NiChart: A machine learning oriented neuro-imaging brain chart, derived from 71,820 MRI scans, and its methodology

Guray Erus[1,&], Ahmed Abdulkadir[1,&] Randa Melhem[1], Elizabeth Mamourian[1], Sindhuja Tirumalai Govindarajan[1], Vishnu Bashyam[1], Dhivya Srinivasan[1], Yuhan Cui[1], Alex Getka[1], George Aidinis[1], Di Wu[1], Kyunglok Baik[1], Zhijian Yang[1], Junhao Wen[1], Ashish Singh[1], Jimit Doshi[1], Ioanna Skampardoni[1,2], Jiong Chen[1], Gyujoon Hwang[1], Mark Bergman[1], Raymond Pomponio[1], Paraskevi Parmpi, Katharina Wittfeld[3,4], Hans J. Grabe[3], Duygu Tosun[5], Murat Bilgel[6], Yang An[6], Daniel S. Marcus[7], Pamela LaMontagne[7], Susan R. Heckbert[8,9], Thomas R. Austin[8,9], Lenore J. Launer[10], Arthur W. Toga[47], Sylvia Villeneuve[48], M. Mallar Chakravarty[48], Brian S. Schwartz[49], Aristeidis Sotiras[11], Mark A. Espeland[12,13], Colin L Masters[14], Paul Maruff[14], Jurgen Fripp[16], Shari Waldstein[50,51], Michele K. Evans[52], Alan Zonderman[52], Tammie Benzinger[7], Aaron Schultz[53], Michael Properzi[53], Reisa Sperling[53,54], Sid O'Bryant[55], Dahyun Yi[56], Luigi Ferrucci[17], Kristine Yaffe[18,] Jeff Williamson[19], Paola Dazzan[21,22], Rene S. Kahn[23], Hugo G. Schnack[24], Geraldo F. Busatto[25], Benedicto Crespo-Facorro[26,27,28,29], Christos Pantelis[30,31], Stephen J. Wood[32,33,34], Chuanjuin Zhuo[35], Ruben C. Gur[37], Raquel E. Gur[37], Nikolaos Koutsouleris[3,21,38], Sterling C. Johnson[39], John C. Morris[40], Timothy J. Hohman[57,58], Marilyn S. Albert[41], David Wolk[44], R. Nick Bryan[42], Susan M. Resnick[6], Mohamad Habes[43], Daniel H. Wolf[37], Russel T. Shinohara[1,36], Theodore D. Satterthwaite[20], Yong Fan[1], Haochang Shou[1,46], Ilya M. Nasrallah[1,45], Christos Davatzikos[1,#,*]

[1] AI²D Center, University of Pennsylvania, Philadelphia, PA, USA

[2] School of Electrical and Computer Engineering, National Technical University of Athens, Greece

[3] Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Germany

[4] German Centre for Neurodegenerative Diseases (DZNE), Site Rostock/Greifswald, Germany

[5] Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, USA

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

[7] Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA

[8] Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA

[9] Department of Epidemiology, University of Washington, Seattle, WA, USA

[10] Neuroepidemiology Section, Intramural Research Program, National Institute on Aging, Bethesda, MD, USA

[11] Department of Radiology and Institute of Informatics, Washington University in St. Louis, St. Louis, MO, USA

[12] Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[13] Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA

[14] Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia

[16] CSIRO Health and Biosecurity, Australian e-Health Research Centre CSIRO, Brisbane, Queensland, Australia

[17] National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

[18] UCSF Weill Institute for Neurosciences, San Francisco, CA, USA

[19] Wake Forest University, School of Medicine, Center for Healthcare Innovation, Winston-Salem, NC, USA

[20] Penn Lifespan Informatics and Neuroimaging Center, University of Pennsylvania, Philadelphia, PA, USA

[21] Department of Psychological Medicine, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK

[22] NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London

[23] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[24] Department of Psychiatry, University Medical Center Utrecht, Utrecht, Netherlands.

[25] Institute of Psychiatry, Faculty of Medicine, University of São Paulo, São Paulo, Brazil.

[26] Mental Health Service, Hospital Universitario Virgen del Rocío, Seville, Spain.

[27] Centro de Investigación Biomédica en Red de Salud Mental, Instituto de Salud Carlos III (CIBERSAM), Sevilla, Spain.

[28] Instituto de Biomedicina de Sevilla (IBiS)-CSIC, Seville, Spain.

[29] Department of Psychiatry, Universidad de Sevilla, Seville, Spain.

[30] Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Royal Melbourne Hospital, Carlton South, VIC, Australia.

[31] Western Hospital Sunshine, Western Health, & Western Centre for Health Research & Education (WCHRE), St. Albans, VIC, Australia.

[32] Centre for Youth Mental Health, University of Melbourne, Melbourne, VIC, Australia.

[33] Orygen, Melbourne, VIC, Australia.

[34] University of Birmingham, Edgbaston, UK.

[35] Department of Psychiatric-Neuroimaging-Genetics and Co-morbidity Laboratory (PNGC-Lab), Nankai University Affiliated Tianjin Anding Hospital; Department of Psychiatry, Tianjin Medical University, Tianjin, China.

36 Penn Statistics in Imaging and Visualization Center, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[37] Neurodevelopment & Psychosis Section, Department of Psychiatry, Perelman School of Medicine, and the CHOP-Penn Lifespan Brain Institute, University of Pennsylvania, Philadelphia, PA, USA.

[38] Max-Planck Institute of Psychiatry, Munich, Germany.

[39] Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

[40] Knight Alzheimer Disease Research Centre, Washington University in St. Louis, St. Louis, MO, USA

[41] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[42] PENN, Philadelphia, PA, USA

[43] Biggs Alzheimer's Institute, University of Texas San Antonio Health Science Centre, TX, USA

[44] Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA

[45] Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA

[46] Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA

[47] Laboratory of Neuro Imaging (LONI), Stevens Neuroimaging and Informatics Institute, Keck School of Medicine

[48] Douglas Research Centre, Department of Psychiatry, McGill University

[49] Department of Environmental Health Sciences, Johns Hopkins University

[50] Department of Psychology, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250

[51] Division of Gerontology, Department of Medicine, University of Maryland School of Medicine & Geriatric Research Education and Clinical Center Baltimore Veterans Affairs Medical Center, USA

[52] Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

[53] Department of Neurology, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA

[54] Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

[55] Department of Internal Medicine, University of North Texas Health Science Center, Fort Worth, Texas, USA

[56] Institute of Human Behavioral Medicine, Medical Research Center, Seoul National University, Seoul, 03080, Republic of Korea
[57] Vanderbilt Memory & Alzheimer's Center, Vanderbilt University Medical Center, Nashville, TN USA
[58] Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN USA

& Equally contributing first co-authors

* Correspondence to Christos Davatzikos, Professor, Department of Radiology, AI^{2}D Center, University of Pennsylvania, christos.davatzikos@pennmedicine.upenn.edu

For the studies listed in Table S.1. See specific acknowledgements in "Acknowledgements"

Keywords

magnetic resonance imaging; big data analytics; machine learning; brain; aging; neurodegeneration; computational anatomy; computational physiology; chronic health conditions

Abstract

Brain Magnetic resonance imaging (MRI) has been widely adopted by studies of brain aging and neurologic disorders including neurodegenerative diseases, which have collectively generated a rich data resource for understanding and quantifying normal and pathologic brain aging. Machine learning (ML) methods have shown great promise for developing individualized MRI-based biomarkers with diagnostic and prognostic value. However, modest diversity and sample sizes of individual studies, as well as variations in MRI scanners and imaging protocols across studies, often limit the power and generalizability of results and derived models. We describe a neuro-imaging brain chart (NiChart), a dimensional neuroimaging system derived from and validated against a diverse harmonized dataset composed of 71,820 MRI time points from 53,757 participants across 23 studies, acquired in over 100 scanners and 3 decades (with an on-going expansion to an additional 12 studies and 9257 individuals). The dimensions of NiChart capture variation of structural brain features, as well as expression of multi-variate imaging signatures derived via ML models, such as imaging signatures of brain age and of Alzheimer's disease. These measures reflect different aspects of heterogeneity in brain aging trajectories in individuals with normal cognition, cognitive deficits, and dementia. Statistical harmonization methods in NiChart enable researchers to minimize confounding inter-scanner and inter-protocol variations and map MRI scans onto the brain chart, thereby gaining access to NiChart's normative statistics and ML-derived imaging signatures and further contributing their own measures to this reference space. We envision NiChart to be a community-based tool that can grow with user input of data and machine learning models from additional studies. To make NiChart easily accessible, we provide both cloud-based and locally installed software tools with a point-and-click graphical user interface. We demonstrate the utility of NiChart with applications in multiple, independently acquired studies of specific focus, including neurodegenerative diseases.

1 Main

Brain MRI has been used extensively to study trajectories of the aging brain both within and between clinical phenotypes over the adult lifespan. Consequently, a wealth of MRI data has been generated by diverse large[1-4] and medium[5-7] scale studies with clinical emphasis on aging and lifespan[8-10], neuro-degeneration[11, 12], psychiatric disorders[13, 14], genetics[15-17], lifestyle[2, 18], cardiovascular risk factors[19, 20], and chronic health conditions[21] collected from populations with different races[22], ethnicities[23], and geographic locations[2, 4]. However, datasets from these studies are typically analyzed independently, which both limits statistical power and, most importantly, restricts the ability to generalize outcomes among diverse disease groups and populations. There is an untapped potential in properly aggregating data across different studies and thereby creating a large, harmonized, and rich database of quantitative brain MRIs. This aggregation would provide a strong foundation for the deeper study of heterogeneity in brain aging and thereby increase the sophistication, reproducibility and generalizability of the resulting models. Pooling large MRI data sets across studies will provide optimal conditions for the large-scale training of machine learning (ML) models. Such models have shown great promise for diagnosis and prognostication[24] and now have begun to offer important insights into disease heterogeneity[25-27], thereby further improving the precision of diagnostics.

The application of common statistical and ML models across diverse studies is challenging in practice, due to variability in the data acquisition and image analysis methods employed in the field. Tools for standardized pre-processing[28-30], harmonization[31], normative modeling[32, 33], and ML models[34, 35] lack comprehensive integration across the entire data analysis chain. Moreover, the training and validation of ML models on small datasets numbering in the hundreds or few thousands of samples[24] raises concerns on robustness and generalization, since the dimensionality of the brain scans is orders of magnitudes higher[36].

Herein, we describe the NiChart platform and its application in the analysis of structural brain MRI data across cross-sectional and prospective studies of aging. This study focuses on adult aging after age 40, excluding children, adolescents, and young adults. The main contributions of NiChart lie in: a) making available normative statistical summarizations, pre-trained harmonization models and compatible ML models from a large and diverse pooled and harmonized dataset; b) developing, validating and freely providing computational tools for users to map their own study data to NiChart's reference space, thereby providing access to high-level imaging signatures via NiChart's ML models and comparing the data with both normative and disease-related populations; c) providing containerized state-of-the-art pre-processing pipelines for easy installation and complete reproducibility on resources ranging from a personal computer to a high performance computing cluster, as well as the NiChart web portal on the Amazon Web Services (AWS) cloud, which eliminates the need for any local installation (https://neuroimagingchart.com). Moreover, through preprocessing and harmonization tools, NiChart enables the growth of a community-driven arsenal of statistical and ML models, facilitating novel methodology dissemination and validation.

An overview of how NiChart integrates multi-center data and an illustration of how the final ML-based imaging signatures can be used to characterize individuals is shown in Fig. 1.

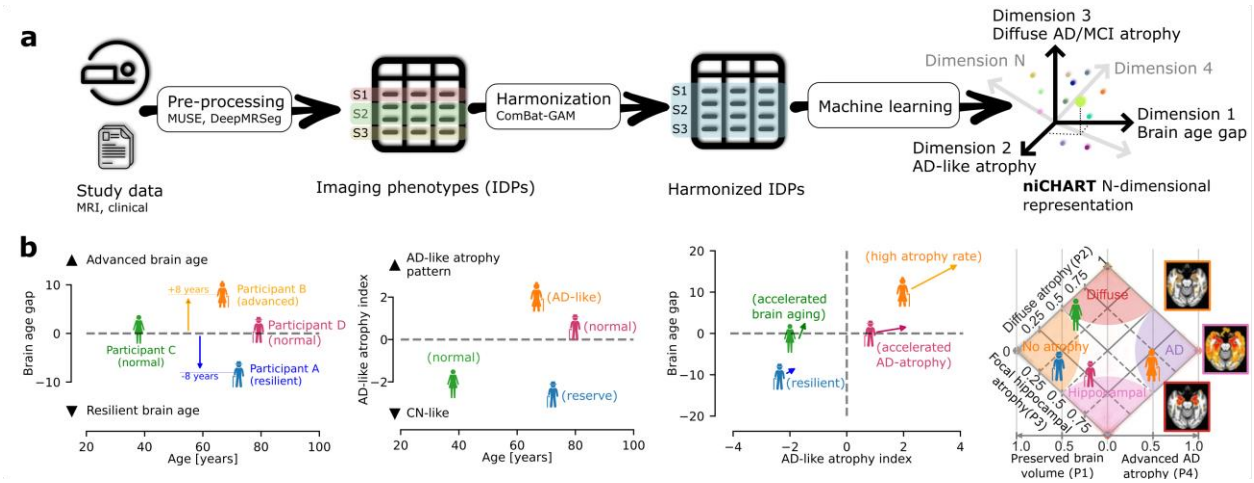

Fig. 1 ***Conceptual view of mapping and interpreting the brain features with NiChart. a.*** *NiChart transforms MRI imaging and demographic data from multiple studies--indicated here as S1, S2, and S3--into a dimensional representation via standard pre-processing to extract imaging derived phenotypes (IDPs), statistical harmonization of IDPs, and application of pre-trained machine learning models for further analysis of prospective data. The resulting charts capture normative ranges and multiple facets of brain health on distinct and interpretable representations.* ***b.*** *From left-to-right, the signatures reflect the brain age (SPARE-BA)[30] and AD-like atrophy (SPARE-AD)[31] indices, and the expression of AD disease sub-types (SMILE-GAN AD-subtypes, see text for further details)[22]. Brain imaging data from four hypothetical participants show varying degrees of expression of imaging signatures associated with these dimensions. Based on this data alone, Participant A, a 72-year-old man, has resilient brain age, lacks AD-like imaging signatures, and expresses the "No atrophy (P1)" AD-subtype pattern, . Participant B, a 68-year-old woman, shows signs of accelerated brain aging, AD-like imaging signatures, and expresses the "AD atrophy (P4)" subtype. Participant C, a 38-year-old woman, has no apparently accelerated brain aging no signs of AD-like imaging signatures, and is classified "No atrophy (P1)" subtype. Participant D, an 80-year-old man has no apparently accelerated brain aging no AD-like imaging signature, but expresses predominantly the "Focal hippocampal atrophy (P3)" subtype, potentially suggesting early signs of emerging AD.*

While recent studies have proposed data analytic pipelines to investigate normative distributions of regional brain volumes with age[32, 33], the focus and scope of the NiChart framework is broader. NiChart integrates state of the art, well-validated methods for image processing and image feature extraction[27, 31, 37-45] on a diverse and large-scale reference dataset currently consisting of 53,757 participants pooled from 23 different studies as part of the iSTAGING Study[26], an initiative to consolidate public and private structural MRI datasets (Table 1). Nonlinear statistical harmonization methods[31] reduce site differences while retaining biological characteristics. Supervised and semi-supervised machine learning (ML) methods were applied to the large, harmonized dataset to develop an ensemble of ML models, including models for accelerated brain aging[46], Alzheimer's disease (AD) related brain atrophy[38], and latent structural sub-types of AD[27], among others. These ML models establish the dimensional brain chart that characterizes structural brain changes due to aging and neurodegenerative diseases and map high-dimensional imaging data from individuals into a low-dimensional chart of expressive, yet interpretable imaging signatures.

2 Results

2.1 Characteristics of NiChart reference datasets and models

The current NiChart reference dataset was constructed from n=71,820 MR studies, including 18,063 longitudinal follow-up scans of adults age 40 years or older. MRI scans were acquired on more than 100 different scanners from 23 studies conducted over approximately 3 decades. Accompanying non-imaging data includes demographic information, major clinical risk factors, and disease-related variables available for all or a subset of included data (

Tab. S 1). Additionally, expansion of NiChart is currently ongoing, with upcoming releases planning to utilize the methods described herein to add data from 12 new studies consisting of 9257 individuals (see supplementary table S1 for details).

Parameters of the statistical harmonization model were estimated from the baseline scans of cognitively unimpaired (CU) individuals, using a dataset that includes n=31,341 samples. ML models were trained and cross-validated using data from subsets selected based on data availability for each specific task. The clinical classification scheme for constructing the task-specific samples is shown in Table 1. A list of all ML models and image processing components that are published alongside this manuscript is listed in Table 2.

Table 1 Simplified diagnostic categories.

Category	Description	Count
CU	*Cognitively unimpaired No apparent sign of cognitive, psychological, or neurological disorder. Sometimes assumed without cognitive evaluation.*	*30,361*
MCI	*Mild cognitive impairment: Formal assessment of MCI due to AD typically at baseline. No consideration of progression to AD. Some variation in the diagnostic workup across studies.*	*1501*
AD	*Dementia due to Alzheimer's disease: Formal clinical assessment of AD. Some variation in the diagnostic workup across studies.*	*1011*
CN (stable)	*No progression from CU to MCI: Participants with longitudinal clinical data who were classified as CU at baseline but did not progress to MCI.*	*3919*
CN to MCI	*Reported progression from CU to MCI: Participants with longitudinal clinical data who were initially classified as CU, and were diagnosed with MCI later in the study.*	*362*
MCI to AD	*Reported progression from MCI to AD: Participants with longitudinal clinical data who were initially diagnosed with MCI, and who were diagnosed with AD later in the study.*	*479*
MCI (censored)	*No progression from MCI to AD: Participants with longitudinal clinical data who were diagnosed with MCI, but never with AD.*	*952*
Hypertension	*Presence or absence of hypertension: Presence of hypertension based on study-specific criteria.*	*1526*

Table 2 NiChart image analysis methods, software and models.

Name	Description	Software Package
MUSE[44]	Anatomical segmentation of T1w MRI using multi-atlas methods.	https://github.com/CBICA/MUSE Installable software package with reference atlases and ROI dictionaries
DLMUSE[43]	Anatomical segmentation of T1w MRI using deep learning	https://github.com/CBICA/DLMUSE Installable Python software package with pre-trained model Also available as a software container and as a cloud application in https://neuroimagingchart.com
DLICV[43]	Skull-stripping of T1w MRI.	https://github.com/CBICA/DLICV Installable Python software package with pre-trained model Also available as a software container and as a cloud application in https://neuroimagingchart.com
DLWML[43]	White matter lesion segmentation using T2/FLAIR MRI.	https://github.com/CBICA/DeepMRSeg Installable Python software package with pre-trained model
NeuroHarmonize[31]	Statistical harmonization of imaging features	https://github.com/rpomponio/neuroHarmonize Installable Python software package
SPARE[38]	Supervised machine learning based prediction of aging and neuro-degenerative diseases	https://github.com/CBICA/spare-model Installable Python software package with pre-trained models (SPARE-AD, SPARE-BA)

		Also available as a software container and as a cloud application in https://neuroimagingchart.com
SmileGAN[27, 41]	Semi-supervised machine learning based prediction of Alzheimer's Disease subtypes	https://pypi.org/project/SmileGAN Installable Python software package with pre-trained model
NiChart	Tool with graphical user interface (GUI) for application of harmonization and machine learning steps on derived MRI features, and for data visualization	https://github.com/CBICA/NiChart; Installable software package with pre-trained models

2.2 Harmonization of NiChart imaging derived phenotypes

Automated processing pipelines (Section 4.2) were used to extract imaging derived phenotypes (IDPs). Anatomical segmentation provided volumes of 145 regions of interest (ROIs)[44]. Other IDPs included lobar white matter hyperintensity volumes, which are known to be associated with cerebrovascular disease[43], structural covariance networks[45], and voxel-based regional volumetric maps[42]. These IDPs were harmonized using neuroComBat-GAM[31], so that systematic biases due to inter-scanner variations were suppressed (Section 4.3). The batch variable, reflecting the scanner group of each sample, was the key target for harmonization. We determined batch values based on differences in studies, scan sites, and scanner settings. Additional covariates such as age, sex, and total intracranial volume (a proxy for head size) were accounted for in the harmonization model, using a smooth parametric generalized additive model (GAM) for nonlinear age effects, since they might otherwise confound disease effects. The harmonization model was estimated from the subset of cross-sectional CN subjects, and the estimated parameters were applied on the complete sample to harmonize all observations (e.g. those of MCI and AD patients, and longitudinal timepoints). An analogous out of-sample statistical harmonization methodology was offered so that a user can place new datasets/studies into NiChart.

Effectiveness of the harmonization process was evaluated by a comparison of the fitted trend lines of the imaging IDPs across image batches before and after harmonization. Age trends fitted with a linear GAM model for all studies indicate that bias related to specific site effects has been removed (Fig. 2). The mean squared error statistic (MSE) was used to assess the goodness of fit of this data across all batches to the selected model before and after harmonization. A smaller MSE value indicates that the data fits the central moment (mean) of the linear GAM model. In 139 out of 145 (95.8%) of the regions of interest used to train the harmonization model, the MSE decreased post-harmonization. 2

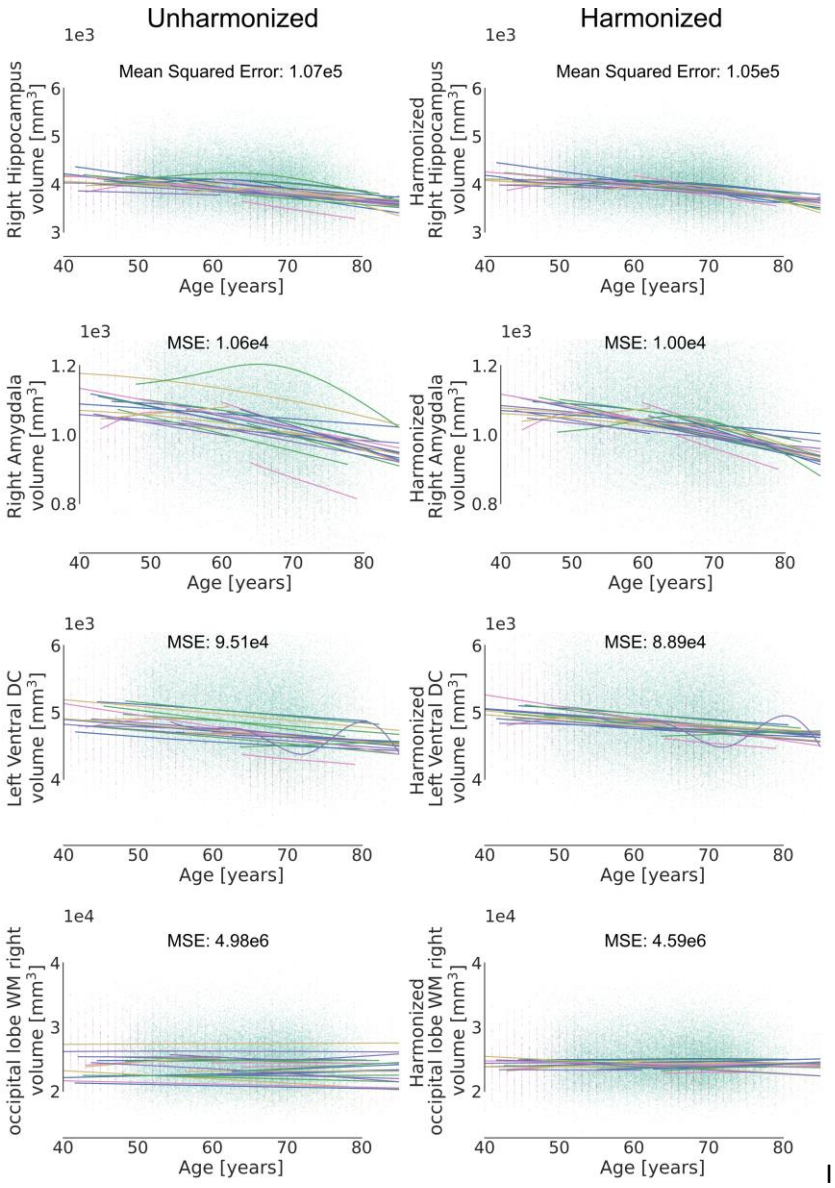

Fig. 2 Cross-sectional NiChart data before (left) and after (right) harmonization. Age trends of representative regions of interest with reduced batch effect after the harmonization process. A comparison of the Mean Squared Error (MSE) across batches pre- and post-, shows that the harmonized regions better fit a model than unharmonized regions. A smaller MSE indicates a better fit of the selected model, in this case, a linear GAM model. The sample that was used for creating the plots included cognitively unimpaired participants at baseline from all iSTAGING studies.

2.3 2Out-of-sample harmonization of a new dataset to NiChart reference model

We implemented an "out-of-sample harmonization" technique for harmonizing a previously unseen study to the harmonization model derived from the current cohort. Out-of-sample harmonization keeps the reference data unchanged, while estimating harmonization parameters that align the new study to the reference data (Section 4.3). We tested the robustness of the out-of-sample harmonization with two independent experiments. The first experiment used two splits of the UK Biobank dataset as independent data batches. The two batches from UK Biobank were collected with the same protocol and within the same population, and therefore, minimal batch differences were expected. On the ROI

level, the difference between the estimated location and scale parameters of the two batches from UK Biobank were smaller than between other batches (Fig. S1 A) and their distributions were narrower than the empirical Bayes priors (Fig. S1 B). The quantiles of the harmonized individual ROIs were consistent (Fig S1 C). The distributions of the extracted SPARE-AD and SPARE-BA scores also matched (Fig. S1 D and E).

In a second experiment we evaluated the accuracy of the out-of-sample harmonization on a study not included in initial NiChart harmonization model, namely the Pre-symptomatic Evaluation of Experimental or Novel Treatments for AD (PREVENT-AD)[47] study. PREVENT-AD consists of 343 cognitively unimpaired (CU) subjects with baseline scans obtained at 55-88 years of age and a total of 870 longitudinal time points. We harmonized PREVENT-AD to the reference harmonized data through out-of-sample harmonization, using baseline MRI scans for the estimation of the harmonization parameters.

Residuals of independent ROI values, after correcting biological covariates and batch effects, showed distributions similar to residuals of the reference dataset. To evaluate the effectiveness of the out-of-sample harmonization, we compared the homogeneity of variance between PREVENT-AD ROI values and reference dataset ROI values, using Levene's test for equality of variance. The values for the reference dataset were calculated from a random sub-sample with the same size as the PREVENT-AD dataset (n=1213). We performed Levene's test for each of the 145 ROIs, independently for PREVENT-AD ROI values before and after harmonization. In this test, 126 out of 145 ROIs (~87% of ROIs) obtained a p-value>0.05 after harmonization, showing that variances of PREVENT-AD and reference groups were similar. Figure 3 shows the distribution of neuro-imaging features before and after harmonization of the PREVENT-AD data, after mapping the values to a two-dimensional embedding using a t-distributed stochastic neighbor embedding (t-SNE, see figure caption for details).

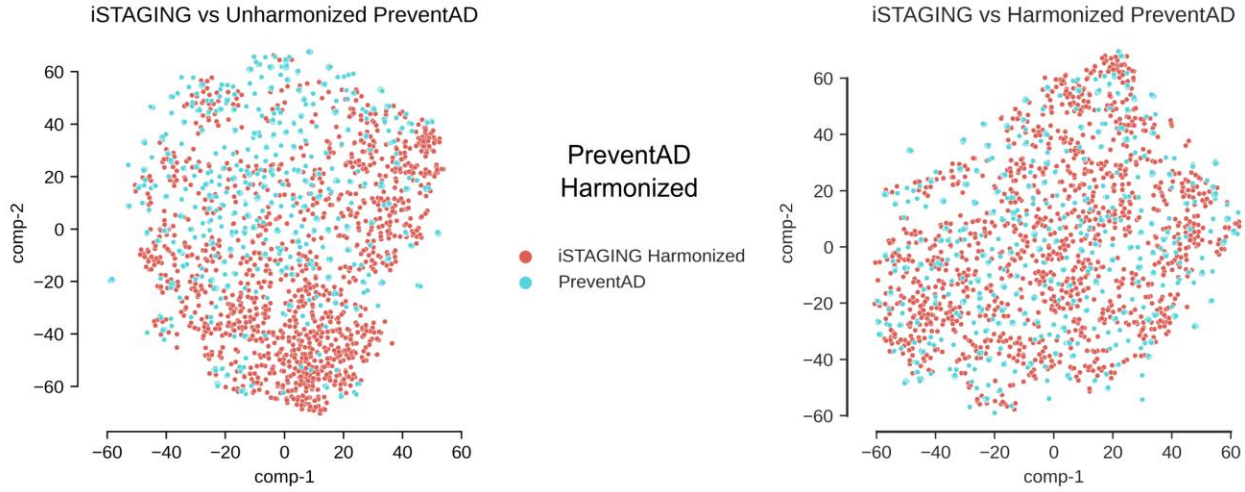

Fig. 3 ***Two-dimensional embedding of neuro-imaging features from PREVENT-AD and iSTAGING.*** *Visualization of the 145-dimensional ROI imaging data in two dimensions using t-distributed stochastic neighbor embedding (t-SNE). Each point represents the embedding derived from the ROI residuals for one participant, after correcting biological covariates (left) or both biological covariates and batch effects (right). PREVENT-AD data was compared to a randomly sampled subset of cognitively unimpaired participants of the iSTAGING data set. The distances between points in the t-SNE plot preserve actual distances between points in the 145-dimensional data. After harmonization*

of batch effects, the Silhouette score[48] was reduced from 0.401 to 0.396, indicating a higher overlap between the two datasets.

2.4 ML-based quantification of pre-clinical neurodegeneration and heterogeneity in AD

Harmonized ROI values were used for constructing ML-based imaging signatures, serving as biomarkers corresponding to dimensions of the NiChart coordinate system. For example, the SPARE-AD index, a supervised learning model derived by contrasting CN and AD individuals, quantifies typical AD-like atrophy patterns in a subject's brain MRI[49]. Previous studies have shown that SPARE-AD is an informative signature of Alzheimer's disease (AD)-related neurodegeneration and has predictive value in terms of clinical progression[50]. SPARE-BA quantifies atrophy patterns associated with brain age[26]. Here, we present illustrative examples of these two indices in UK Biobank data, to demonstrate the potential of NiChart dimensions to identify early brain changes in the general population and to predict progression of disease.

The UK Biobank study did not include participants with known cognitive impairment at baseline. However, a small subset of subjects was diagnosed with MCI or clinical AD related dementia (AD-dementia) during follow up. Fig 4a shows these subjects projected to the NiChart dimensions, in comparison to complete UK Biobank sample. These individuals had on average 0.9 unit higher SPARE-AD index and a 7.8-year larger brain age gap, showing that SPARE-AD scores of CN subjects at baseline are sensitive to later MCI and AD-dementia diagnosis.

In the second analysis, NiChart used MRIs of 2088 individuals in the iSTAGING dataset who had baseline MCI and were diagnosed with AD-dementia during follow-up. The goal of this analysis was to investigate disease heterogeneity within this sample. To do this, we focused on the four distinct imaging patterns of mild cognitive impairment (MCI) and Alzheimer's disease (AD) captured by SMILE-GAN's AD-subtype signatures. The 4 SMILE-GAN AD-subtypes are referred to as P1 (preserved volume), P2 (mild diffuse atrophy), P3 (focal medial temporal lobe atrophy), and P4 (widespread AD-like atrophy)[27]. The 4 subtype scores (P1-P4) showed distinct trajectories of clinical progression (4b).

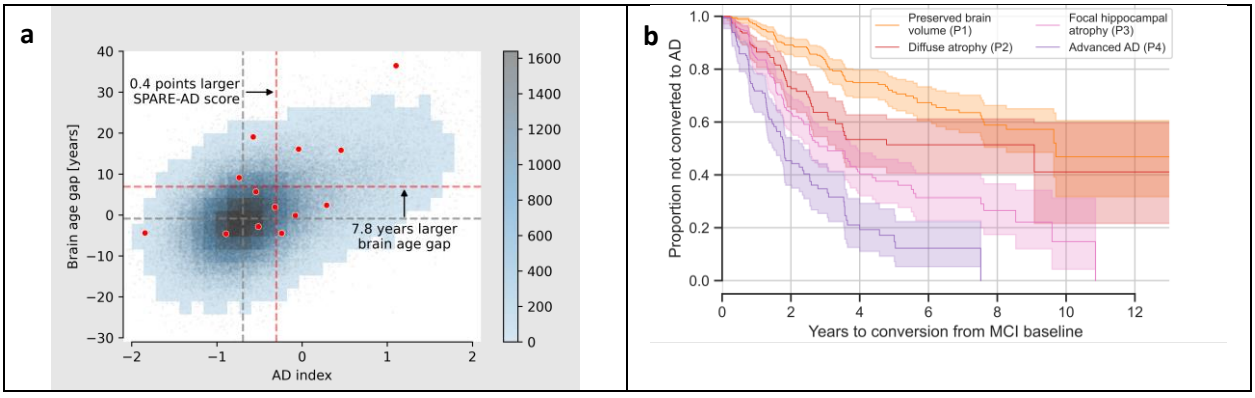

Fig. 4 ***Cross-sectional signatures and their clinical correlates. a.*** *Placement on the AD-like atrophy index/brain age gap 2-dimensional coordinate system of participants who had an incidental report of dementia (red dots) in the UK*

Biobank. The participants to the top and the right of the center indicated by the crosshair of the gray dashed lines have advanced brain aging and greater AD-like brain atrophy relative to the average of the population. ***b.*** *Predictive value of four ML-defined imaging signatures of brain atrophy, in terms of progression from MCI to clinical AD[27]. Expression of these imaging signatures at baseline predicts subsequent clinical trajectories.*

2.5 Temporal trajectories of AD-related atrophy patterns

NiChart provides an opportunity to analyze temporal trajectories of individuals within the dimensional system of ML-based imaging indices. We investigated trajectories of SPARE-AD and SPARE-BA scores from different participant groups defined by clinical disease categories. Participant groups exhibited differences in the rate of progression along NiChart dimensions (Figure 5). Cognitively unimpaired participants who did not progress to MCI (CN-stable) had the slowest trajectories along the SPARE-AD and SPARE-BA dimensions . Participants who progressed from MCI to AD-dementia had, on average, SPARE-AD rates of change four standard deviations higher than CN-stable subjects (Fig 5.b). AD-dementia patients had SPARE-AD rates of change like that in the MCI to AD-dementia converter group, but a lower intercept. While the SPARE-BA score was less sensitive in detecting differences between AD-dementia diagnostic groups, it did show group differences, suggesting the importance of multi-dimensional characterization of disease in NiChart.

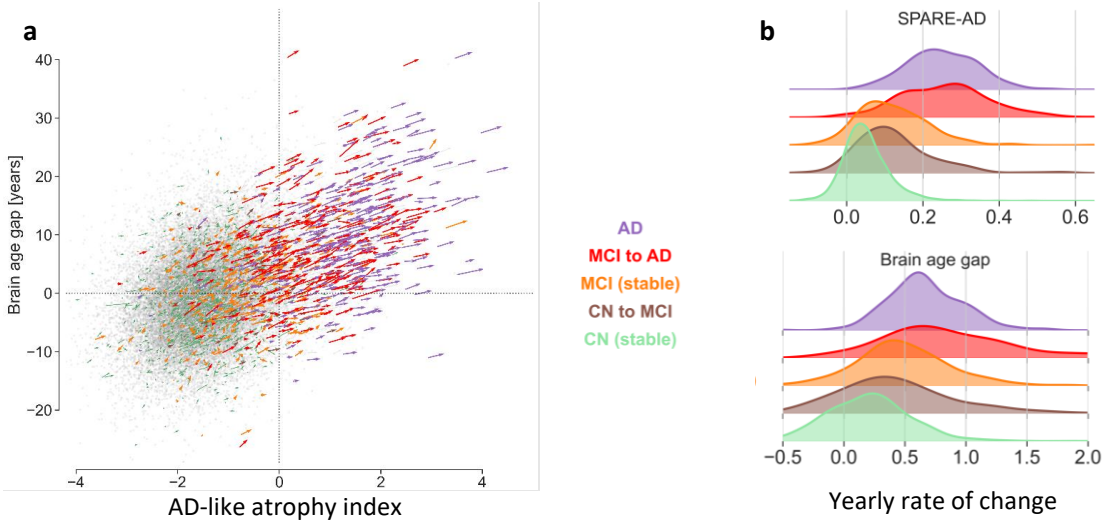

Fig. 5 Longitudinal rate of change of SPARE-AD and SPARE-BA signatures for participants with five or more MRI visits and adjudication for cognitive status along the Alzheimer's disease (AD) spectrum. ***a.*** *Each arrow represents the vector, magnitude, and direction, of the fitted annual change of one participant in the coordinate system spanned by the SPARE-AD and SPARE-BA signatures. The analysis show that cognitively unimpaired participants (CU, green) and participants with mild cognitive impairment (MCI) at baseline who do not progress to AD-dementia (MCI stable, orange) have a lower annual change in their SPARE scores compared to participants who did progress to AD-dementia (MCI to AD-dementia, red). AD-dementia cases had the highest SPARE-AD at baseline but lower longitudinal progression, likely due to a ceiling effect.* ***b.*** *Diagnostic groups vary in rate of change of SPARE scores. Histograms demonstrate that participants who progress from MCI to AD-dementia (red) have the highest increase in SPARE-BA and SPARE-AD scores.*

3 Discussion

We have described NiChart, a brain chart and associated software, which provides a framework for standardization of structural neuroimaging measurements throughout the adult lifespan. The main strengths of NiChart are the 1) large and diverse datasets from which normative brain aging models were derived; 2) advanced set of image analysis methods and associated software, including multi-atlas segmentation and statistical harmonization pipelines; 3) ability to harmonize new studies into the NiChart coordinate system, thereby enabling the dynamic, community-driven enrichment of normal and abnormal brain aging models; 4) emphasis on ML imaging signatures, which have recently offered great promise as individualized diagnostic and prognostic markers and as tools for unravelling disease-related heterogeneity; 5) easy access via a cloud-based drag-and-drop interface (https://neuroimagingchart.com). The current implementation of NiChart focuses on morphometric brain features, as well as on ML imaging signatures of brain aging and AD. These models were derived from a large, diverse dataset of MRI scans of 71,820 MRI time points from 53,757 participants and 23 studies.

The dimensions of NiChart capture and model variation of *in vivo* structural brain features at different levels of granularity: from basic features, such as total gray matter volume, and regional volumes, to more complex ones such as ML models of MCI/AD and brain age. Some of these latter features reflect different aspects of the heterogeneity of typical and diseased brain aging trajectories that were previously not provided by cross-sectional brain charts[32, 33]. Moreover, the containerized NiChart software provides integrated tools allowing multi-atlas ROI labeling, harmonization, and ML model application.

Most importantly, NiChart provides the foundation for community-driven construction of additional models of diagnostic and predictive value across neurologic diseases and psychiatric disorders tailored to the needs of individual users. In particular, after mapping data of independent studies into the NiChart coordinate system via consistent image processing and harmonization tools, new ML models derived from out-of-sample cohorts will be applicable using NiChart pipelines, and hence will offer the ability to measure novel neuroanatomical dimensions.

In the search for better biomarkers, brain charts became a useful tool in neuroscience in recent years. For example, in AD, deviations of morphometric features, such as the volume of any disease relevant anatomical region, from estimated normative aging trajectories may be indicative of underlying abnormal neurobiological processes which can be related to genetics, lifestyle, or adverse events, such as traumatic brain injury. Even more importantly, ML derived imaging signatures, which are a focus of the current platform, have the potential to be more predictive, sensitive, and specific than individual ROI variables because they often encode subtle, yet complex patterns related to disease/exposures[39, 51, 52]

This study has several limitations that should be considered when interpreting the results. The framework for harmonization of new data sets relies on the availability of a reference sample of comparable participants to adjust for study-specific batch effects. Incomplete or poor quality adjunct data may impair translation of some models that use that data. In our current formulation, we used cognitively unimpaired individuals in this process, with age, sex and ICV as adjusted covariates. It is important to note that a limitation of this process is that the sample must lie within the age range of the

main training dataset. We assume that any external study shares the same covariate model as the training dataset in the out-of-sample harmonization process, therefore, having a similar age distribution will ensure better harmonization results. However, comparable cohorts for out of sample harmonization could be defined based on other features in the iSTAGING cohort, such as race, clinical conditions (e.g. diabetes), genetics (e.g. by using individuals with certain genetic characteristics), and other factors that allow researchers to most accurately map their data into NiChart anchored by appropriate sub-cohorts of NiChart.

NiChart's ML models, built from structural MRI scans, were specifically trained for AD and brain aging. New analytic models, especially ML models, as well as functional connectivity data and derived models, are currently being added to NiChart in the same systematic fashion, thereby enriching our understanding and quantification of brain aging. The broader iSTAGING dataset includes data from populations with diabetes[21, 53], hypertension[20], obesity, low socioeconomic status[22], exposure to harmful chemicals like organolead[54], and populations from diverse ethnic backgrounds[23]. Accordingly, upcoming NiChart releases will enrich the ML toolbox of NiChart with statistical and ML models tailored to these variables.

NiChart aims to provide significant contributions toward precision diagnostics. By providing the means for unravelling heterogeneity of brain aging, and by offering a rich set of dimensions that probe complementary aspects of brain structure, NiChart offers the opportunity to refine our ability to capture spatial imaging patterns, and hence to differentiate among individuals following different brain aging trajectories. It also allows us to investigate disease patterns in relation to typical brain aging patterns in the absence of specific disease, which provide a background against which disease related changes occur. Ultimately, the position and trajectory of an individual within this coordinate system might offer a rich, yet compact representation of brain structure and function. Moreover, modifying factors, including treatments, can be evaluated more precisely in terms of their ability to modify an individual's trajectory within the NiChart reference space.

4 Methods

4.1 Software implementation

Different components of the NiChart framework are implemented as modular, minimally coupled software packages to facilitate installation, usage and maintenance.

Image processing, data harmonization and ML methods, and associated pre-trained models are provided as open source software. We also provide supplementary tools for visualization of image features and biomarkers derived from user datasets together with ranges and distributions from the NiChart reference datasets, to allow the users to compare their selected data to different disease or demographic groups, or to normative distributions.

The web portal leverages deep-learning (DL)-based image pre-processing modules, software containerization and AWS parallelization to empower users with limited technical expertise or processing power with a platform to process their data using a simple drag-and-drop interface.

All NiChart distributions -- source code, containerized, or web-accessible -- are based on the same codebase available via GitHub, ensuring consistent, reproducible, comparable results.

4.2 NiChart Image Processing

Structural MRI images were pre-processed with a containerized pipeline to compute regional volumes[44]. A deep convolutional neural network, DeepMRSeg[43], was used to compute the intra-cranial volume and to segment white matter lesions from T1-weighted and T2-FLAIR images. Before computing regional volumes, the T1-weighted scan of each subject was preprocessed for correction of intensity inhomogeneities. We applied a multi-atlas skull stripping algorithm using a large set of reference atlases for robust and accurate extraction of the brain tissues[55]. To maintain longitudinal consistency, the brain mask at each scan time point was combined with a probabilistic brain mask that was propagated from the baseline image through deformable registration using ANTS[56] (software available at http://stnava.github.io/ANTs/). Baseline SPGR and MPRAGE scans were selected independently during this process. Skull stripped T1 images were bias corrected using *FAST[57].* We segmented the brain into a set of anatomical regions of interest (ROIs) using a multi-atlas label fusion method, MUSE[44], which was the top-ranking method in the MICCAI-SATA challenge on deep brain segmentation[58], and produced more consistent results at varying age and across field strengths[59].

Instead of applying MUSE independently to each time point, i.e., by registering/warping ROI labels from reference atlases directly to each time point, we followed a pseudo-4D approach, by propagating atlases that were warped to baseline image space into each follow-up time point, thus imposing a more unified registration path before the fusion of reference labels. Regional volumes are calculated for each ROI, as well as in larger anatomical regions obtained by grouping single ROIs within a hierarchical representation.

4.3 Statistical Harmonization of Imaging IDPs

The ComBat-GAM model[31] builds and expands upon the linear ComBat model for mitigating study (batch) differences by estimating the biological variations attributed to age, sex, and intracranial volume and unwanted batch variations expressed via additive and multiplicative effects in the residuals for each of the 145 primary brain morphometry measures (Figure 6). Nonlinear age effects were accounted for

using a smooth parametric generalized additive model (GAM) with thin plate regression splines basis expansion.[46] The batch-specific location and scale parameters that capture systematic technical variability not explained by the biological covariates are estimated on the residuals Z_{ijr} for batch i, subject j and IDP r, after removing the biological effects, where Z_{ijr}= $\frac{1}{\hat{\sigma}_r}\left(y_{ijr} - \hat{\alpha}_r - f(X|\beta_r)\right)$. The ComBat model uses empirical Bayes adjustment to estimate the batch-specific location shift γ_{ir}^* and scale shift δ_{ir}^{2*}. It allows shrinkage and shared information borrowed across imaging IDPs[60] and express the batch parameters as a weighted sum between within-batch mean and variance for a specific IDP, and the overall statistics across IDPs:

$$\gamma_{ir}^* = \frac{n_i\bar{\tau}_i^2\hat{\gamma}_{ir} + \delta_{ir}^{2*}\bar{\gamma}_i}{n_i\bar{\tau}_i^2 + \delta_{ir}^{2*}}; \quad \delta_{ir}^{2*} = \frac{\bar{\theta}_i + \frac{1}{2}\sum_j (Z_{ijr} - \gamma_{ir}^*)^2}{\frac{n_j}{2} + \bar{\lambda}_i - 1} \qquad (1)$$

where $\bar{\gamma}_i = \frac{1}{R}\sum_r \hat{\gamma}_{ir}$, $\hat{\gamma}_{ir} = \frac{1}{n_i}\sum_j Z_{ijr}$. Once the batch parameters are estimated, they are removed from the raw IDPs to obtain harmonized values as:

$$y_{ijr} = \frac{\hat{\sigma}_r}{\hat{\delta}_{ir}}\left(Z_{ijr} - \hat{\gamma}_{ir}^*\right) + \hat{\alpha}_r + f(X|\beta_r).$$

Out-of-sample Harmonization. Note that in the previous steps, only the mean covariate model $f(X|\beta_r)$ and pooled residual variance $\hat{\sigma}_r$ across studies are needed. We can estimate batch parameters for the data from the new batch without having access to the original reference data, with the assumption that the mean covariate model and the pooled residual variance are the same. Hence, we developed the out-of-sample data harmonization strategy for new study batches k using the similar Empirical Bayes steps as in formula (1) after obtaining the standardized residuals Z_{kjr} as $\frac{1}{\hat{\sigma}_r}\left(y_{ijr} - \hat{\alpha}_r - f(X|\beta_r)\right)$. To minimize the sample bias due to non-technical sources, the subpopulation used to estimate the batch parameters should be as close to the reference population as possible. Not all studies within iSTAGING had an equally thorough screening for these criteria. Participants with conditions that affect brain morphometry to a small extent (such as diabetes and hypertension) were therefore not excluded and had varying prevalence across studies. This induced some heterogeneity in the reference population. Of note, in studies without a comparable reference population, this may lead to a sample bias, yet within-sample comparisons remain valid.

In addition to removing batch effects, the covariate model of the ComBat-GAM harmonization is also useful to compute (partial) residuals to reduce biological variance such as that from age, sex, and intracranial volume. The cross-sectional model fitted to data from multiple studies covers the age range between 20 and 102 years and can thus be applied to data across the entire adult lifespan.

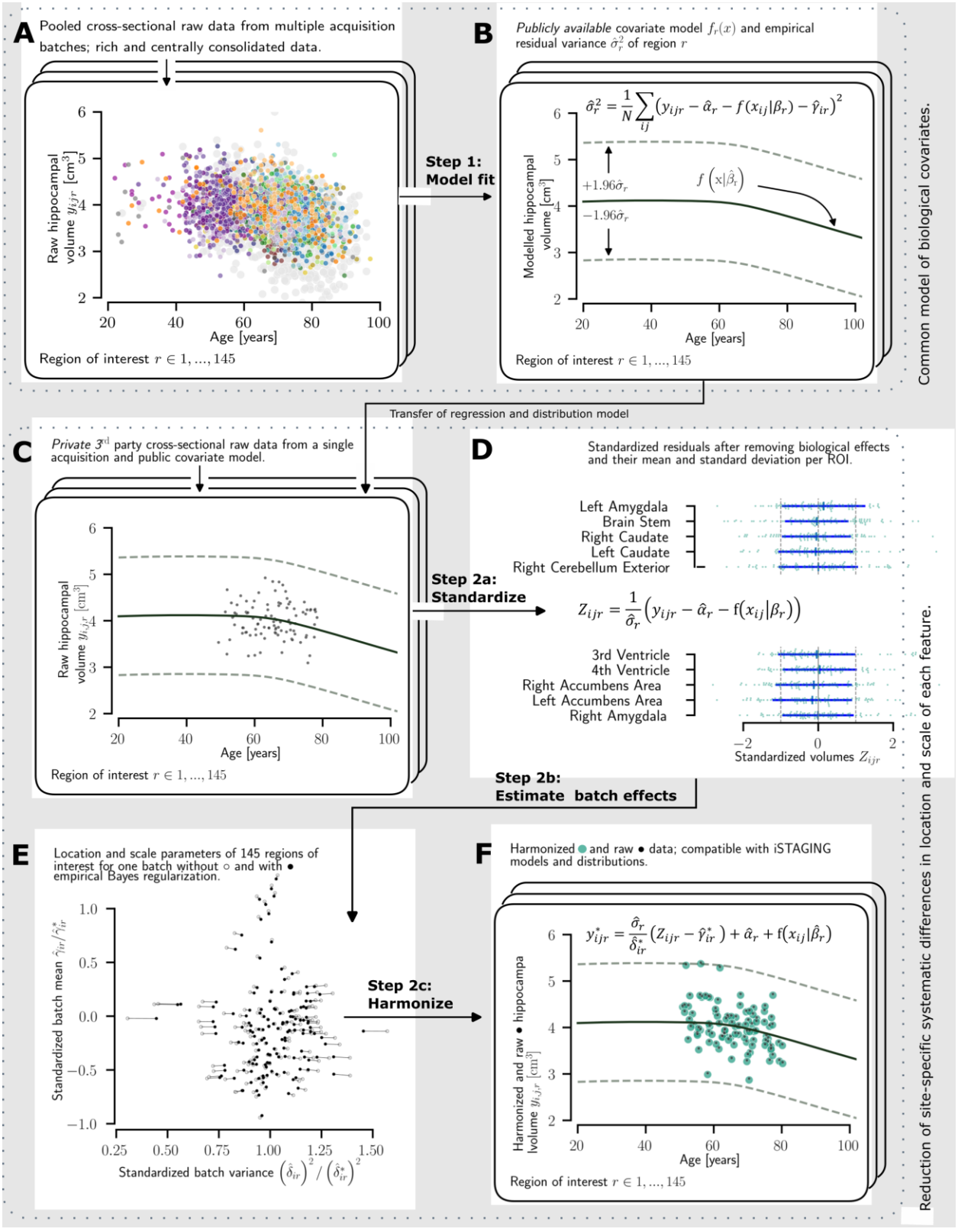
A
Pooled cross-sectional raw data from multiple acquisition batches; rich and centrally consolidated data.
Raw hippocampal volume y_{ijr} [cm^3]
Age [years]
Region of interest $r \in 1, ..., 145$
Step 1: Model fit
B
Publicly available covariate model $f_r(x)$ and empirical residual variance $\hat{\sigma}_r^2$ of region r
$\hat{\sigma}_r^2 = \frac{1}{N}\sum_{ij}\left(y_{ijr} - \hat{\alpha}_r - f(x_{ij}|\beta_r) - \hat{\gamma}_{ir}\right)^2$
$+1.96\hat{\sigma}_r$
$-1.96\hat{\sigma}_r$
$f\left(x|\hat{\beta}_r\right)$
Modelled hippocampal volume [cm^3]
Age [years]
Region of interest $r \in 1, ..., 145$
Common model of biological covariates.
Transfer of regression and distribution model
C
*Private 3*rd party cross-sectional raw data from a single acquisition and public covariate model.
Raw hippocampal volume y_{ijr} [cm^3]
Age [years]
Region of interest $r \in 1, ..., 145$
Step 2a: Standardize
D
Standardized residuals after removing biological effects and their mean and standard deviation per ROI.
Left Amygdala
Brain Stem
Right Caudate
Left Caudate
Right Cerebellum Exterior
$Z_{ijr} = \frac{1}{\hat{\sigma}_r}\left(y_{ijr} - \hat{\alpha}_r - f(x_{ij}|\beta_r)\right)$
3rd Ventricle
4th Ventricle
Right Accumbens Area
Left Accumbens Area
Right Amygdala
Standardized volumes Z_{ijr}
Step 2b: Estimate batch effects
E
Location and scale parameters of 145 regions of interest for one batch without ○ and with ● empirical Bayes regularization.
Standardized batch mean $\hat{\gamma}_{ir}/\hat{\gamma}^*_{ir}$
Standardized batch variance $\left(\hat{\delta}_{ir}\right)^2 / \left(\hat{\delta}^*_{ir}\right)^2$
Step 2c: Harmonize
F
Harmonized ● and raw ● data; compatible with iSTAGING models and distributions.
$y^*_{ijr} = \frac{\hat{\sigma}_r}{\hat{\delta}^*_{ir}}\left(Z_{ijr} - \hat{\gamma}^*_{ir}\right) + \hat{\alpha}_r + f(x_{ij}|\hat{\beta}_r)$
Harmonized and raw ● hippocampal volume y_{ijr} [cm^3]
Age [years]
Region of interest $r \in 1, ..., 145$
Reduction of site-specific systematic differences in location and scale of each feature.

Figure 6 ***Elements of the statistical harmonization. a.*** *A large and diverse pool of cognitively unimpaired participants from multiple studies spanning the adult life span was used in* ***Step 1*** *to fit (B) a model $f(x|\hat{\beta})$ of biological covariates age, sex, and intracranial volume and estimate the residual variance $\hat{\sigma}_r^2$ for each region of interest r. The learned model is combined with prospective data from cognitively unimpaired participants from a single batch* ***c*** *and the data is standardized by subtracting from each participant the mean $f(x|\hat{\beta}_r)$ and dividing by $\hat{\sigma}_r$ yielding the residuals shown in* ***d****. Since the effects of known biological covariates were removed, the batch-specific deviation of the residuals from a unit Normal distribution is assumed to be due to systematic technical batch effects, such as from a specific scanner sequence.* ***e*** *To avoid over-correction, the estimates of location and shift parameters are regularized via empirical Bayes[47] resulting in harmonized features (****f****) that are compatible with iSTAGING models and with models trained with data that was harmonized as described. The raw data from the initial model fit (****a-b****) is not required for harmonization of new data (****c-f****).*

4.4 ML models

Using the harmonized and standardized regional volumes and demographic and clinical data as input features, we trained a set of machine learning models to derive the indices of advanced aging, AD-typical atrophy, and MCI/AD subtypes. These models map the multi-variate input data to a scalar score or a probability distribution that summarizes the expression of brain atrophy patterns associated with the target disease or condition. The output score is used for charting the NiChart dimensional system2. The methodologies underpinning the NiChart ML models are generic and can be extended to generate novel measures for other diseases or risk factors. For example, near-future planned extensions include ML models for cardio-metabolic risk factors and neuro-psychiatric disorders.

SPARE-BA score. To build this model, we identified cognitively unimpaired controls based on study-specific criteria. We excluded participants with known neurological abnormalities, but not those with common co-pathologies — such as hypertension and psychiatric disorders including depression – to obtain a model that captured aging of cognitively unimpaired participants as opposed to clinically measurable disease-free aging. To train the model, we included the sex-, and ICV adjusted regional volumes from the first MRI visit. To avoid biasing the models with the relatively larger sample of the UK Biobank dataset, a smaller subset was randomly selected. Although complex deep learning architectures can be used, we have found that most informative brain age models are obtained by modest fits[37]. Moreover, we have determined that conventional ML methods, like support vector regression (SVR), provide more continuously distributed brain age residuals. We therefore trained a SVR model using a five-fold nested cross-validation for hyperparameter optimization. Within each fold, we adjusted linearly for the age bias that leads to overestimation of young age and underestimation of old age. The final brain age score was pooled from the 25 SVR models, after bias correction and subtraction of the chronological age. The cross-validated mean absolute error (MAE) of the age prediction was 5.6 years. Non-linear SVR (MAE=5.4 years) and deep learning (MAE=2.9 years) predicted the age more accurately, but were less sensitive to detect unspecific neurodegeneration[48] and are therefore less useful for the purpose of charting.

SPARE-AD score. To compute this score[38], we constructed a training set that consists of a sub-set of CU controls with the same criteria as for the brain age prediction, and participants with probable dementia due to AD based on data available from the source studies. The model was trained using the age-, sex-, ICV and harmonized regional volumes from either the first visit—for participants classified as AD, or last visit—for participants classified as CU, as input features. We trained an ensemble consisting of five repeated five-fold linear support vector classifiers with nested hyperparameter optimization. The

index of AD-typical atrophy is the average decision value of the ensemble. The cross-validated balanced classification accuracy of CU controls versus probable AD was 0.89 and that of CU versus MCI was 0.71.

Pathology-tuned SPARE-AD and SPARE-BA models. We trained alternative AD and brain age models using samples restricted by the amyloid status of subjects, in order to facilitate AD-focused studies[39]. In particular, the training sets included only amyloid-positive AD cases as the disease sample, and only amyloid-negative CU controls, and hence resulted in biologically-, rather than clinically-defined imaging signature of AD.

SMILE-GAN AD-subtypes. We trained a weakly supervised latent variable model to characterize four patterns of MCI/AD with a deep generative adversarial artificial neural network, as previously described[27]. The inputs to the network were harmonized regional volumes, and the covariates were age, sex, and ICV, with clinical labels of CN for the reference dataset and MCI/dementia for the target data sets. The covariates were used to initially regress out effects of no interest. The output of the model was a probabilistic score that indicates the likelihood of belonging to a disease subtype.

5 Acknowledgements

The iSTAGING study is a multi-institutional effort funded by NIA through RF1 742 AG054409. iSTAGING statistical harmonization work was supported by R0MH123550. NiChart software development was funded by the NIH Brain Initiative project under U24NS130411. The Baltimore Longitudinal Study of Aging neuroimaging study is funded by the Intramural Research Program, National Institute on Aging, National Institutes of Health and by HHSN271201600059C. The BIOCARD study is in part supported by NIH grant U19-AG033655. The PHENOM study is funded by NIA grant R01MH112070 and by the PRONIA project as funded by the European Union 7th Framework Program grant 602152. Other supporting funds are 5U01AG068057, 1U24AG074855, R01MH119219 and S10OD023495. Data were provided [in part] by OASIS OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AA was funded through grants 191026. The PREVENT-AD study was funded through McGill University, the FRQ-S, the CIHR, Pfizer Canada, Brain Canada, the Alzheimer's Society of Canada, the Canada Fund for Innovation, the J.L. Levesque Foundation, and the Lemaire Foundation. This research has been conducted using the UK Biobank Resource under Application Number 35148. ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012) are funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the

University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. MESA is supported by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS). Brain MRI is supported by grant R01 HL127659 from the National Heart, Lung, and Blood Institute with additional support from the National Institute on Aging and grant R01 AG080821. The Study of Health in Pomerania (SHIP) is part of the Community Medicine Research net (CMR) (http://www.medizin.uni-greifswald.de/icm) of the University Medicine Greifswald, which is supported by the German Federal State of Mecklenburg- West Pomerania. MRI scans in SHIP and SHIP-TREND have been supported by a joint grant from Siemens Healthineers, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania.

Conflicts of Interest:

HJG has received travel grants and speakers honoraria from Neuraxpharm, Servier, Indorsia and Janssen Cilag.

6 References

1. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NEUROIMAGE* **166,** 400–424; 10.1016/j.neuroimage.2017.10.034 (2018).
2. Bookheimer, S. Y. *et al.* The Lifespan Human Connectome Project in Aging: An overview. *NEUROIMAGE* **185,** 335–348; 10.1016/j.neuroimage.2018.10.009 (2019).
3. Harms, M. P. *et al.* Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NEUROIMAGE* **183,** 972–984; 10.1016/j.neuroimage.2018.09.060 (2018).
4. Ellis, K. A. *et al.* The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International psychogeriatrics* **21,** 672–687; 10.1017/S1041610209009405 (2009).
5. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* **19,** 1523–1536; 10.1038/nn.4393 (2016).

6. Mueller, S. G. *et al.* The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America* **15,** 869-77, xi-xii; 10.1016/j.nic.2005.09.008 (2005).

7. Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C. & Buckner, R. L. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J Cogn Neurosci* **22,** 2677–2684; 10.1162/jocn.2009.21407 (2010).

8. Habes, M. *et al.* Advanced brain aging: relationship with epidemiologic and genetic risk factors, and overlap with Alzheimer disease atrophy patterns. *Translational Psychiatry* **6,** e775; 10.1038/tp.2016.39 (2016).

9. Waldstein, S. R. *et al.* Differential Associations of Socioeconomic Status With Global Brain Volumes and White Matter Lesions in African American and White Adults: the HANDLS SCAN Study. *Psychosomatic Medicine* **79,** 327–335; 10.1097/PSY.0000000000000408 (2017).

10. Jack, C. R. *et al.* The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of magnetic resonance imaging : JMRI* **27,** 685–691; 10.1002/jmri.21049 (2008).

11. Chand, G. B. *et al.* Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. *Brain : a journal of neurology* **143,** 1027–1038; 10.1093/brain/awaa025 (2020).

12. Eavani, H. *et al.* Heterogeneity of structural and functional imaging patterns of advanced brain aging revealed via machine learning methods. *Neurobiology of aging* **71,** 41–50; 10.1016/j.neurobiolaging.2018.06.013 (2018).

13. Smith, S. M. *et al.* Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *eLife Sciences Publications, Ltd* (2020).

14. Morris, M. C. *et al.* MIND diet slows cognitive decline with aging. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **11,** 1015–1022; 10.1016/j.jalz.2015.04.011 (2015).

15. Sweeney, M. D. *et al.* Vascular dysfunction-The disregarded partner of Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **15,** 158–167; 10.1016/j.jalz.2018.07.222 (2019).

16. Trofimova, O. *et al.* Brain tissue properties link cardio-vascular risk factors, mood and cognitive performance in the CoLaus|PsyCoLaus epidemiological cohort. *Neurobiology of aging* **102,** 50–63; 10.1016/j.neurobiolaging.2021.02.002 (2021).

17. Espeland, M. A. *et al.* Sex-Related Differences in Brain Volumes and Cerebral Blood Flow Among Overweight and Obese Adults With Type 2 Diabetes: Exploratory Analyses From the Action for Health in Diabetes Brain Magnetic Resonance Imaging Study. *The journals of gerontology. Series A, Biological sciences and medical sciences* **75,** 771–778; 10.1093/gerona/glz090 (2020).

18. Austin, T. R. *et al.* Association of Brain Volumes and White Matter Injury With Race, Ethnicity, and Cardiovascular Risk Factors: The Multi-Ethnic Study of Atherosclerosis. *Journal of the American Heart Association* **11,** e023159; 10.1161/JAHA.121.023159 (2022).

19. Arbabshirani, M. R., Plis, S., Sui, J. & Calhoun, V. D. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Promises and pitfalls. *NEUROIMAGE* **145,** 137–165; 10.1016/j.neuroimage.2016.02.079 (2017).

20. Ferreira, D. *et al.* Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. *Sci Rep* **7,** 46263; 10.1038/srep46263 (2017).

21. Habes, M. *et al.* The Brain Chart of Aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **17,** 89–102; 10.1002/alz.12178 (2021).

22. Yang, Z. *et al.* A deep learning framework identifies dimensional representations of Alzheimer's Disease from brain structure. *Nat Commun* **12,** 7065; 10.1038/s41467-021-26703-z (2021).

23. 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; 10.1038/s41592-022-01458-7 (2022).

24. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16,** 111–116; 10.1038/s41592-018-0235-4 (2019).

25. Fischl, B. FreeSurfer. *NEUROIMAGE* **62,** 774–781; 10.1016/j.neuroimage.2012.01.021 (2012).

26. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NEUROIMAGE* **208,** 116450; 10.1016/j.neuroimage.2019.116450 (2020).

27. Bethlehem, R. *et al.* Brain charts for the human lifespan. *bioRxiv,* 2021.06.08.447489; 10.1101/2021.06.08.447489 (2021).

28. Rutherford, S. *et al.* Charting Brain Growth and Aging at High Spatial Precision. *bioRxiv,* 2021.08.08.455487; 10.1101/2021.08.08.455487 (2021).

29. Varoquaux, G. & Cheplygina, V. How I failed machine learning in medical imaging -- shortcomings and recommendations, 3/18/2021.

30. Franke, K. & Gaser, C. *Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? Front. Neurol.*, 10, 789. Available at https://www.frontiersin.org/articles/10.3389/fneur.2019.00789/pdf (Frontiers, 2019).

31. Davatzikos, C., Xu, F., An, Y., Fan, Y. & Resnick, S. M. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain : a journal of neurology* **132,** 2026–2035; 10.1093/brain/awp091 (2009).

32. Hwang, G. *et al.* Disentangling Alzheimer's disease neurodegeneration from typical brain aging using machine learning, 9/8/2021.

33. Wen, J. *et al.* Multidimensional representations in late-life depression: convergence in neuroimaging, cognition, clinical symptomatology and genetics, 10/21/2021.

34. Doshi, J. *et al.* MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *NEUROIMAGE* **127,** 186–195; 10.1016/j.neuroimage.2015.11.073 (2016).

35. Doshi, J., Erus, G., Habes, M. & Davatzikos, C. DeepMRSeg: A convolutional deep neural network for anatomy and abnormality segmentation on MR images, 7/3/2019.

36. Davatzikos, C., Genc, A., Xu, D. & Resnick, S. M. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *NEUROIMAGE* **14,** 1361–1369; 10.1006/nimg.2001.0937 (2001).

37. Sotiras, A., Resnick, S. M. & Davatzikos, C. Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization. *NEUROIMAGE* **108,** 1–16; 10.1016/j.neuroimage.2014.11.045 (2015).

38. Franke, K., Ziegler, G., Klöppel, S. & Gaser, C. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *NEUROIMAGE* **50,** 883–892; 10.1016/j.neuroimage.2010.01.005 (2010).

39. Williamson, J. D. *et al.* The Action to Control Cardiovascular Risk in Diabetes Memory in Diabetes Study (ACCORD-MIND): rationale, design, and methods. *The American journal of cardiology* **99,** 112i-122i; 10.1016/j.amjcard.2007.03.029 (2007).

40. Nasrallah, I. M. *et al.* Association of Intensive vs Standard Blood Pressure Control With Cerebral White Matter Lesions. *JAMA* **322,** 524–534; 10.1001/jama.2019.10551 (2019).

41. Doshi, J., Erus, G., Ou, Y., Gaonkar, B. & Davatzikos, C. Multi-atlas skull-stripping. *Academic Radiology* **20,** 1566–1576; 10.1016/j.acra.2013.09.010 (2013).

42. Avants, B. B. *et al.* The Insight ToolKit image registration framework. *Front. Neuroinform.* **8,** 44; 10.3389/fninf.2014.00044 (2014).

43. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging* **20,** 45–57; 10.1109/42.906424 (2001).

44. Asman, A. *et al.* (eds.). *Miccai 2013 segmentation algorithms, theory and applications (SATA) challenge results summary* (2013).

45. Srinivasan, D. *et al.* A comparison of Freesurfer and multi-atlas MUSE for brain anatomy segmentation: Findings about size and age bias, and inter-scanner stability in multi-site aging studies. *NEUROIMAGE* **223,** 117248; 10.1016/j.neuroimage.2020.117248 (2020).

46. Wood, S. N. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65,** 95–114; 10.1111/1467-9868.00374 (2003).

47. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8,** 118–127; 10.1093/biostatistics/kxj037 (2007).

48. Bashyam, V. M. *et al.* MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain : a journal of neurology* **143,** 2312–2324; 10.1093/brain/awaa160 (2020).

49. Marcus, D. S. *et al.* Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci* **19,** 1498–1507; 10.1162/jocn.2007.19.9.1498 (2007).

Supplementary Materials A: Data overview and briefs

The ML models from dimensional charts of neuroimaging signatures were trained and evaluated on 71,820 MRI time points from 53,757 participants and 23 studies that were consolidated and harmonized through the iSTAGING pipelines. The centralized processing and analysis of data from multiple data sources enabled the construction of a cross-sectional model of biological covariates spanning the adult lifespan, which was not possible with data coming only from individual studies. In the consolidation process, we also harmonized non-imaging variables, for instance by simplifying study-specific continuous variables into a binary variable. For example, study- and method-specific measures of amyloid were converted into a binary A+/A- categorization. T1 weighted structural imaging data was processed with the MUSE pipeline to obtain tabular data of regional volumes of tissue types and the intracranial volume. T2/FLAIR weighted images were used to extract regional volumes of white matter hyperintensities.

The consolidation addressed several logistical challenges. While the image data processing can run on standard computational hardware, the large amount of data required the use of a high-performance computational cluster with large storage capacity. To keep the data manageable, we stored the data from all studies in the same file structure along with processing logs. To facilitate reproducibility and distribution, we packaged the processing pipelines into a container image. With this container, the preparation of new MRI pre-processing is simplified to a single call per image, lowering the barrier of entry for researchers to a manageable level.

Some of the details of the consolidate data are illustrated in Table 1 whereas more details are provided in Supplementary Material A.

Tab. S 1 ***Dataset characteristics.***

	Clinical Emphasis						Years MRI	Age	Sex	Imaging			Non-Imaging Data										Counts		FU
Study	HA	AD	CVD	Di	R/S	Oth	1995-2020	20-100 years	% female	T1	T2	fMRI	APOE	AB	Tau	Sm	Dia	Ht	Cho	Mem	ExF	PT	MRI	fMR	0-20 years
iSTAGING studies included in harmonization and model training																									
ADNI									45	++	++	++	+	++	++	+	+	++	+	++	++	2444	10854	0	
AIBL									54	++	++	–	+	+	–	+	++	+	++	++	++	1264	1978	0	
BIOCARD									60	++	++	–	+	++	++	++	++	+	–	++	++	300	1127	0	
BLSA									51	++	++	++	+	++	–	++	+	++	++	++	++	1190	3937	2091	
CARDIA									52	++	++	++	+	–	–	++	++	++	++	++	++	894	1379	2057	
OASIS									55	++	–	++	+	++	+	++	+	+	++	+	+	1757	2827	3621	
PENN									56	++	+	+	+	+	+	+	+	+	+	+	+	1106	1295	320	
SHIP									51	+	+	–	+	–	–	++	++	++	++	++	–	3311	3311	0	
UKBIOBANKv1.6									52	++	++	++	+	–	–	++	++	++	+	+	++	21313	22721	21313	
WHIMS									100	++	++	–	+	–	–	+	+	++	+	+	–	1421	2145	0	
WRAP									69	++	++	–	+	+	–	++	++	++	++	++	++	325	735	0	
Added to future harmonization model																									
GSP									58	+	+	–	+	–	–	–	–	–	–	–	–	1570	1570	0	
HCP-YA									54	+	+	++	+	–	–	+	+	+	–	–	+	1206	1113	1200	
Variation already covered by harmonization model																									
HABS									59	++	++	+	+	+	–	++	++	++	++	++	++	290	513	600	
HCP-Aging									56	+	+	++	+	–	–	+	+	+	+	+	+	715	715	725	
PreventAD									71	++	++	–	+	+	+	–	++	++	+	–	+	348	1213	0	
UKBIOBANKv1.7									53	++	++	++	+	–	–	++	++	++	+	+	++	18271	18271	18271	
Extra iSTAGING studies																									
ACCORD									44	+	+	–	–	–	–	+	++	++	++	+	+	605	610	0	
DupontLead									0	+	+	–	+	–	–	++	+	+	+	++	++	638	638	0	
FITBIR									28	++	+	–	–	–	–	+	+	+	+	+	+	1496	3341	0	
HABS-HD									62	++	++	–	+	+	+	++	++	++	++	++	++	3816	5724	0	
HANDLS									55	+	+	–	–	–	–	+	+	++	+	+	+	238	233	211	
LookAHEAD									71	+	+	–	–	–	–	++	++	++	++	++	++	319	311	0	
MESA									53	++	++	++	–	–	–	++	++	++	++	–	–	1043	1041	1062	
PHENOM									36	+	+	–	–	–	–	–	–	–	–	–	–	2752	2752	0	
SPRINT									38	++	++	–	–	–	–	++	++	++	++	++	++	785	1332	1319	

HA: healthy aging; AD: Alzheimer's disease; CVD: cardio-vascular disease; Di: Diabetes; R/S: race/socio-economics; Oth: other; AB: Amyloid beta; Sm: smoking; Ht: hypertension; Cho: cholesterol; Mem: memory; ExF: executive functions; PT: number of participants; FU: follow-up interval; +: single/cross-sectional measurement; ++: longitudinal measurements

ADNI [61, 62] The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a public-private partnership coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. Sixty-three sites throughout the United States of America and Canada have participated in ADNI since the study began in 2004.

With the primary objective of identifying and diagnosing Alzheimer's Disease (AD) at the earliest stage, tracking the progression of the disease using biomarkers, and supporting advancement in prevention and treatment of AD, ADNI has collected serial structural and functional magnetic resonance imaging (MRI) of the brain, positron emission tomography (PET) using F-fluorodeoxyglucose (FDG), Pittsburgh compound-B (PiB), and AV-45 tracers, biological samples (blood, urine, and in a subset CSF), clinical evaluations and neuropsychological assessments of over 1,000 subjects; with AD, age-matched controls, or with Mild Cognitive Impairment (MCI) and a high risk for developing AD.

The entirety of the data set is available after registration and signing of a data use agreement.

AIBL [63] The Australian Imaging, Biomarker, and Lifestyle (AIBL) study began in Perth and Melbourne, Australia between 2006-2008, as a longitudinal prospective research study of Alzheimer's Disease led by the Australian Commonwealth Scientific Industrial and Research Organization (CSIRO). AIBL study

enrolled 1,112 participants above the age of 60, with Alzheimer's Disease (20%), Mild Cognitive Impairment (10%), and healthy controls (70%). The first assessment included cognitive testing, blood collection, and health and lifestyle questionnaires. MRI and PET neuroimaging using the Pittsburgh Compound B (PiB) tracer were initially acquired in only 287 AIBL participants. Following longitudinal assessments, occurring at 18-month intervals, have added neuroimaging to all participants and CSF collection for analysis of biomarkers. The cognitive assessment in AIBL was designed to cover main domains of cognition affected by Alzheimer's disease. This 90-minute assessment was conducted in-clinic by trained staff, most of whom were neuropsychologists. A part of the data has been made publicly available via [https://aibl.csiro.au/adni/index.html].

BIOCARD [16] The Biomarkers of Cognitive Decline Among Normal Individuals (BIOCARD) study was initiated by the Geriatric Branch of the National Institute of Mental Health in 1995. To identify biomarkers associated with the progression of cognitive impairment, BIOCARD recruited 354 cognitively healthy adults including those with a first-degree relative affected with AD. BIOCARD participants were initially admitted for a three-day long evaluation at the Clinical Center at the National Institutes of Health (NIH). At this time, a detailed physical examination, neuropsychological assessment, brain MRI, and blood and cerebrospinal fluid (CSF) samples were obtained. Subsequently, annual evaluations were conducted at the NIH until the study closed in 2005. The Johns Hopkins School of Medicine re-opened the BIOCARD study in 2009, continuing with annual clinical and cognitive follow-up for the original BIOCARD cohort. Amyloid PET imaging using the Pi-B tracer began in 2015, and Tau PET imaging began in 2017. In BIOCARD, Johns Hopkins University continued to administer all the cognitive tests from the original protocol from NIH. The annual neuropsychological assessment covers every major cognitive domain (memory, executive function, language, visuospatial ability, attention, speed of processing, and psychomotor speed). This assessment includes TMT and CVLT cognitive tests of executive function and memory.

BLSA [64] The Baltimore Longitudinal Study of Aging (BLSA) is an ongoing longitudinal study led by the National Institute of Aging (NIA). The BLSA began in Baltimore, Maryland, USA in 1958, with the aim of characterizing the process of aging in general, and healthy aging in particular. The BLSA acquires physiological measures, biomarkers, risk factors, blood samples, and evaluated cognitive function in individuals healthy at enrollment over the course of their lifetime. The neuroimaging sub-study of BLSA began in 1994, with annual or semi-annual MR and PET imaging studies, cognitive testing, and clinical evaluations of a subset of 158 BLSA participants. In 2009, MR imaging assessments were extended to all eligible BLSA participants, who also receive full cognitive assessments, with the current follow-up visit schedule: age less than 60, every 4 years, age 60-79, every 2 years, age 80 and older, every year. 3T MRI data and select variables from the BLSA will be available for download from LONI/GAAIN (https://www.gaaindata.org/partner/BLSA). Investigators can apply for data access via https://www.blsa.nih.gov/.

CARDIA [19] The Coronary Artery Risk Development in Young Adults (CARDIA) is an ongoing longitudinal study of the development and trajectories of cardiovascular disease in a bi-racial population-based cohort. Data collection began in 1985 with follow-up visits in 5-year intervals, at four sites in the USA. Brain MRI began in CARDIA year 25 and continued in year 30 and year 35, with a subset of less than 1,000 of over 5,000 subjects participating in the brain MRI protocol. Non-imaging data from the CARDIA study is available through the NHLBI's BIOLINCC [https://biolincc.nhlbi.nih.gov/studies/cardia/].

GSP [17] The Brain Genomics Superstruct Project (GSP) aims to examine the link between brain function, behavior, and genetic variation. Data collection included a saliva-based genomic data collection, structural and function MR imaging, and ongoing online assessments including demographics, health questionnaires, cognitive and behavioral assessment in 1,570 young and healthy participants.

HABS [9] The Harvard Brain Aging Study (HABS) is a longitudinal observational study of healthy and pathological aging from cognitively unimpaired participants age 65 and older. MR and PET imaging protocols were designed to be compatible with ADNI. Clinical assessments include demographics, neurocognitive and neuropsychological assessments, blood work, APOE genotyping, and physiological measurements. Data is publicly available via [https://habs.mgh.harvard.edu/researchers/].

HCP-YA [65] The Human Connectome Project Young Adult study (HCP-YA) aimed to create a standard map of normal connections between brain regions. Toward this goal, they collected over 1,100 high resolution MR imaging studies in healthy young adult twins and siblings. The MR imaging data is publicly available via [https://www.humanconnectome.org/study/hcp-young-adult].

HCP-Aging [8, 10] the HCP was later extended to cover the entire lifespan, through the HCP-Aging study, which covers ages 36-100 with high resolution MR imaging studies in over 1,200 adult participants.

OASIS [6] is the latest release in the Open Access Series of Imaging Studies (OASIS) that aimed at making neuroimaging datasets freely available to the scientific community. Previously released data for OASIS-Cross-sectional[49] and OASIS-Longitudinal[7] have been utilized for hypothesis driven data analyses, development of neuroanatomical atlases, and development of segmentation algorithms. OASIS-3 is a longitudinal neuroimaging, clinical, cognitive, and biomarker dataset for normal aging and Alzheimer's Disease.

The OASIS datasets hosted by central.xnat.org provide the community with open access to a significant database of neuroimaging and processed imaging data across a broad demographic, cognitive, and genetic spectrum an easily accessible platform for use in neuroimaging, clinical, and cognitive research on normal aging and cognitive decline. All data is available via [www.oasis-brains.org].

PENN [66] The PENN dataset was collected at the University of Pennsylvania's Memory Center, including participants with memory complaints and early stages of cognitive decline. Data collection is compatible with the standard National Alzheimer's Coordinating Center's uniform data set.

PREVENT-AD [47] The Pre-symptomatic Evaluation of Experimental or Novel Treatments for AD (PREVENT-AD) is an observational study of cognitively unimpaired participants with a sibling or family member who has or had Alzheimer's disease. N= 349 study participants were enrolled at McGill University from the year 2010, some of whom participated in a clinical trial for an anti-inflammatory drug, naproxen sodium. Data collection includes longitudinal MR imaging, CSF biomarker and genetic assays, and cognitive and health screenings. Data is publicly available via [https://openpreventad.loris.ca/].

SHIP [67, 68] The Study of Health in Pomerania (SHIP) is a population-based cohort including adults from 22-90 years old with MR imaging collected from 2008-2012. Additional data collection includes demographics, health history, cognitive testing, and genotyping.

UK BIOBANK [69] UK Biobank is a major national and international health resource, and a registered charity in its own right, with the aim of improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses – including cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia. UK Biobank recruited 500,000 people aged between 40-69 years in 2006-2010 from across the country to take part in this project. They have undergone measures, provided blood, urine and saliva samples for future analysis, detailed information about themselves and agreed to have their health followed. Over many years this will build into a powerful resource to help scientists discover why some people develop a disease and others do not. The UK Biobank dataset is publicly available for researchers via [https://www.UK Biobank.ac.uk/].

WHIMS [70] Brain MRI scans were obtained in a subset of 1,403 women aged 71–89 years who participated in the Women's Health Initiative Memory Study (WHIMS). WHIMS is an ancillary study to the Women's Health Initiative, which consisted of two randomized, placebo-controlled trials: 0.625 mg conjugated equine estrogens (CEE) with or without 2.5 mg medroxyprogesterone acetate (MPA) in one daily tablet. Scans were performed, on average, 3.0 years post-trial for the CEE + MPA trial and 1.4 years post-trial for the CEE-Alone trial; average on-trial follow-up intervals were 4.0 years for CEE + MPA and 5.6 years for CEE-Alone.

WRAP [5] The Wisconsin Registry for Alzheimer's Prevention (WRAP) is a longitudinal observational study of cognitively unimpaired adults. Participants are primarily the healthy children of late-onset Alzheimer's Disease patients at the Memory Assessment Clinic at the University of Wisconsin-Madison. Since 2001, WRAP has followed 1,561 subjects at bi-annual assessments, including molecular PET and structural MR imaging, CSF collection, laboratory tests, genetic screening, cognitive testing, and self-reported histories of lifestyle and health.

ACCORD [71] The Action to Control Cardiovascular Risk in Diabetes (ACCORD) clinical trial evaluated intensive glycemic control on a cohort of participants with type 2 diabetes. The Memory in Diabetes (MIND) sub-study further evaluated cognitive and structural brain outcomes in a subset of participants from the two randomized groups of ACCORD study. Non-imaging data from the ACCORD clinical trial is available through the NHLBI's BIOLINCC [https://biolincc.nhlbi.nih.gov/studies/accord/].

OrganoLead [72] The aim of this study was to evaluate the relationship between lifetime cumulative lead exposure with cognitive function and brain structure in a population of former employees at a chemical manufacturing plant in the USA. Brain MR imaging and tibia lead was collected in over 500 men between the age of 40-70 years. Additional data collection included demographics, health history, and cognitive testing.

FITBIR [73] The Federal Interagency TBI Research (FITBIR) informatics system is hosted in collaboration between the NIH and the UK Department of Defense, as a public dataset for research into traumatic brain injury (TBI). The consortium includes multiple studies with varying data collection procedures, all focused on TBI. Data is publicly available via [https://fitbir.nih.gov/].

HANDLS [74] The Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study is an ongoing longitudinal study of race and socioeconomic-related health disparities based in Baltimore. Brain MRI was introduced in the third wave of the HANDLS protocol, between the years 2009-2013.

LookAHEAD [21] The Action for Health in Diabetes (LookAHEAD) trial studies the long-term effect of intensive lifestyle intervention in participants with type 2 diabetes, with outcomes of cardiovascular morbidity and mortality. The randomized control trial began in 2001 with brain MRI added in a subset of the initial trial population on average 10 years after randomization. Data from the randomized controlled clinical trial is available (including measures of brain structure and cognitive function) through the NIH's NIDDK central repository [https://repository.niddk.nih.gov/studies/look-ahead/].

MESA [23] The Multi-Ethnic Study of Atherosclerosis (MESA) is a longitudinal study of subclinical cardiovascular disease that began in the year 2000. This population has diverse racial and ethnic characteristics, including 25% Black, 15% Chinese-American, 19% Hispanic, and 41% White participants. The MESA cohort was sub-sampled to collect brain MR imaging at six sites at the time of the sixth exam, between the years 2018-2019.

SPRINT [75] The Systolic Blood Pressure Intervention Trial (SPRINT) is a randomized control trial designed to evaluate how intensive blood pressure treatment affects the cardiovascular system, kidneys, and brain. The clinical trial ran from 2010-2015, including brain MR imaging in a subset of N= 670 participants. MR data was collected at seven sites according to a standardized protocol. Non-imaging data from the SPRINT clinical trial is available through the NHLBI's BIOLINCC [https://biolincc.nhlbi.nih.gov/studies/sprint/].

PHENOM Consortium The PHENOM (Psychosis Heterogeneity Evaluated Via Dimensional Neuroimaging) consortium is an international consortium pooling MRI and clinical data across several institutions world-wide [FIXME add citation https://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.21070686]. The main focus on the consortium is to study the use of machine learning and MRI as a means for deriving personalized imaging signatures of schizophrenia, and of predicting future clinical progression. An associated goal is to use state of the art weakly supervised machine learning methods to dissect the neuroanatomical heterogeneity in psychosis. In addition to offering healthy control subjects at younger ages, this consortium helps establish imaging signatures of psychosis into NiChart.

Supplementary materials B: Results

6.1 Out-of-sample harmonization of UK Biobank batches

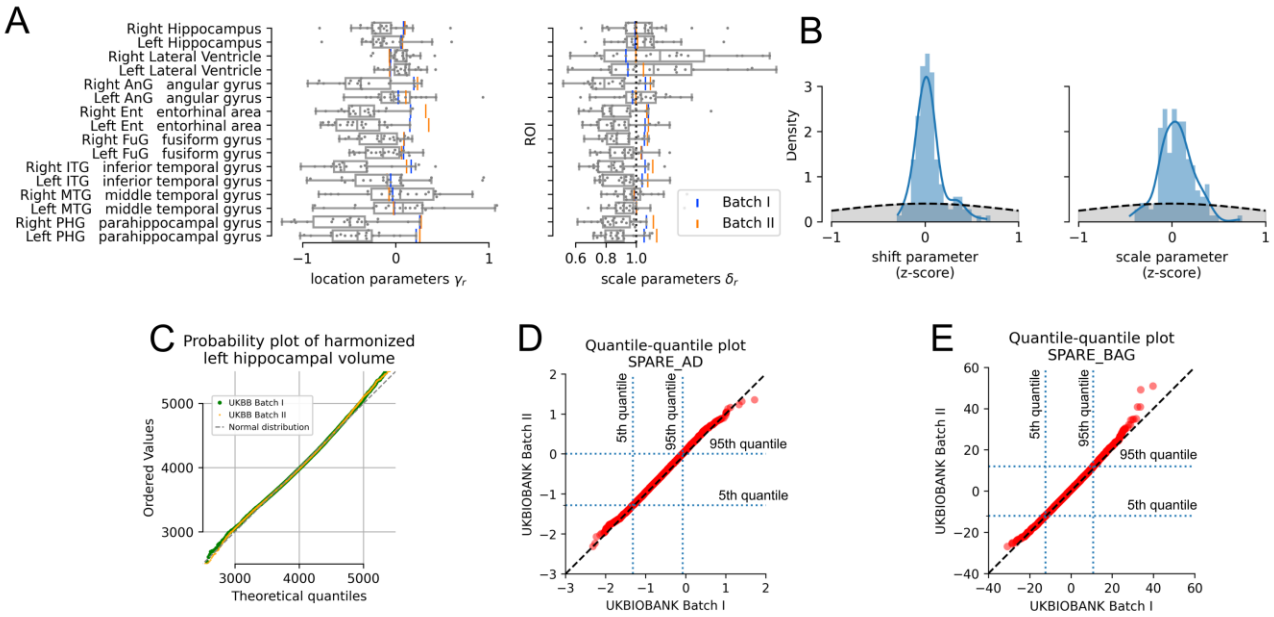

Fig. S 1 Figure 1 (A) Subset of regions of interest (ROIs) showing boxplots estimated location and scale parameters of the reference NiChart data set. UK Biobank Batch I and Batch II are highlighted. (B) Density plot of differences of estimated location and scale parameter of UK Biobank Batch and Batch II in light blue. The gray area limited by the dashed black line represents the probability density of the empirical Bayes prior. (C) Theoretical quantiles of a Normal distribution of the volume of the right hippocampus and the empirical quantiles of the two UK Biobank batches. (D) Quantile-quantile plots of the SPARE-AD. (E) Quantile-quantile plots of SPARE-BA.

Supplementary Materials C: Software implementation

The graphical user interface associated with the application programming interface makes the tasks required for NiChart (multiple preprocessing steps, harmonization, analysis using pre-trained ML models) accessible without the need for programming. The code is written in Python and is compatible with major computer operating systems including Windows, macOS, and Linux. NiChart is structured around an extensible set of plugins, all of which have access to the same data that is provided by the user. Plugins are called through a well-defined Application Programming Interface (API), and each plugin can run calculations and provide visualizations. Critically, all statistical and machine learning models referred to in this manuscript are published alongside with the software. By design, input data tables have only a few mandatory data fields for minimal operations. The fields include participant identity, date of acquisition—which can be a dummy date to distinguish visits, age at acquisition, sex, and at least one set of imaging features obtained from standardized pre-processing to calculate at least one brain aging chart. The mandatory columns can be complemented with additional variables such as cognitive scores or physiological markers for additional plugins and downstream analyses. Typical actions performed by plugins include reading from the main data table, calculation of new variables, interactive display of original or derived data, and addition of derived data to the data table. The input/output plugin provides functionality to import and export data from and to several tabular formats to enable interoperability with other up- and downstream data analysis tools. Much of the functionality of loading data, harmonization, and extraction of imaging signatures is intuitively executable as it is actionable through labeled buttons like “Load harmonization model” or “Add harmonized data to data frame”. The initial software release includes plugins and associated statistical and machine learning models for extracting multiple neuro-imaging signatures and statistical normative ranges of control populations. The plugins are developed and operate independently of each other, albeit assuming a certain structure of the data table. In the standalone package, all computations are executed locally and thus no personal health information is exposed to any third party. The modular application programming interface exposes the core and plugins’ functionality to developers and advanced data scientists.

Supplementary References

1. Ferrucci, L., *The Baltimore Longitudinal Study of Aging (BLSA): a 50-year-long journey and plans for the future.* J Gerontol A Biol Sci Med Sci, 2008. **63**(12): p. 1416-9.
2. Fowler, C., et al., *Fifteen years of the Australian Imaging, Biomarkers and Lifestyle (AIBL) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to Alzheimer’s disease.* Journal of Alzheimer's disease reports, 2021. **5**(1): p. 443-468.
3. Jack, J.C.R., et al., *The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods.* Journal of Magnetic Resonance Imaging, 2008. **27**(4): p. 685-691.
4. Miller, K.L., et al., *Multimodal population brain imaging in the UK Biobank prospective epidemiological study.* Nature neuroscience, 2016. **19**(11): p. 1523-1536.
5. Johnson, S.C., et al., *The Wisconsin Registry for Alzheimer's Prevention: a review of findings and current directions.* Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2018. **10**: p. 130-142.
6. Marcus, D.S., et al., *Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults.* Journal of cognitive neuroscience, 2007. **19**(9): p. 1498-1507.

7. Resnick, S.M., et al., *Postmenopausal hormone therapy and regional brain volumes: the WHIMS-MRI Study.* Neurology, 2009. **72**(2): p. 135-42.
8. Bookheimer, S.Y., et al., *The lifespan human connectome project in aging: an overview.* Neuroimage, 2019. **185**: p. 335-348.
9. Dagley, A., et al., *Harvard aging brain study: dataset and accessibility.* Neuroimage, 2017. **144**: p. 255-258.
10. Harms, M.P., et al., *Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects.* Neuroimage, 2018. **183**: p. 972-984.
11. Villemagne, V.L., et al., *Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study.* Lancet Neurology, 2013. **12**(4): p. 357-367.
12. Weintraub, D., et al., *Neurodegeneration across stages of cognitive decline in Parkinson disease.* Arch Neurol, 2011. **68**(12): p. 1562-8.
13. Wright, I.C., et al., *Meta-Analysis of Regional Brain Volumes in Schizophrenia.* American Journal of Psychiatry, 2000. **157**(1): p. 16-25.
14. Zanetti, M.V., et al., *Neuroanatomical pattern classification in a population-based sample of first-episode schizophrenia.* Progress in Neuro-Psychopharmacology and Biological Psychiatry, 2013. **43**: p. 116-125.
15. Smith, S.M., et al., *Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations.* elife, 2020. **9**: p. e52677.
16. Albert, M., et al., *Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to ApoE genotype.* Current Alzheimer Research, 2014. **11**(8): p. 773-784.
17. Holmes, A.J., et al., *Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures.* Scientific data, 2015. **2**(1): p. 1-16.
18. Wadden, T.A., et al., *The Look AHEAD Study: A description of the lifestyle intervention and the evidence supporting it.* Obesity, 2006. **14**(5): p. 737-752.
19. Cermakova, P., et al., *Subclinical cardiac dysfunction and brain health in midlife: CARDIA (Coronary Artery Risk Development in Young Adults) brain magnetic resonance imaging substudy.* Journal of the American Heart Association, 2017. **6**(12): p. e006750.
20. Nasrallah, I.M., et al., *Association of Intensive vs Standard Blood Pressure Control With Cerebral White Matter Lesions.* Jama-Journal of the American Medical Association, 2019. **322**(6): p. 524-534.
21. Espeland, M.A., et al., *Sex-related differences in brain volumes and cerebral blood flow among overweight and obese adults with type 2 diabetes: exploratory Analyses from the Action for Health in Diabetes Brain Magnetic Resonance Imaging Study.* The Journals of Gerontology: Series A, 2020. **75**(4): p. 771-778.
22. Waldstein, S.R., et al., *Differential Associations of Socioeconomic Status With Global Brain Volumes and White Matter Lesions in African American and White Adults: the HANDLS SCAN Study.* Psychosomatic Medicine, 2017. **79**(3): p. 327-335.
23. Austin, T.R., et al., *Association of Brain Volumes and White Matter Injury With Race, Ethnicity, and Cardiovascular Risk Factors: The Multi-Ethnic Study of Atherosclerosis.* Journal of the American Heart Association, 2022. **11**(7): p. e023159.
24. Arbabshirani, M.R., et al., *Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls.* Neuroimage, 2017. **145**: p. 137-165.
25. Ferreira, D., et al., *Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications.* Scientific reports, 2017. **7**(1): p. 46263.
26. Habes, M., et al., *The Brain Chart of Aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans.* Alzheimers Dement, 2021. **17**(1): p. 89-102.

27. Yang, Z., et al., *A deep learning framework identifies dimensional representations of Alzheimer's Disease from brain structure.* Nat Commun, 2021. **12**(1): p. 7065.
28. Adebimpe, A., et al., *ASLPrep: a platform for processing of arterial spin labeled MRI and quantification of regional brain perfusion.* Nature methods, 2022. **19**(6): p. 683-686.
29. Esteban, O., et al., *fMRIPrep: a robust preprocessing pipeline for functional MRI.* Nature methods, 2019. **16**(1): p. 111-116.
30. Fischl, B., *FreeSurfer.* Neuroimage, 2012. **62**(2): p. 774-81.
31. Pomponio, R., et al., *Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan.* Neuroimage, 2020. **208**: p. 116450.
32. Bethlehem, R.A., et al., *Brain charts for the human lifespan.* Nature, 2022. **604**(7906): p. 525-533.
33. Rutherford, S., et al., *Charting brain growth and aging at high spatial precision.* elife, 2022. **11**: p. e72904.
34. Davatzikos, C., *Machine learning in neuroimaging: Progress and challenges.* Neuroimage, 2019. **197**: p. 652-656.
35. Mateos-Pérez, J.M., et al., *Structural neuroimaging as clinical predictor: A review of machine learning applications.* NeuroImage: Clinical, 2018. **20**: p. 506-522.
36. Varoquaux, G. and V. Cheplygina, *How I failed machine learning in medical imaging--shortcomings and recommendations.* arXiv preprint arXiv:2103.10292, 2021.
37. Bashyam, V.M., et al., *MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14468 individuals worldwide.* Brain, 2020. **143**(7): p. 2312-2324.
38. Davatzikos, C., et al., *Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index.* Brain, 2009. **132**(Pt 8): p. 2026-35.
39. Hwang, G., et al., *Disentangling Alzheimer's disease neurodegeneration from typical brain ageing using machine learning.* Brain Communications, 2022. **4**(3): p. fcac117.
40. Wen, J., et al., *Multi-scale semi-supervised clustering of brain images: Deriving disease subtypes.* Med Image Anal, 2022. **75**: p. 102304.
41. Yang, Z., J. Wen, and C. Davatzikos, *Smile-GANs: Semi-supervised clustering via GANs for dissecting brain disease heterogeneity from medical images.* arXiv:2006.15255 [cs, eess, q-bio, stat], 2020.
42. Davatzikos, C., et al., *Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy.* Neuroimage, 2001. **14**(6): p. 1361-9.
43. Doshi, J., et al., *DeepMRSeg: A convolutional deep neural network for anatomy and abnormality segmentation on MR images*. 2019, arXiv:1907.02110. p. arXiv:1907.02110 %8 July 01, 2019 %! DeepMRSeg: A convolutional deep neural network for anatomy and abnormality segmentation on MR images.
44. Doshi, J., et al., *MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection.* NeuroImage, 2016. **127**: p. 186-195.
45. Sotiras, A., S.M. Resnick, and C. Davatzikos, *Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization.* Neuroimage, 2015. **108**: p. 1-16.
46. Franke, K., et al., *Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters.* Neuroimage, 2010. **50**(3): p. 883-92.
47. Tremblay-Mercier, J., et al., *Open science datasets from PREVENT-AD, a longitudinal cohort of pre-symptomatic Alzheimer's disease.* NeuroImage: Clinical, 2021. **31**: p. 102733.
48. Rousseeuw, P.J., *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.* Journal of computational and applied mathematics, 1987. **20**: p. 53-65.
49. Davatzikos, C., et al., *Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification.* Neurobiol Aging, 2011. **32**(12): p. 2322 e19-27.

50. Davatzikos C, e.a. *Imaging Signatures of AD, Brain Aging and White Matter Hyperintensities Show Dissociable Associations with Cognitive Decline in Healthy Subjects Longitudinally Followed in Adni*. in *AAIC Annual Conference*. 2018. Chicago, IL: Poster.
51. Dwyer, D.B., P. Falkai, and N. Koutsouleris, *Machine learning approaches for clinical psychology and psychiatry.* Annual review of clinical psychology, 2018. **14**: p. 91-118.
52. Rathore, S., et al., *A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages.* NeuroImage, 2017. **155**: p. 530-548.
53. Williamson, J.D., et al., *Cognitive Function and Brain Structure in Persons With Type 2 Diabetes Mellitus After Intensive Lowering of Blood Pressure and Lipid Levels: A Randomized Clinical Trial.* JAMA Intern Med, 2014.
54. Stewart, W.F., et al., *Past Adult Lead Exposure is Linked to Neurodegeneration Measured by Brain MRI.* Neurology, 2006. **66**: p. 1476-1484.
55. Doshi, J., et al., *Multi-Atlas Skull-Stripping.* Academic radiology, 2013. **20**(12): p. 1566-1576.
56. Avants, B.B., et al., *Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain.* Medical Image Analysis, 2008. **12**(1): p. 26-41.
57. Smith, S.M., et al., *Advances in functional and structural MR image analysis and implementation as FSL.* Neuroimage, 2004. **23**(S1): p. 208-219.
58. Asman, A. *MICCAI 2013 Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA)*. 2013; MICCAI SATA 2013 competition]. Available from: http://masi.vuse.vanderbilt.edu/submission/leaderboard.html.
59. Srinivasan, D., et al., *A comparison of Freesurfer and multi-atlas MUSE for brain anatomy segmentation: Findings about size and age bias, and inter-scanner stability in multi-site aging studies.* Neuroimage, 2020. **223**: p. 117248.
60. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
61. Weiner, M.W., et al., *The Alzheimer's disease neuroimaging initiative: progress report and future plans.* Alzheimer's & Dementia, 2010. **6**(3): p. 202-211. e7.
62. Jack Jr, C.R., et al., *The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods.* Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2008. **27**(4): p. 685-691.
63. Pietrzak, R.H., et al., *Trajectories of memory decline in preclinical Alzheimer's disease: results from the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing.* Neurobiology of aging, 2015. **36**(3): p. 1231-1238.
64. Ferrucci, L., *The Baltimore Longitudinal Study of Aging (BLSA): a 50-year-long journey and plans for the future*. 2008, Oxford University Press. p. 1416-1419.
65. Glasser, M.F., et al., *The human connectome project's neuroimaging approach.* Nature neuroscience, 2016. **19**(9): p. 1175-1187.
66. Wolk, D.A., et al., *Amyloid imaging in Alzheimer's disease: comparison of florbetapir and Pittsburgh compound-B positron emission tomography.* Journal of Neurology, Neurosurgery & Psychiatry, 2012. **83**(9): p. 923-926.
67. Habes, M., et al., *Relationship between APOE genotype and structural MRI measures throughout adulthood in the study of health in Pomerania population-based cohort.* American Journal of Neuroradiology, 2016. **37**(9): p. 1636-1642.
68. Hegenscheid, K., et al. *Whole-body magnetic resonance imaging of healthy volunteers: pilot study results from the population-based SHIP study*. in *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. 2009. © Georg Thieme Verlag KG Stuttgart· New York.

69. Alfaro-Almagro, F., et al., *Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank.* Neuroimage, 2018. **166**: p. 400-424.
70. Resnick, S.M., et al., *Postmenopausal hormone therapy and regional brain volumes: the WHIMS-MRI Study.* Neurology, 2009. **72**(2): p. 135-142.
71. Williamson, J.D., et al., *The action to control cardiovascular risk in diabetes memory in diabetes study (ACCORD-MIND): rationale, design, and methods.* The American journal of cardiology, 2007. **99**(12): p. S112-S122.
72. Schwartz, B.S., et al., *Evaluation of cumulative lead dose and longitudinal changes in structural MRI in former organolead workers.* Journal of occupational and environmental medicine/American College of Occupational and Environmental Medicine, 2010. **52**(4): p. 407.
73. Wintermark, M., et al., *Traumatic brain injury imaging research roadmap.* American Journal of Neuroradiology, 2015. **36**(3): p. E12-E23.
74. Waldstein, S.R., et al., *Differential associations of socioeconomic status with global brain volumes and white matter lesions in African American and white adults: the HANDLS SCAN study.* Psychosomatic medicine, 2017. **79**(3): p. 327.
75. Nasrallah, I.M., et al., *Association of intensive vs standard blood pressure control with cerebral white matter lesions.* Jama, 2019. **322**(6): p. 524-534.