatomical and functional scans, DeepPrep exhibited acceleration factors of 13.3 and 7.7 times compared to those of fMRIPrep, respectively (Extended Data Fig. 2). In the HPC environment, the trade-off between preprocessing time and computational resource utilization, measured in terms of CPU hours and associated expenses, becomes a critical consideration. Achieving shorter preprocessing time by recruiting more CPUs comes with higher expense on hardware (Fig. 2c), which can be a substantial concern for users. To explore this trade-off relationship, we conducted an analysis by recruiting one, two, four, eight and 16 CPUs for processing a participant. fMRIPrep exhibited a characteristic curve, illustrating the trade-off between processing time and expense (Fig. 2c). By contrast, DeepPrep demonstrated stability in both processing time and expenses due to its computational flexibility in dynamically

allocating computational resources to match the specific requirements of individual task processes. Importantly, the computational expenses associated with DeepPrep were found to be at least 5.8 times and up to 22.1 times lower than those of fMRIPrep (Fig. 2c). Additionally, we demonstrated the scalability of DeepPrep by preprocessing the entire UKBB neuroimaging dataset, consisting of over 54,515 scans, within 6.5 d in an HPC cluster.

The DeepPrep and fMRIPrep outputs were also compared in the Mindboggle-101, MSC and CoRR-HNU datasets, covering a range of aspects including anatomical parcellation and segmentation (Extended Data Fig. 3), anatomical morphometrics (Extended Data Fig. 4), temporal signal-to-noise ratio (tSNR) (Extended Data Fig. 5), spatial normalization (Extended Data Fig. 6), task-evoked responses (Extended Data Fig. 7), functional connectivity (Extended Data Fig. 8) and test-retest reliability in functional connectomes (Extended Data Fig. 9). Overall, these comparisons indicated that DeepPrep consistently yielded preprocessing results that were either similar or superior to those generated by fMRIPrep when assessed using various metrics. These results support the accuracy of DeepPrep while concurrently maximizing efficiency.

Next, we screened 53 clinical samples from three clinical datasets that could not be successfully processed by FreeSurfer within 48 h due to distorted brains or imaging noises. These samples were preprocessed using both DeepPrep and fMRIPrep. DeepPrep exhibited a substantially higher pipeline completion ratio (100.0%), a higher

acceptable ratio (58.5%; see Methods for definitions of completion and acceptable ratios) and shorter processing time for each participant ( $27.4 \pm 5.2$  min) than fMRIPrep (Fig. 2d; completion ratio of 69.8%,  $\chi^2$  test:  $\chi^2 = 16.6$ ,  $P = 4.7 \times 10^{-5}$ ; accuracy ratio of 30.2%,  $\chi^2 = 8.6$ ,  $P = 0.003$ ; processing time of  $369.6 \pm 205.0$  min, two-tailed paired  $t$ -test,  $t_{(36)} = 10.3$ ,  $P = 3.0 \times 10^{-12}$ ). The occurrences of fMRIPrep preprocessing failures and errors could be attributed to three main causes: segmentation errors, surface reconstruction failures and surface registration errors (Fig. 2e). These causes correspond to processing modules where deep learning algorithms replace conventional algorithms, indicating the robustness of DeepPrep in handling complicated clinical samples.

In summary, DeepPrep demonstrates high efficiency and robustness in processing large-scale neuroimaging datasets and complex clinical samples. This success can be attributed to the integration of workflow managers for handling big data and the utilization of deep learning algorithms. With time, we plan to expand DeepPrep into a comprehensive platform for processing multimodal neuroimaging. While the current version focuses primarily on the most time-consuming workflows in structural MRI and fMRI, future versions will integrate additional modalities, such as arterial spin labeling and diffusion imaging. These enhancements will benefit the broader neuroimaging community.

### Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-025-02599-1>.

### References

1. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
2. Markiewicz, C. J. et al. The OpenNeuro resource for sharing of neuroscience data. *eLife* **10**, e71774 (2021).
3. Bearden, C. E. & Thompson, P. M. Emerging global initiatives in neurogenetics: the Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) Consortium. *Neuron* **94**, 232–236 (2017).
4. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
5. Esteban, O. et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
6. Cieslak, M. et al. QSIPrep: an integrative platform for preprocessing and reconstructing diffusion MRI data. *Nat. Methods* **18**, 775–778 (2021).
7. Adebimpe, A. et al. ASLPrep: a platform for processing of arterial spin labeled MRI and quantification of regional brain perfusion. *Nat. Methods* **19**, 683–686 (2022).
8. Ren, J. et al. Personalized functional imaging-guided rTMS on the superior frontal gyrus for post-stroke aphasia: a randomized sham-controlled trial. *Brain Stimul.* **16**, 1313–1321 (2023).
9. Siegel, J. S., Shulman, G. L. & Corbetta, M. Measuring functional connectivity in stroke: approaches and considerations. *J. Cereb. Blood Flow Metab.* **37**, 2665–2678 (2017).
10. Henschel, L. et al. FastSurfer — a fast and accurate deep learning based neuroimaging pipeline. *Neuroimage* **219**, 117012 (2020).
11. Ren, J. et al. Fast cortical surface reconstruction from MRI using deep learning. *Brain Inform.* **9**, 6 (2022).
12. Ren, J. et al. SUGAR: spherical ultrafast graph attention framework for cortical surface registration. *Med. Image Anal.* **94**, 103122 (2024).
13. Hoffmann, M., Hoopes, A., Greve, D. N., Fischl, B. & Dalca, A. V. Anatomy-aware and acquisition-agnostic joint registration with SynthMorph. *Imaging Neurosci.* **2**, 1–33 (2024).
14. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
15. Kurtzer, G. M., Sochat, V. & Bauer, M. W. Singularity: scientific containers for mobility of compute. *PLoS ONE* **12**, e0177459 (2017).
16. Gorgolewski, K. J. et al. BIDS apps: improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput. Biol.* **13**, e1005209 (2017).
17. Esteban, O. et al. MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* **12**, e0184661 (2017).
18. Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* **6**, 171 (2012).
19. Gordon, E. M. et al. Precision functional mapping of individual human brains. *Neuron* **95**, 791–807 (2017).
20. Zuo, X. N. et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* **1**, 140049 (2014).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

## Methods

DeepPrep offers a computationally efficient, robust and scalable neuroimaging preprocessing pipeline, incorporating recent advancements in deep learning for neuroimaging processing and the workflow manager. To achieve optimal acceleration, it is recommended to deploy DeepPrep in a computing environment equipped with GPUs, although it remains executable on CPU-only systems (see Supplementary Note 2 for minimum system requirements). Additionally, version control of the project's codes and documents is managed through GitHub.

### Preprocessing workflow for anatomical images

**Overall description.** The anatomical preprocessing workflow in DeepPrep closely follows the FreeSurfer pipeline while efficiently replacing several of the most time-consuming steps with deep learning algorithms. Specifically, volumetric anatomical segmentation, CSR and cortical surface registration are accomplished using FastSurferCNN<sup>10</sup>, FastCSR<sup>11</sup> and SUGAR<sup>12</sup>, respectively. FastSurferCNN is a deep learning architecture designed for anatomical segmentation of brain tissues, enabling computationally efficient segmentation of the entire brain into 95 distinct cortical and subcortical regions<sup>10</sup>. FastCSR is a deep learning model designed for the time-consuming task of CSR through the utilization of implicit representations of cortical surfaces<sup>11</sup>. SUGAR is a deep learning framework designed for both rigid and nonrigid cortical surface registration<sup>12</sup>. Detailed information on the training and test procedures for each algorithm is provided in Supplementary Note 3. The remaining steps in the workflow remain consistent with fMRIPrep, including spherical mapping, morphometric estimation and statistics. These steps ensure the continued reliability and accuracy of the overall process while harnessing the benefits of deep learning algorithms to enhance computational efficiency. The preprocessing workflow consists of several essential steps as follows:

**Motion correction.** If multiple T1w images are available for each participant or each session, FreeSurfer's command 'recon-all -motioncor' is employed to correct head motions across the scans. This process yields an average T1w image to minimize the impact of head motions on data quality.

**Segmentations.** The whole brain is segmented into 62 cortical and 33 subcortical regions using FastSurferCNN<sup>10</sup>, a deep learning model for accurate and rapid anatomical segmentations. In addition, FreeSurfer's command 'mri\_cc' is used to segment the corpus callosum into five subcortical regions, yielding a total of 38 subcortical regions.

**Skull stripping and bias field correction.** A brain mask is generated through the combination of all these segmented brain regions to achieve accurate and robust skull stripping. The T1w images undergo N4 bias field correction using SimpleITK with reference to a brain mask. The normalized and skull-stripped T1w images were then fed into the subsequent steps.

**Cortical surface reconstruction.** The white matter and pial cortical surfaces are reconstructed based on the anatomical segmentation. This process uses FastCSR<sup>11</sup>, a deep learning-based model designed to accelerate CSR. FastCSR leverages an implicit representation of the cortical surface through the level-set representation and uses a topology-preserving surface extraction method to yield white and pial surfaces represented by triangulated meshes.

**Cortical surface registration.** The reconstructed surface is inflated to a sphere with minimal distortion using the FreeSurfer command 'mris\_sphere'. Both rigid and nonrigid registrations for the spherical cortical surface are performed to align anatomical landmarks and morphometrics, facilitated by the deep learning cortical surface registration framework SUGAR<sup>12</sup>. Individual spherical surfaces are aligned to the fsaverage template surfaces by default.

**Cortical surface parcellation.** The cortical surface parcellation is generated based on the cortical surface registration using the FreeSurfer command 'recon-all -cortparc', yielding 68 cortical regions. Subsequently, the cortical parcellation is projected to the volumetric segmentation by assigning voxels their closest cortical labels using the command 'mri\_surf2volseg', thereby replacing the cortical parcellation derived from FastSurferCNN.

### Preprocessing workflow for functional images

**Overall description.** The functional preprocessing workflow in DeepPrep incorporates advanced registration methods, SynthMorph<sup>13,21</sup>, to replace the most time-consuming step, spatial normalization. SynthMorph is a deep learning-based approach tailored for volumetric registration, also known as spatial normalization, aiming to achieve a robust, accurate and computationally efficient alignment of previously unseen brain images. The workflow is also complemented by modules from existing tools, including AFNI<sup>22</sup>, FSL<sup>23</sup> and fMRIPrep<sup>5</sup>, to form a comprehensive functional image-preprocessing method. The fMRI preprocessing workflow consists of several essential steps as follows:

**Motion correction and slice-timing correction.** The head motion parameters of the BOLD fMRI signals are estimated with MCFLIRT from FSL<sup>24</sup>, with the reference volume generated from averaging nonsteady state volumes or dummy scans. Slice-timing correction is included in our processing pipeline for fMRI data using 3dTshift from AFNI<sup>22</sup> when slice-timing information is available in the BIDS metadata. This is an optional step and can be deactivated if the BIDS metadata does not specify slice times.

**Susceptibility distortion correction.** DeepPrep incorporates susceptibility distortion correction workflows (SDCflows)<sup>25</sup> to correct susceptibility distortions. SDCflows offers versatile workflows designed to preprocess various MRI schemes, enabling the estimation of  $B_0$  field inhomogeneity maps directly associated with distortion. This distortion correction is applied to the fMRI data when the appropriate field map information is available within the BIDS metadata. Distortion correction is an optional step.

**Co-registration.** A rigid registration is performed using FreeSurfer's boundary-based registration to align motion-corrected fMRI volumes to native T1w images for each participant. The registration optimizes the boundary-based loss function to align the boundary between gray and white matter across different imaging modalities.

**Spatial normalization.** The spatial normalization step aims to normalize individual brain images to a standard template, such as the MNI152NLin6Asym volumetric template and FreeSurfer's fsaverage6 surface template. Traditionally, this step is highly time-consuming due to the requirement of nonrigid surface and volumetric registration, often taking hours for computations. However, in DeepPrep, a deep learning algorithm called SynthMorph<sup>13,21</sup> is used to achieve both rigid and nonrigid volumetric registration within 1 min. Additionally, for both rigid and nonrigid cortical surface registration, SUGAR<sup>12</sup> is used to achieve accurate alignment in anatomical landmarks and morphometrics in seconds. Subsequently, preprocessed BOLD fMRI volumes are projected to the MNI152NLin6Asym template and fsaverage6 template surfaces by default by applying deformation matrices derived from the registrations. The pipeline also flexibly supports normalization to other volumetric templates of human brains managed by TemplateFlow<sup>26</sup>.

In summary, this preprocessing workflow uses a combination of conventional methods and advanced deep learning algorithms to efficiently and accurately preprocess structural and functional images for neuroimaging analysis.

### The workflow manager and containers

In DeepPrep, we used Nextflow<sup>14,27</sup> (<https://www.nextflow.io>), an open-source workflow manager widely used in life sciences, and Docker (<https://www.docker.com>) and Singularity<sup>15</sup> containers to establish a scalable, portable and reproducible neuroimaging preprocessing pipeline across diverse computing environments.

Nextflow enhances DeepPrep by optimizing computational resource utilization, especially in HPC clusters and cloud infrastructures, while ensuring cross-platform portability and reproducibility. Specifically, DeepPrep established and executed the pipeline using key Nextflow components, including the process, channel, workflow and executor, with some modifications aimed at improving GPU utilization efficiency. Initially, we disassembled the pipeline into 83 discrete minimal executable steps, enabling parallelization and substantial improvements in computational efficiency. The disassembly also allowed for precise monitoring and control of system resource consumption for each individual step. Each step was defined as a Nextflow process, ensuring that input and output data adhered to expected standards and specifying the system resource requirements for each process. Data flow pipelines were delineated using the Nextflow channel component, and the complete workflow was specified through the Nextflow workflow component, integrating predefined processes and data flow pipelines.

Once the pipeline was established, DeepPrep employed the Nextflow executor component, serving as an abstraction layer between the processes and the underlying execution system. This abstraction facilitated the portable execution of the pipeline across diverse platforms and infrastructures. However, in cases where the execution system was set to 'local', the existing version of Nextflow lacked GPU resource-monitoring capabilities. Given the benefits in computational efficiency of DeepPrep from multiple GPU-intensive processes, concurrent execution on a single GPU-equipped machine could result in errors due to insufficient GPU video memory (VRAM). To address this challenge, we devised a GPU scheduling module, using two distributed mutual exclusion locks based on Redis (<https://redis.io>). This module categorized GPU processes into two groups: those consuming less than half of the VRAM and those requiring more than half. Processes with lower VRAM requirements could proceed upon acquiring a single lock, while those with higher demands needed to secure both locks simultaneously. This module effectively fixed GPU-related resource conflicts and substantially improved GPU utilization efficiency before Nextflow's GPU-monitoring implementation.

In addition, we prioritized the reproducibility of our preprocessing pipeline. To this end, DeepPrep underwent rigorous version control and was fully compatible with execution in both Docker and Singularity containers. Singularity, designed for scientific computational tasks such as neuroimaging processing, offered the flexibility to execute across a spectrum of computing infrastructures and platforms, ranging from local machines to HPC clusters and cloud environments. Notably, Singularity emphasized security, a particular concern in scientific computing, particularly in HPC settings. It operated within a user space model, obviating the need for elevated privileges, thereby substantially reducing security risks typically associated with superuser access.

### DeepPrep report

DeepPrep automatically generates a descriptive HTML report for each participant and session (Supplementary Fig. 2). The report commences with a concise summary of key imaging parameters extracted from the BIDS meta-information. Subsequently, the report provides an overview of the overall CPU and GPU processing times for data preprocessing. Key processing steps and results for structural images are visually presented, including segmentation, parcellation, spatial normalization and co-registration. The normalization and co-registration outcomes are demonstrated through dynamic 'before' versus 'after' animations.

Additionally, the report includes a carpet plot, showcasing both the raw and preprocessed fMRI data, along with a tSNR map. Finally, the report concludes with a comprehensive boilerplate method text, offering a clear and consistent description of all preprocessing steps employed, accompanied by appropriate citations.

### Standard output

The preprocessed structural MRI data are organized to align with the results of FreeSurfer, encompassing the normalized and skull-stripped brain, reconstructed cortical surfaces and morphometrics, volumetric segmentation, cortical surface parcellation and their corresponding statistics. Additionally, transformation files for surface spherical registration are included, as illustrated in Supplementary Fig. 1, depicting the data structure. For the preprocessed fMRI data, the naming adheres to the BIDS specification for derived data. The default output spaces for the preprocessed fMRI consist of three options: (1) the native BOLD fMRI space, (2) the MNI152NLin6Asym space and (3) the fsaverage6 surface space. However, users have the flexibility to specify other output spaces, including the native T1w space and various volumetric and surface templates available on TemplateFlow. The main outputs of the preprocessed data include:

1. Preprocessed fMRI data
2. Reference volume for motion correction
3. Brain masks in the BOLD native space, including nuisance masks, such as ventricle and white matter masks
4. Transformation files for between T1w and the fMRI reference and between T1w and the standard templates
5. Head motion parameters and the tSNR map
6. Confound matrix.

### Evaluation data

We employed over 55,000 scans from seven datasets<sup>1,8,18–20,28,29</sup> to evaluate the performance of DeepPrep (Supplementary Table 2). The evaluation datasets are distinct from training datasets for deep learning models to avoid information leakage. The training datasets for these models are listed in Supplementary Table 3. A large-scale dataset (the UKBB dataset<sup>1</sup>, 49,300 participants with 5,215 participants including repeated scans, a total of 54,515 anatomical and functional scans, 25,639 women, age =  $64.37 \pm 7.98$  years (mean  $\pm$  s.d.)) was used to evaluate computational efficiency, a dataset with manual annotations of the anatomical regions (the Mindboggle-101 dataset<sup>18</sup>,  $n = 101$ , 44 women, age =  $28.4 \pm 8.6$  years) was used to evaluate anatomical parcellation, and two repeated measures datasets (the MSC dataset<sup>19</sup> (ten participants and ten sessions per participant, five women, age =  $29.1 \pm 3.3$  years) and the CoRR-HNU dataset<sup>20</sup> (30 participants and ten sessions per participant, 15 women, age = 20–30 years)) were used to evaluate accuracy and test–retest reliability. Moreover, to evaluate the robustness in preprocessing clinical samples with anatomical distortions, we included three in-house patient datasets, including the BTH-Glioma dataset ( $n = 19$ , six women, age =  $43.2 \pm 11.3$  years), the SHH-DoC dataset ( $n = 15$ , eight women, age =  $41.2 \pm 7.2$  years) and the CRR-Stroke dataset ( $n = 19$ , six women, age =  $54.0 \pm 11.7$  years). Written informed consents were obtained for all participants. See Supplementary Note 1 for more detailed descriptions of each dataset.

### Comparison to alternative preprocessing tools

We compared the processing time, along with the anatomical and functional outcomes, of DeepPrep (version 24.1.1) and fMRIPrep (version 24.0.0).

**Processing time and computational costs.** The preprocessing times of both DeepPrep and fMRIPrep were recorded for comparison and evaluated in an identical hardware environment (a workstation with an Intel Core i9-10980XE 3.00 GHz  $\times$  36 CPU, an NVIDIA RTX 3090 GPU

with 24 GB of VRAM and 64 GB of RAM). We processed 100 participants from the UKBB dataset and then assessed the serial processing time for single participants (Fig. 2a and Extended Data Fig. 2). To ensure a fair comparison, we allocated identical hardware resources (eight CPUs, one GPU and 16 GB of RAM) for both pipelines. Additionally, to evaluate batch-processing capabilities on a single workstation, we preprocessed the maximum possible number of participants on the same workstation, without any hardware limitations, for both pipelines (Fig. 2b). To examine the scalability of DeepPrep in large-scale datasets, we also preprocessed 54,515 scans from 49,300 participants in the UKBB dataset using our pipeline in an HPC cluster of Changping Laboratory. The cluster is equipped with 1,920 CPU cores (2.90 GHz) and 20 NVIDIA RTX 3090 GPUs. To estimate computational expenses in the HPC environment, we calculated CPU hours, derived by multiplying the number of recruited CPU cores by the duration of usage. Of note, due to the reliance on GPU processing for DeepPrep, we empirically converted one GPU hour into an equivalent of 15 CPU hours to enable a standardized comparison, according to the cloud computing charge (visit <https://aws.amazon.com/ec2/pricing/on-demand/> for pricing information on r7g.medium instances for CPU processing and g5.xlarge instances for GPU processing. Last visited on 21 October 2024).

**Assessments of structural imaging preprocessing.** Structural images were preprocessed using both DeepPrep and fMRIPrep. The preprocessed structural results were compared between the two pipelines, including anatomical segmentation and parcellation and similarities in the morphometrics, as elaborated below.

**Accuracy of anatomical parcellation and segmentation.** The surface-based parcellation and volumetric segmentation of structural images were automatically delineated by different pipelines. We first assessed the similarity, measured by the Dice coefficient, between the parcellations and segmentations automatically generated by different preprocessing pipelines and manually delineated annotations that were considered as the ‘ground truth’. Higher Dice coefficients indicate greater similarity and higher accuracy. To visually demonstrate differences in parcellation and segmentations, we directly estimated the difference percentages in Dice coefficients from two pipelines.

**Similarity to the morphometric atlas.** To assess surface registration performance, we compared similarities in the morphometrics, including sulcal depth and curvature, between the aligned surfaces and the FreeSurfer ‘fsaverage’ template surface. We employed the mean squared error to measure the dissimilarity, with lower mean squared error values indicating better alignment. We used the metrics to assess registration performance because the primary objective of surface registration is to align these anatomical features between individual cortical surfaces and atlases<sup>12,30,31</sup>.

**Assessments of functional imaging preprocessing.** Preprocessed functional images were compared between DeepPrep and fMRIPrep in multiple aspects, including spatial normalization, tSNR, task activations and functional connectivity. To ensure a fair comparison, we applied identical post-processing pipelines for task and resting state fMRI for both preprocessing pipelines. For the task fMRI, we used FSL FILM<sup>32</sup> (FMRIB’s improved linear model) to perform a standard general linear model for the first-level analysis after high-pass temporal filtering (100 s). Second-level inference was performed using FLAME<sup>33</sup> (FMRIB’s local analysis of mixed effects), based on the first-level outcomes. For post-processing resting state fMRI, we included bandpass filtering (0.01–0.08 Hz) and regression of nuisance variables, including six motion parameters, white matter signal, ventricular signal, whole-brain signal and their first-order temporal derivatives.

**Temporal signal-to-noise ratio.** The voxel-wise tSNR was calculated to assess the amount of informative signal relative to the noise level in preprocessed images, following a previous report<sup>34</sup>. The tSNR was estimated using the tSNR of Nipype, which first applied quadratic detrending and then calculated the temporal mean and divided by its temporal standard deviation for each voxel of each participant. The individual voxel-wise tSNR map was averaged across all participants.

**Spatial normalization.** After normalizing the BOLD fMRI images to the MNI152NLin6Asym volumetric template, we calculated the standard deviation map of averaged BOLD time series across participants to examine the performance of spatial normalization. A higher standard deviation around the brain outline indicates a lower performance of spatial normalization.

**Motor task activation map.** The MSC motor task<sup>19</sup> is a block design fMRI task with conditions of tongue, left hand, right hand, left foot and right foot, adapted from the paradigm used in the Human Connectome Project<sup>35</sup>. Task-evoked activations were modeled by a general linear model for each voxel and each session for the first-level analyses. The second-level analyses averaged data across session for each participant based on the fixed effect model. The third-level analyses grouped data across participants based on the mixed effect model. The task post-processing analyses were performed using FSL<sup>23</sup>.

**Seed-based cortico-cortical and cortico-cerebellar functional connectivity.** To assess the similarity in resting state functional connectivity (RSFC) derived from different pipelines, we examined typical long-range RSFC of a cortico-cortical and a cortico-cerebellar circuit. Specifically, we identified cortical and cerebellar seeds from previous studies<sup>36</sup>. For the cortico-cortical circuit, a seed in the left-hemisphere posterior cingulate cortex (MNI coordinates = -2, -53, 26) was selected. For the cortico-cerebellar circuit, a seed in the left angular gyrus (MNI coordinates = -49, -63, 45) was selected. Each seed region of interest (ROI) was a 6-mm sphere centered around the MNI coordinates of each seed. The RSFC was estimated by Pearson’s correlation coefficient between average BOLD time series within each seed ROI. To normalize the distribution of correlation coefficients, these  $r$  values were converted to  $z$  values with Fisher’s  $r$ -to- $z$  transformation. To demonstrate both the surface-based and volumetric RSFC, we performed surface-based analyses for the posterior cingulate cortex seed and volumetric analyses for the angular gyrus seed. The similarity in RSFC was assessed in the MSC dataset.

**Test–retest reliability in functional connectomes.** To assess test–retest reliability in functional connectomes, we used a volumetric atlas comprising 300 cortical and subcortical ROI<sup>37</sup> and a cortical surface atlas comprising 300 cortical ROI<sup>38</sup> to estimate the whole-brain and cortical functional connectomes, respectively. Test–retest reliability was measured by calculating the similarity between functional connectomes obtained from two different segments in the repeated measures datasets, namely, the MSC dataset and the CoRR-HNU dataset. Furthermore, we investigated the dependence of reliability on scanning duration by randomly selecting and concatenating two or more sessions to measure the functional connectome and its corresponding test–retest reliability, following the procedure described previously<sup>19</sup>.

### Application to data in the clinical setting

To assess the pipelines’ robustness in handling clinical samples with distorted brains, we included a total of 53 samples from three clinical datasets<sup>8,28,29</sup>, comprising patients with various conditions, including stroke, glioma and disorders of consciousness (see Supplementary Note 1 for more details). Notably, these clinical samples all failed to be completely preprocessed by FreeSurfer within 48 h on a CPU. To compare the pipelines’ performance, we employed both DeepPrep and

fMRIprep to preprocess all samples and recorded the processing time, completion and accuracy rates. The error reasons were also analyzed.

**Completion and acceptable ratios.** Completion is defined as successfully finishing the process within 48 h on eight CPUs and one GPU, with interpretable outputs such as meaningful reconstructed surfaces, registration files and preprocessed volumetric data, although the outputs are not guaranteed to be accurate. The completion ratio was determined as the proportion of clinical datasets that successfully completed all preprocessing steps, resulting in interpretable output files. On the other hand, the acceptable ratio was defined as the proportion of samples with preprocessing results deemed acceptable for subsequent processing, considering the rationality of surface reconstruction, anatomical segmentation, registration quality and functional connectivity. The assessment of preprocessing results was conducted independently by three experienced neuroimaging experts through visual inspection (J.R., W.C. and H.L.). The experts were blinded to the label of the preprocessing pipeline to minimize subjective biases. A sample was classified as acceptable when at least two neuroscientists reached a consensus on its output quality.

**Preprocessing error categories.** During the assessment of failed and inaccurate samples, we conducted a systematic step-by-step inspection of the preprocessing outputs to identify the underlying reasons for preprocessing errors. Based on this analysis, the failures were categorized into three distinct groups: segmentation errors, surface reconstruction failures and surface registration errors. Segmentation errors were observed in samples where clear inaccuracies were present in the segmentation of anatomical tissues. These errors were more commonly found in samples with extensive lesions or substantial head motion during T1w imaging scanning. Surface reconstruction failures were characterized by either the failure to reconstruct cortical surfaces or an excessively prolonged processing time (exceeding 48 h) to produce results. These issues often arose due to challenges in accurately fixing surface topology in distorted brains. Surface registration errors indicated misalignment of morphometric features, such as sulci and gyri. Such misalignments could lead to inaccuracies in the registration of brain structures.

### Meta-analysis of large-scale neuroimaging studies

To summarize recent trends in big data of neuroimaging, we conducted a meta-analysis of large-scale neuroimaging studies. First, we performed a PubMed (<http://www.pubmed.gov>) search on studies published before 30 June 2023, using the keywords (large scale OR large sample size) AND (neuroimage OR \*MRI). Next, we collected further studies by reviewing the reference lists of relevant papers, including studies that employed open-source big data or reviews focusing on large-sample size studies. The inclusion criteria for large-scale datasets were as follows: (1) the study recruited a sample larger than 1,000 healthy participants or 400 participants from specific populations, such as patients, twins or preterm brains; (2) the study was peer reviewed; and (3) the study contained MRI data with at least one type of structural and functional images. These literature searches yielded a total of 42 studies, comprising over 200,000 participants. Each sample was then categorized into one of three types of projects based on its funding resources. The first type was named ‘big project’, as it was supported by a single funding source. The second type was named ‘consortium’, as it involved international collaboration with multiple funding sources. The last type was named ‘data sharing platform’, as it shared data on the same or different topics from different groups.

### Statistical analysis

The  $\chi^2$  test was used for comparisons in completion and accuracy ratios. Processing time, tSNR and similarities in functional connectomes of two pipelines were statistically compared using two-tailed

paired-sample *t*-tests. False discovery rate correction was used to account for multiple testing.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Data from the UKBB dataset used here are available at <https://www.fmrib.ox.ac.uk/ukbiobank/>, with the access application 99038. Data from the MSC dataset are available at <https://openneuro.org/datasets/ds000224/versions/1.0.3>. Data from the CoRR-HNU dataset are available at [http://fcon\\_1000.projects.nitrc.org/indi/CoRR/html/hnu\\_1.html](http://fcon_1000.projects.nitrc.org/indi/CoRR/html/hnu_1.html). Data from the Mindboggle-101 dataset are available at <http://mindboggle.info/data>. Clinical datasets are available on request from H.L., J.R., X.W., Y.W. and T.J. They have not yet been made available through public databases, as data collection is still ongoing. The 300 volumetric cortical and subcortical ROI are available at [https://greenelab.ucsd.edu/data\\_software](https://greenelab.ucsd.edu/data_software). The Schaffer’s 300 ROI of the cortical surfaces used here are available for access at [https://github.com/ThomasYeoLab/CBIG/tree/master/stable\\_projects/brain\\_parcellation/Schaefer2018\\_LocalGlobal](https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal). Meta-analysis data for Extended Data Fig. 1 were collected from PubMed (<http://www.pubmed.gov>). Source data are provided with this paper.

### Code availability

DeepPrep is available under the Apache License at <https://github.com/pBFSLab/DeepPrep>. Docker images corresponding to every new release of DeepPrep are automatically generated and made available on Docker Hub (<https://hub.docker.com/r/pbfs/dep>). Software documentation is available at <https://deepprep.readthedocs.io>. DeepPrep code is also available through a Code Ocean compute capsule: <https://doi.org/10.24433/CO.2579051.v1> (ref. 39).

### References

- Hoffmann, M. et al. SynthMorph: learning contrast-invariant registration without acquired images. *IEEE Trans. Med. Imaging* **41**, 543–558 (2022).
- Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR Biomed.* **10**, 171–178 (1997).
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
- Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
- Esteban, O., Markiewicz, C., Blair, R., Poldrack, R. & Gorgolewski, K. SDCflows: susceptibility distortion correction workFLOWS. *Zenodo* <https://doi.org/10.5281/zenodo.3758524> (2020).
- Ciric, R. et al. TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. *Nat. Methods* **19**, 1568–1571 (2022).
- Wratten, L., Wilm, A. & Goke, J. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nat. Methods* **18**, 1161–1168 (2021).
- Tang, Z. et al. Efficacy and safety of high-dose TBS on poststroke upper extremity motor impairment: a randomized controlled trial. *Stroke* **55**, 2212–2220 (2024).
- Cui, W. et al. Personalized fMRI delineates functional regions preserved within brain tumors. *Ann. Neurol.* **91**, 353–366 (2022).
- Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195–207 (1999).
- Yeo, B. T. et al. Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE Trans. Med. Imaging* **29**, 650–668 (2010).

32. Woolrich, M. W., Ripley, B. D., Brady, M. & Smith, S. M. Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage* **14**, 1370–1386 (2001).
33. Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M. & Smith, S. M. Multilevel linear modelling for fMRI group analysis using Bayesian inference. *Neuroimage* **21**, 1732–1747 (2004).
34. Smith, S. M. et al. Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144–168 (2013).
35. Barch, D. M. et al. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
36. Thomas Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
37. Seitzman, B. A. et al. A set of functionally-defined brain regions with improved representation of the subcortex and cerebellum. *Neuroimage* **206**, 116290 (2020).
38. Schaefer, A. et al. Local–global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
39. Ren, J. et al. DeepPrep: an accelerated, scalable, and robust pipeline for neuroimaging preprocessing empowered by deep learning. *Code Ocean* <https://doi.org/10.24433/CO.2579051.v1> (2024).

## Acknowledgements

We thank Y. Li for drawing the logo of DeepPrep and the icons in Fig. 1b. This work was supported by Changping Laboratory (H.L.), the China Postdoctoral Science Foundation (2022M720529 to J.R., 2023M730175 to W.C.), the Beijing Natural Science Foundation (JQ23040 to Y.W.) and the Natural Science Foundation of China (82072786 to Y.W.).

## Author contributions

Conceptualization: J.R., N.A., H.L.; data curation: X.W., Y.W., T.J.; funding acquisition: H.L., J.R., W.C., Y.W.; methodology: J.R., N.A.; project administration: J.R.; resources: H.L., J.R.; software and documentation: N.A., C.L., Y.Z., Z.S., J.R., W.Z., N.G., W.C., Q.H., W.W.; supervision: H.L., J.R.; validation: J.R., W.C., H.L., N.G.; visualization: J.R., Z.S., S.L.; writing (original draft): J.R.; all authors reviewed, edited and approved the manuscript.

## Competing interests

H.L. is the chief scientist of Neural Galaxy Inc. The other authors declare no competing interests.

## Additional information

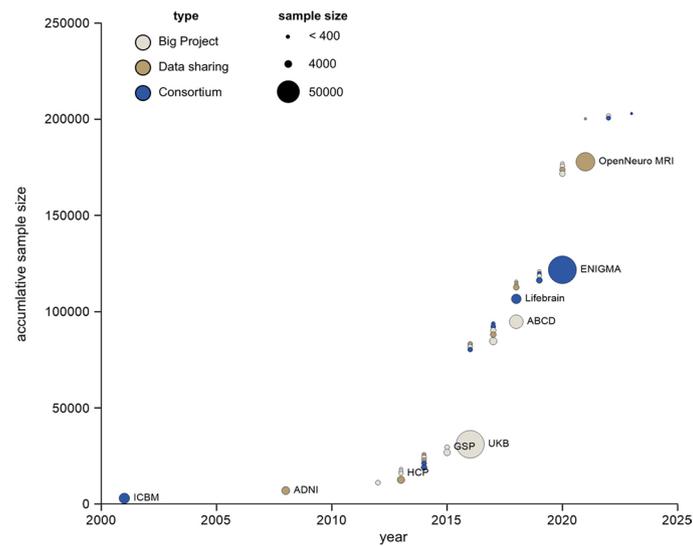
**Extended data** is available for this paper at <https://doi.org/10.1038/s41592-025-02599-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41592-025-02599-1>.

**Correspondence and requests for materials** should be addressed to Jianxun Ren or Hesheng Liu.

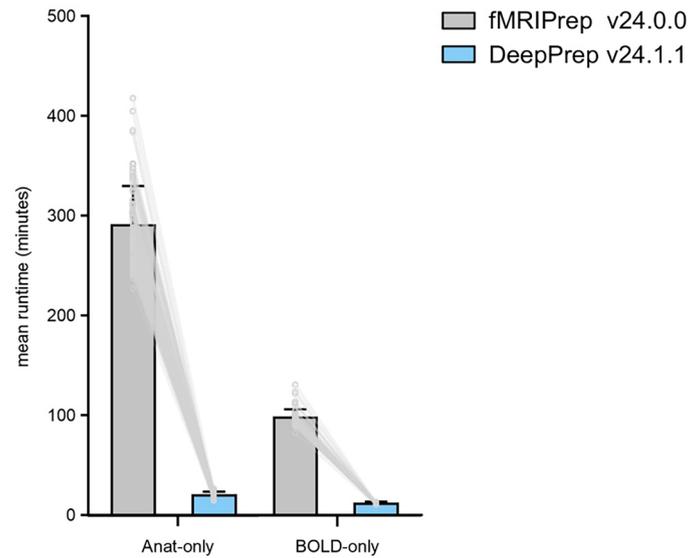
**Peer review information** *Nature Methods* thanks Melanie Ganz and Kaustubh Patil for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Nina Vogt, in collaboration with the *Nature Methods* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



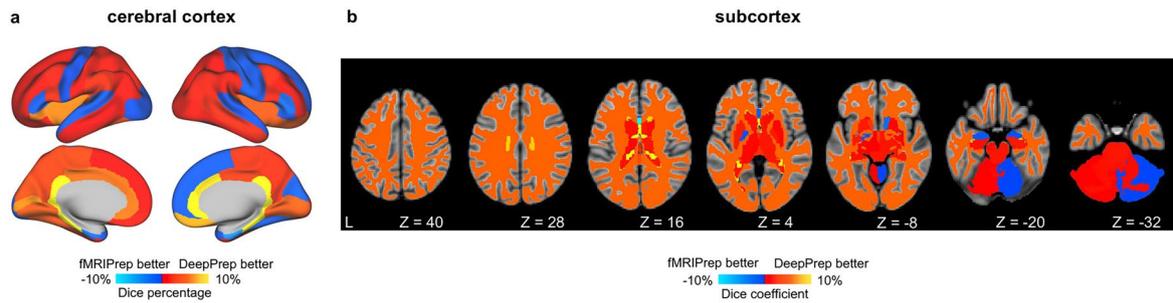
**Extended Data Fig. 1 | Exponential growth of publicly available neuroimaging datasets from 2000 to 2023.** The figure depicts the cumulative sample size of publicly accessible neuroimaging datasets. A total of forty-two large-scale neuroimaging datasets were selected through a comprehensive meta-analysis spanning from the year 2000 to 2023. These datasets are categorized into three representative types: 'big project' (white circles), supported by a single funding source, 'consortium' (blue circles), resulting from international collaborations

involving multiple funding sources, and 'sharing data platform' (sand circles), such as OpenNeuro, which serves as a data repository. The size of each circle in the plot corresponds to the sample size of the respective dataset. The project names of some representative datasets are labeled next to the corresponding circles in the dataset. Stacked circles represent cases where multiple different datasets were released in a single year. The Y-axis represents the cumulative sample size.



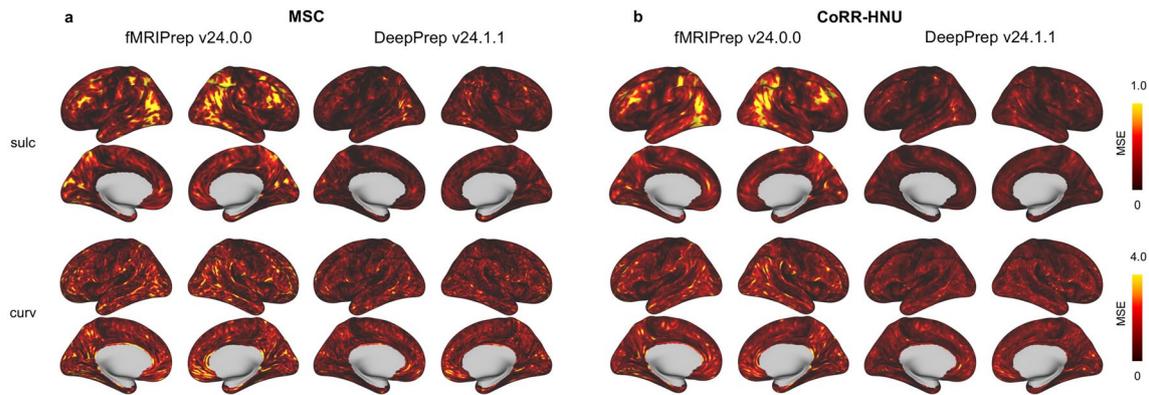
**Extended Data Fig. 2 | DeepPrep substantially accelerates anatomical and functional imaging preprocessing.** In addition to the comparison of processing time for the entire pipeline between fMRIPrep and DeepPrep, we assessed separate processing time for anatomical and functional images ( $n = 100$  subjects from the UKBB dataset). When considering anatomical preprocessing with the ‘-anat-only’ flag, DeepPrep completed the task within  $20.9 \pm 2.5$  minutes, marking a substantial 14.0-fold increase in speed compared to fMRIPrep, which took

$291.6 \pm 38.4$  minutes (two-tailed paired t-test,  $t_{(99)} = 71.3$ ,  $p = 6.9 \times 10^{-87}$ ). Similarly, when processing functional images with the ‘-bold-only’ flag, DeepPrep exhibited significant efficiency, finishing the task in  $12.9 \pm 0.5$  minutes, a 7.7-fold improvement over fMRIPrep, which required  $99.1 \pm 7.1$  minutes (two-tailed paired t-test,  $t_{(99)} = 120.4$ ,  $p = 3.7 \times 10^{-109}$ ). During functional image processing, preprocessed anatomical images were provided at the outset. Data are presented as mean  $\pm$  s.t.d.



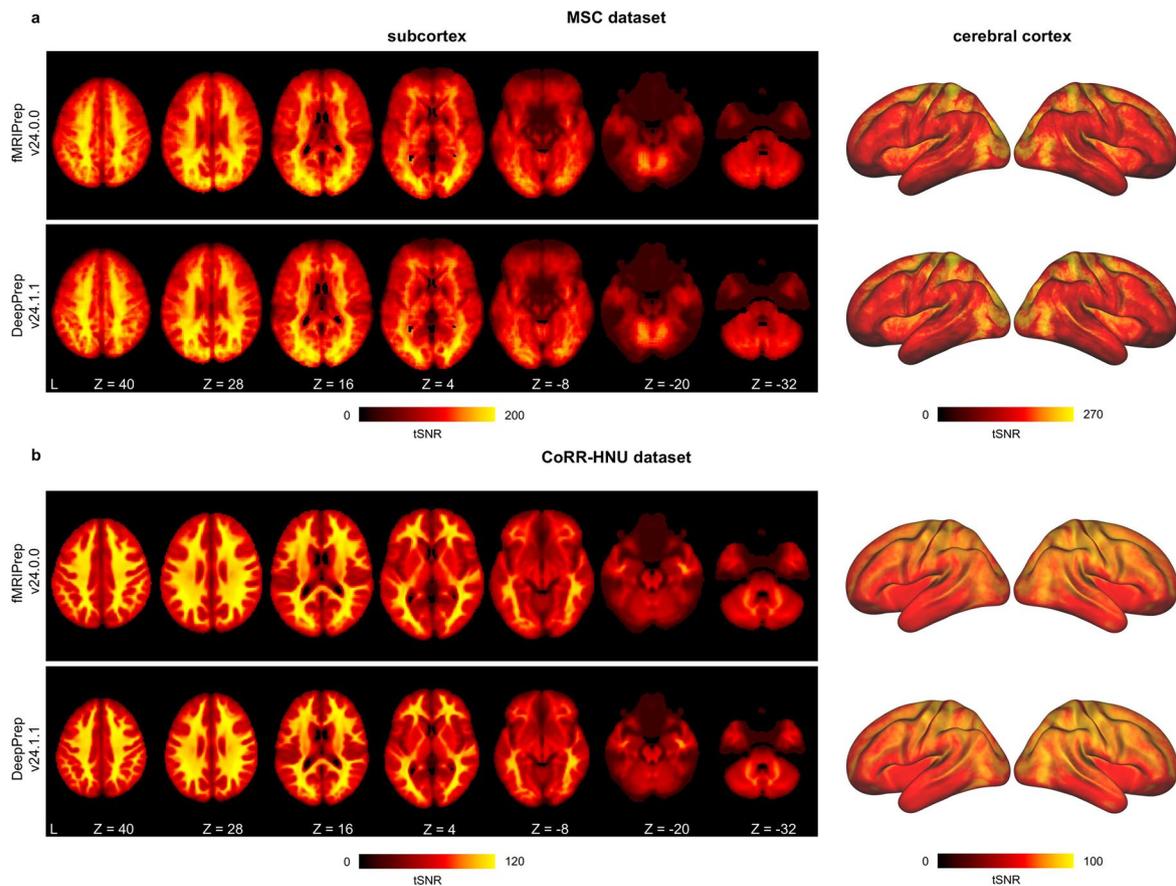
**Extended Data Fig. 3 | Comparable performance in brain tissue segmentation of DeepPrep and fMRIPrep.** DeepPrep's performance in brain tissue segmentation was assessed and compared with that of fMRIPrep. To demonstrate segmentation accuracy, both DeepPrep and fMRIPrep were applied to the Mindboggle-101 dataset. This dataset includes anatomical images and manually segmented brain regions by experts as the 'ground truth' of brain tissue segmentation. The accuracy of automatic segmentations produced by both pipelines was evaluated by calculating Dice coefficients, which measure the similarities of brain regions between the automatic and manual segmentations.

Comparison of segmentation accuracy between DeepPrep and fMRIPrep was conducted by evaluating the percentage difference in Dice coefficients for each brain region. Positive percentages indicate higher accuracy achieved by DeepPrep, while negative percentages indicate higher accuracy for fMRIPrep. **a)** Fifty-two out of 62 cortical regions display positive percentages. **b)** Similar comparisons were performed for 39 subcortical regions, where 31 regions show positive percentages. All statistics are reported in Supplementary Table 4 and Table 5.



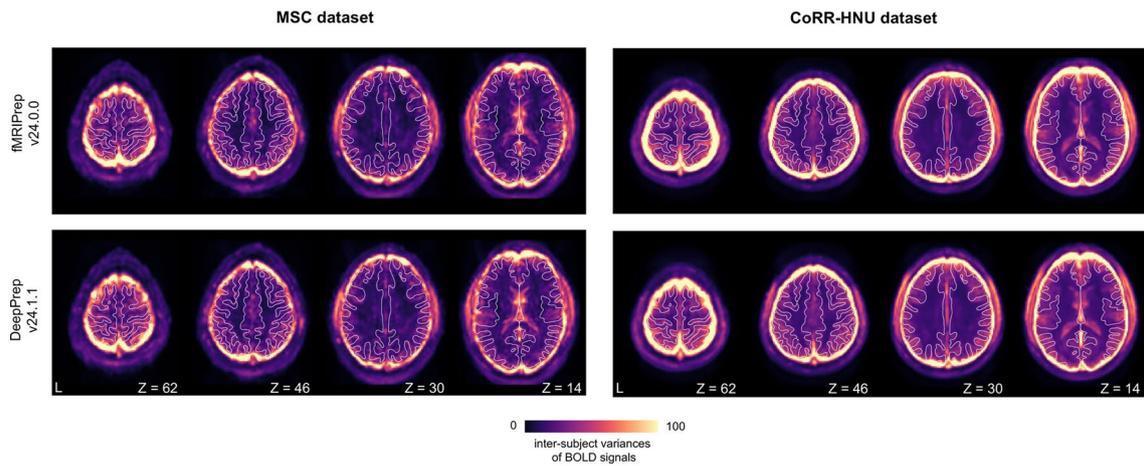
**Extended Data Fig. 4 | Improved similarity of cortical morphometric patterns with target atlases by DeepPrep.** Both fMRIPrep and DeepPrep were applied to preprocess anatomical images from two datasets: **a**) the MSC dataset and **b**) the CoRR-HNU dataset. Both preprocessing pipelines reconstructed individual cortical surfaces and aligned surface morphometrics with target atlases.

To evaluate the performance in surface reconstruction and registration, we conducted a comparison using two morphometric measures: sulcal depth (top row), curvature (bottom row). The assessment was based on the mean squared error (MSE) when contrasting these measures with the target atlases.



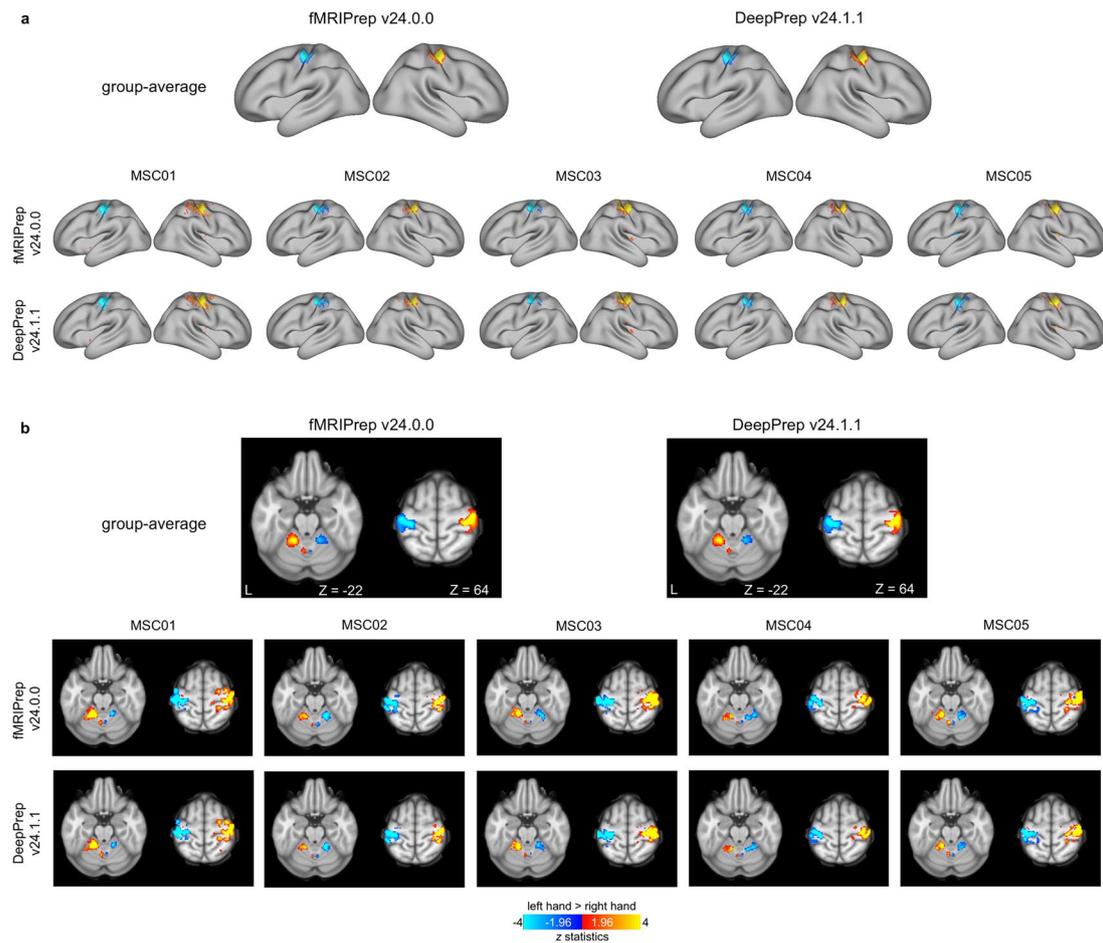
**Extended Data Fig. 5 | Enhanced temporal signal-to-noise ratio with DeepPrep.** Following spatial normalization of BOLD-fMRI signals to the MNI152NLin6Asym volumetric template, we calculated the temporal signal-to-noise ratio (tSNR) for the BOLD-fMRI signals. DeepPrep exhibited comparable tSNR values to fMRIprep in whole brains (left) and cortical surfaces

(right, fasverage6) in both the **(a)** MSC (average tSNR of the cerebral cortex, fMRIprep =  $159.3 \pm 26.8$ , DeePrep =  $158.9 \pm 26.1$ , two-tailed paired-sample t-test,  $t_{(9)} = 1.1$ ,  $p = 0.32$ ) and **(b)** CoRR-HNU datasets (average tSNR, fMRIprep =  $65.4 \pm 2.3$ , DeePrep =  $65.9 \pm 2.2$ ,  $t_{(29)} = 8.2$ ,  $p = 5.0 \times 10^{-9}$ ).



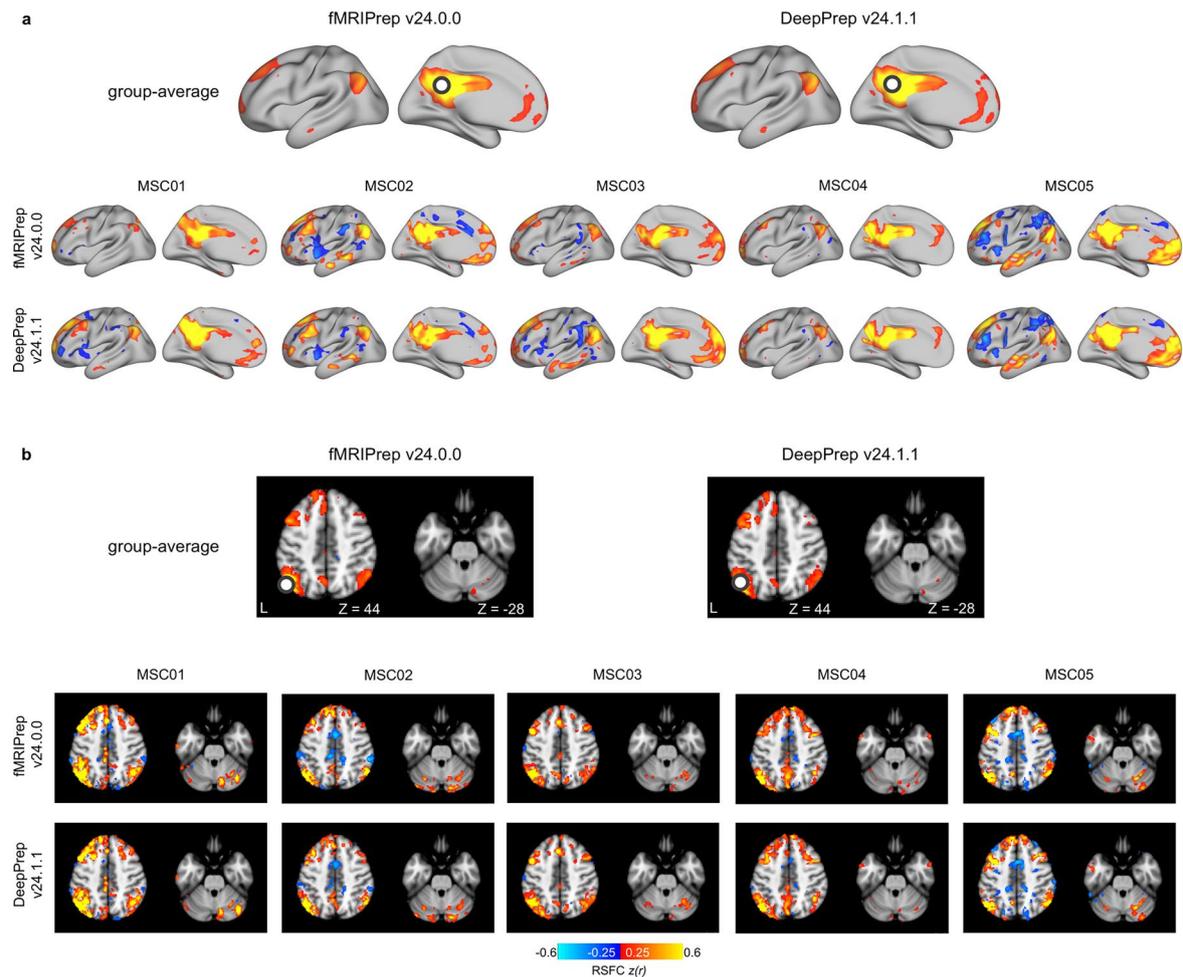
**Extended Data Fig. 6 | Comparable inter-subject variances in average BOLD signals between DeepPrep and fMRIPrep.** Following spatial normalization, we assessed the voxel-wise inter-subject variances in average BOLD signals from the

MSC and CoRR-HNU datasets. Lighter colors in the figures indicate larger inter-subject variances. The white matter boundaries are delineated in white on the MNI152NLin6Asym volumetric template.



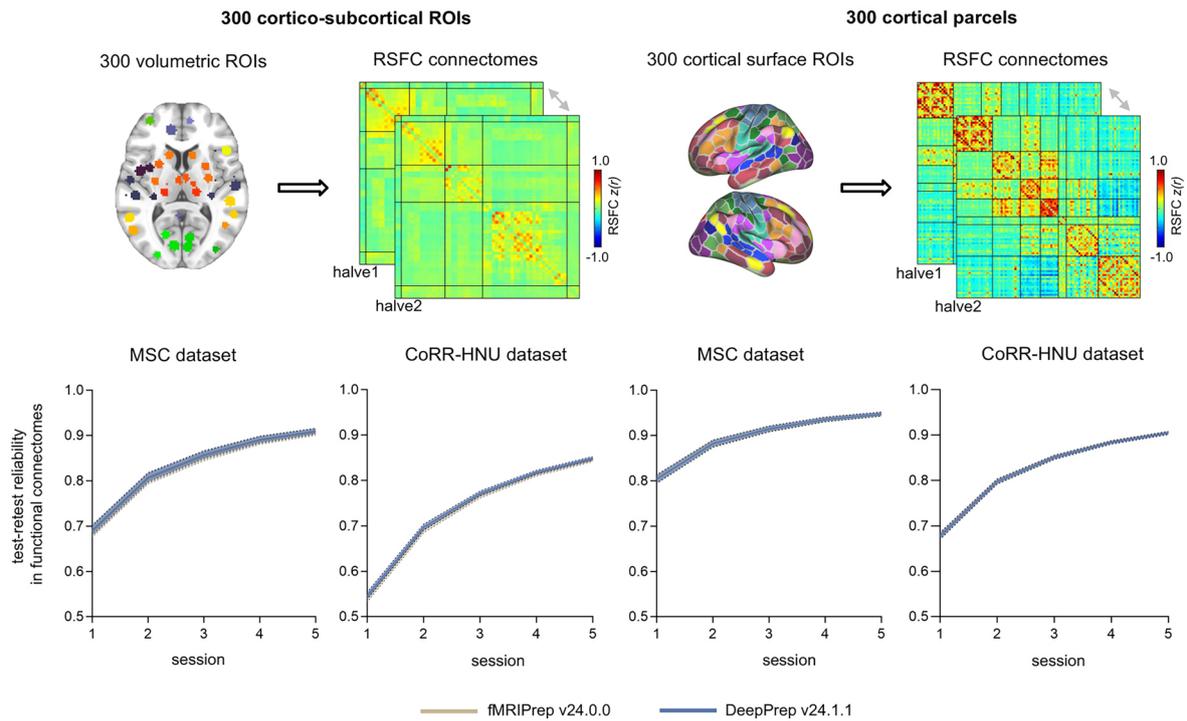
**Extended Data Fig. 7 | Comparable patterns and strength of task-evoked responses between DeepPrep and fMRIPrep.** DeepPrep and fMRIPrep were employed for preprocessing the motor task-evoked fMRI data from the MSC dataset. Subsequent processing steps, including high-pass filtering and general linear modeling, were consistent between the two pipelines. **a)** Contrast maps

derived from the surface-based pipelines depicting the activation patterns for the left hand vs. right hand task are presented on a cortical surface template. **b)** Volumetric contrast maps yielded by the volumetric pipelines are shown in a volumetric template.



**Extended Data Fig. 8 | Comparable seed-based functional connectivity between DeepPrep and fMRIPrep.** DeepPrep and fMRIPrep were utilized to preprocess resting-state fMRI data from the MSC dataset. Subsequent processing steps, including band-pass filtering (0.01-0.08 Hz) and confound regression, were consistent between the two pipelines. A 6-mm spherical seed ROI was placed in the left posterior cingulate cortex (MNI coordinate = -2, -53, 26)

for surface-based functional connectivity analyses and the left angular gyrus (MNI coordinate = -49, -63, 45) for volumetric functional connectivity analyses, respectively. The seed locations are represented by the white circles. The seed-based functional connectivity maps derived from surface and volumetric pipelines are displayed within **a**) a cortical surface template and **b**) a volumetric template, respectively.



**Extended Data Fig. 9 | Comparable test-retest reliability of functional connectomes between DeepPrep and fMRIPrep.** To estimate whole-brain and cortical functional connectomes, we employed 300 volumetric ROIs encompassing both cortical and subcortical regions, as well as Schaefer's 300 cortical surface ROIs, respectively. Test-retest reliability was quantified by measuring the similarity between functional connectomes derived from two independent segments randomly sampled from one or multiple sessions in two

repeated measured datasets, namely the MSC and CoRR-HNU dataset. To elucidate the relationship between test-retest reliability and data amounts, we assessed reliability using one to five sessions. No significant differences in the test-retest reliability of functional connectomes were observed between DeepPrep and fMRIPrep (two-tailed paired-sample *t*-tests, FDR-corrected *p*-values > 0.05). Data are presented as mean  $\pm$  s.t.d.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** Data was sourced from multiple sites, as explained in the manuscript and in the Data section below. Data was collected from MRI scanners, no software or code was used.

**Data analysis** No commercial software were used for data analysis. All data analyses were made using open source code, developed by the project or other projects. DeepPrep is available under the apache license at <https://github.com/pBFSLab/DeepPrep> as well as Code Ocean <https://doi.org/10.24433/CO.2579051.v1>. Docker images corresponding to every new release of DeepPrep are automatically generated and made available on Docker Hub (<https://hub.docker.com/r/pbfs/DeepPrep>). Software documentation is available at <https://deepprep.readthedocs.io>. DeepPrep incorporates SimpleITK v2.3.0, AFNI 24.0.00, FastSurfer v1.1.0, SUGAR 1.0.0, SynthMorph 2, FastCSR v1.0.0, FreeSurfer 7.2.0, SDCFlow 2.8.1, and niworkflows 1.10.0, nipype 1.8.6, TemplateFlow 23.2.0, Nextflow 23.10.4, Redis 6.0.16, and python-redis-lock, 4.0.0. For evaluation purposes, we also used FreeSurfer 6.0.0, FSL 6.0.5.1, and fMRIprep v24.0.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data from the UKBB dataset used here are available at <https://www.fmrib.ox.ac.uk/ukbiobank/>, with the access application 99038. Data from the MSC dataset are available at <https://openneuro.org/datasets/ds000224/versions/1.0.3>. Data from the CoRR-HNU dataset are available at [http://fcon\\_1000.projects.nitrc.org/indi/CoRR/html/hnu\\_1.html](http://fcon_1000.projects.nitrc.org/indi/CoRR/html/hnu_1.html). Data from the Mindboggle-101 dataset are available at <http://mindboggle.info/data>. Clinical datasets are available on request from H.L., J.R., W.X., Y.W., and T.J. They have not yet been made available through public databases as data collection is still ongoing. The 300 volumetric cortical and subcortical ROIs are available from [https://greenelab.ucsd.edu/data\\_software](https://greenelab.ucsd.edu/data_software). The Schaffer's 300 ROIs of the cortical surfaces used here are available for access from [https://github.com/ThomasYeoLab/CBIG/tree/master/stable\\_projects/brain\\_parcellation/Schaefer2018\\_LocalGlobal](https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal). Meta-analysis data for Extended Data Fig. 1 were collected from PubMed (<http://www.pubmed.gov>).

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

we used data from either open repositories or collected for other projects that have their own policies for reporting. We did not analyze subjects by either sex or gender.

Population characteristics

The population of all datasets (pre-exclusion) contained 55,450 scans from 49,865 (25,864 female) subjects. For each dataset the population characteristics are summarized below:  
 UKBB: 49,300 (25,639 Female). Mean age=67.4, standard deviation=8.0 years  
 MSC: 10 (5 Female). Mean age=29.1, standard deviation=3.3 years  
 CoRR-HNU: 30 (15 Female). Mean age=24.0, standard deviation=2.4 years  
 Mindboggle-101: 101 (44 Female). Mean age=28.4, standard deviation=8.6 years  
 BTH-Glioma: 168 (71 Female). Mean age=43.2, standard deviation=11.3 years  
 SHH-DoC: 38 (18 Female). Mean age=41.2, standard deviation=7.2 years  
 CRRC-Stroke: 218 (72 Female). Mean age=54.0, standard deviation=11.7 years

Recruitment

No participants were recruited for this study. All data comes from previous studies or public repositories.

Ethics oversight

No new data were collected specifically for this study; all data were acquired with IRB approval at their original institutions. The IRB of Beijing Tiantan Hospital, Capital Medical University approved the BTH-glioma study. The IRB of Shanghai Huashan Hospital, Fudan University approved the SHH-DoC study. The IRB of China Rehabilitation Research Center approved the CRRC-Stroke study. The UKBB, MSC, CoRR-HNU, and Mindboggle-101 datasets are publicly available, de-identified data resources.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No sample size calculation was performed, but we collected a large and heterogeneous data set spanned enough variation to adequately test the software, including seven different datasets acquired with a wide range of scanners, demographics, brain conditions, and acquisition parameters, for a total of 55,450 scans from n=49,865 (n=25,864 females) scans. The acquisition parameters are presented in Supplementary Note 1 and Table 2.

Data exclusions

To evaluate the robustness of preprocessing pipelines in intractable clinical cases, We excluded 23, 149, and 199 cases from the SHH-DoC, BTH-Glioma, CRRC-Stroke datasets, respectively, if these cases can be successfully processed by FreeSurfer within 48 hours. We excluded four cases due to either poor data quality or the absence of necessary anatomical annotations in the Mindboggle-101 dataset.

Replication

DeepPrep reproducibly preprocessed seven datasets with distinct population characteristics and scanning parameters. All processing used a containerized pipeline. A computational capsule is available on Code Ocean.

Randomization	Participants were not divided into separate experimental groups.
Blinding	When evaluating processing accuracy in clinical datasets, the neuroimaging experts were blinded to the label of the preprocessing pipeline to minimize subjective biases.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

## Magnetic resonance imaging

### Experimental design

Design type	resting-state and block-design task
Design specifications	We only used the motor task data from the MSC dataset and used resting-state fMRI data from other datasets. During the motor task, subjects received visual cues instructing them to either close and relax their hands, flex and relax their toes, or wiggle their tongue. Each task block began with a 2.2-second cue indicating the specific movement to be performed. Subsequently, participants executed 12 movements during each block, with the total duration of movements per block being 13.2 seconds. Each task run comprised two blocks for each type of movement and three blocks of resting fixation, each lasting 15.4 seconds.
Behavioral performance measures	No behavioral performance measures or used for this study

### Acquisition

Imaging type(s)	Structural (T1w) and functional
Field strength	3 Tesla
Sequence & imaging parameters	Data were sourced from different sites. The descriptions of imaging parameters are shown in Supplementary Note 1
Area of acquisition	Head/Brain
Diffusion MRI	<input type="checkbox"/> Used <input checked="" type="checkbox"/> Not used

### Preprocessing

Preprocessing software	DeepPrep v24.1.1 or fMRIPrep v24.0.0
Normalization	Volumetric linear and nonlinear spatial normalizations were performed using SynthMorph. Surface-based linear and nonlinear normalizations were performed using SUGAR.
Normalization template	There are multiple template options in DeepPrep, as implemented by Templateflow. The default volumetric and surface templates for DeepPrep are MNI152Nlin6Asym and fsaverage6, respectively. This default templates were used for processing the test datasets.
Noise and artifact removal	Before functional connectivity estimation, we remove artifacts and noises through the bandpass filtering (0.01-0.08Hz) and regression of nuisance variables, including 6 motion parameters, white-matter signal, ventricular signal, whole-brain signal, and their temporal derivatives.
Volume censoring	To directly compare the outputs from different pipelines, we did not apply volume censoring, but DeepPrep can automatically provide censoring label in the confound estimation.

## Statistical modeling & inference

Model type and settings For the motor task fMRI, we used the FSL FLIM (FMRIB's Improved Linear Model) to perform a standard general linear model for the first level analysis after high-pass temporal filtering (100s). The second level inference was performed using the FLAME (FMRIB's Local Analysis of Mixed Effects), based on the first-level outcomes.

Effect(s) tested The only task under analysis was the motor task, and the only contrast analyzed was "left hand - right hand".

Specify type of analysis:  Whole brain  ROI-based  Both

Anatomical location(s) In most of analyses, we performed whole-brain analyses, including anatomical segmentation, task activations, functional connectome. In the ROI-based functional connectivity analysis, we used two seeds in the posterior cingulate cortex (PCC; MNI coordinate = -2, -53, 26) and the left angular gyrus (AG, MNI coordinate = -49, -63, 45) from previous studies. Each seed ROI was a 6-mm spheres centered around the MNI coordinates of each seed.

Statistic type for inference  
(See [Eklund et al. 2016](#)) voxel-wise

Correction FDR

## Models & analysis

n/a | Involved in the study

Functional and/or effective connectivity

Graph analysis

Multivariate modeling or predictive analysis

Functional and/or effective connectivity Resting-state functional connectivity was estimated by Pearson's correlation coefficient of BOLD-fMRI signals