A Low-dimensional Manifold Representation of the Human Brain Aging Continuum

Zifei Liang[*,1], Collin J. Szczepanski[*,1], Chenyang Li[1], Jiayi Li[1,2], Henry Rusinek[1], Yulin Ge[1], Jiangyang Zhang[1,3]

[1] Bernard and Irene Schwartz Center for Biomedical Imaging, Department of Radiology, New York University Grossman School of Medicine, New York, NY 10016, USA

[2] Department of Computer Science, New York University, New York, NY 10016, USA

[3] Neuroscience Institute, New York University Grossman School of Medicine, New York, NY 10016, USA

[*] Zifei Liang and Collin J. Szczepanski contributed equally to this work

Corresponding to:

Jiangyang Zhang, Ph.D.

Center for Biomedical Imaging,

Department of Radiology,

New York University Grossman School of Medicine,

327 E 30th Street, Room 733

New York, NY 10016, USA

Jiangyang.zhang@nyulangone.org

Abstract

The structural changes accompanying brain aging exhibit complex, multifaceted patterns that challenge traditional analytical approaches and impede accurate assessment of individual brain health. While previous studies have documented these age-related changes, integrating findings across multiple brain structures to assess individual brain age status has proven difficult due to the high-dimensional nature of neuroimaging data. Using structural MRI data from the Human Connectome Project Aging Dataset, we demonstrate that age-related regional volume changes can be mapped onto a low-dimensional manifold that reflects underlying neuroanatomical constraints. This manifold representation provides a transparent framework for both assessing an individual's overall brain aging status and identifying specific regional changes that drive this assessment. By examining data within neighborhoods on the manifold, we identified distinct structural changes within anatomically homogeneous subgroups, including pronounced frontal atrophy that was observed primarily in male participants. These local patterns can be aggregated to create a comprehensive picture of how an individual brain ages, providing new insights into the heterogeneous nature of brain aging.

Main

Brain aging is a heterogeneous process, marked by varying rates of changes across age, brain regions, and individuals. This variability extends to neurodegenerative diseases, such as Alzheimer's disease (AD) and AD-related dementia, which disrupts the course of normal aging in region-specific manners[1–3]. Magnetic resonance imaging (MRI) based studies have consistently identified robust age-related alterations in multiple brain regions (e.g., ventricular enlargement[4] and hippocampal atrophy[5]). However, the spatiotemporal heterogeneity of aging processes creates complex regional patterns that vary substantially between individuals[6,7], making it challenging to provide personalized assessments of cerebral and cognitive health or to make diagnosis or prognosis along the normal-to-pathological aging continuum.

To address this challenge, researchers have utilized large-scale MRI datasets and machine learning algorithms to estimate the biological age of the brain (i.e., brain age[8]), which shows promise as a personalized marker for assessing overall brain aging status. Brain age gap (the difference between predicted brain age and chronological age) correlates more strongly with cognitive performance and health outcomes than chronological age alone, with typical prediction accuracies of mean absolute error (MAE) around 3-7 years in healthy adults[9–11]. However, these metrics often lack biological interpretability and provide little insight into underlying regional changes that contribute to the overall assessment. Normative trajectory models have been developed to track multiple brain regions across the lifespan, offering standard regional aging curves along with individual centile scores[12–16]. Despite these advancements, capturing complex multi-regional trajectories remains challenging due to numerous brain regions, inter-regional correlations[17], and individual variation. In response, data-driven approaches have identified distinct clusters of brain regions with interconnected aging dynamics, forming potential bases to

examine the aging process[18,19]. Building on these developments, the opportunity has arisen for a multiscale metric that combines the comprehensible overall brain age with specific regional aging patterns. Such a metric could provide both clinical utility and biological interpretability for personalized brain aging assessments.

While MRI data inherently exist in a high-dimensional space ,aging-related changes in the brain likely occur within a lower-dimensional manifold embedded within this space[20]. This concept, known as the manifold hypothesis[21], suggests that both inter-individual variations and age-related changes follow underlying anatomical constraints rather than occurring across all possible dimensions. Previous studies have demonstrated the existence of such manifolds and used them to classify normal aging versus AD subjects[22,23]. However, developing interpretable brain age metrics using these more compact manifold representations remains unexplored.

In this study, we developed a manifold representation of the brain aging continuum and an associated computational framework based on structural MRI data from the Human Connectome Project Aging (HCP-A) dataset[24]. Specifically, we characterized age-related anatomical changes by analyzing data within neighborhoods on the manifold, where subjects exhibited similar regional volumes. We then integrated findings from these anatomically homogeneous subgroups across the entire manifold to reveal large-scale patterns of brain aging. This framework enabled us to provide overall brain age estimates and identify the specific anatomical changes that most strongly influence these estimates. Unlike conventional deep learning–based brain age models, which often operate as opaque systems[25], our approach offers complete transparency and biological interpretability at the regional level.

Results

A two-dimensional (2D) representation of the aging brain continuum

T_1-weighted MRI data from 690 HCP-A subjects were parcellated into 267 regions using an automated multi-atlas pipeline (**Fig. 1a**), producing a 267x1 regional volume vector $\boldsymbol{v}$ for each subject (**Fig. 1b**). We utilized uniform manifold approximation and projection (UMAP)[26] to construct a mapping φ that projects all subjects' $\boldsymbol{v}$s onto a 2D manifold, called aging manifold here (**Fig. 1c**). Manifold stability was assessed by systematically varying UMAP parameters (**Extended Data Fig. 1**). The distribution of HCP-A subjects on the manifold showed gradients corresponding to age (**Fig. 1d**) and sex (**Fig. 1e**), consistent with well-established sex and age-related differences in brain morphology. Similar gradients were also found in results from principal component analysis (PCA) (**Extended Data Fig. 2a-b**), although the subject distribution was less uniform, likely because that PCA was driven primarily by the large differences between young and old subjects[27]. Normalizing $\boldsymbol{v}$ by individual intracranial volume removed the sex-related gradient while preserving the age-related gradient (**Extended Data Fig. 3a-b**). Several outliers that deviated from the age-related gradient were observed (**Fig. 1d**); however, the overall brain morphology of these outliers aligned with their manifold neighbors better than age-matched subjects, suggesting mismatches between chronological ages and expected regional volumes. By stratifying male and female HCP-A subjects into multiple age groups, we computed average aging trajectories for each sex based on mean regional volumes and their 95% confidence intervals on the manifold (**Fig. 1f**). The substantial scatter of individual subjects around these average trajectories highlighted the considerable variability within the HCP-A dataset.

The immediate question was whether the aging manifold, as a representation of plausible regional volume configurations alone the aging continuum, could be generalized to other datasets. While the manifold was constructed using the HCP-A dataset, structural MRI data from other studies can also be mapped onto it using the UMAP-derived φ. To test this generalizability, we mapped selected data from the HCP-Young adult (HCP-YA)[28] and Alzheimer's disease neuroimaging initiative (ADNI)[29] datasets onto the manifold (**Fig. 1g**). HCP-YA subjects (ages 25-40 years) colocalized with younger HCP-A subjects, while ADNI subjects (ages 55-90 years) positioned alongside older subjects at the opposite end of the manifold. A similar finding was also observed in the PCA-based results (**Extended Data Fig. 2c**). Furthermore, the manifold framework demonstrated flexibility, as it can be readily extended by incorporating structural MRI data from additional datasets (e.g., HCP-YA data as shown in **Extended Data Fig. 3c**) or non-MRI data modalities (e.g., cognitive test results as demonstrated in **Extended Data Fig. 3d-e**).

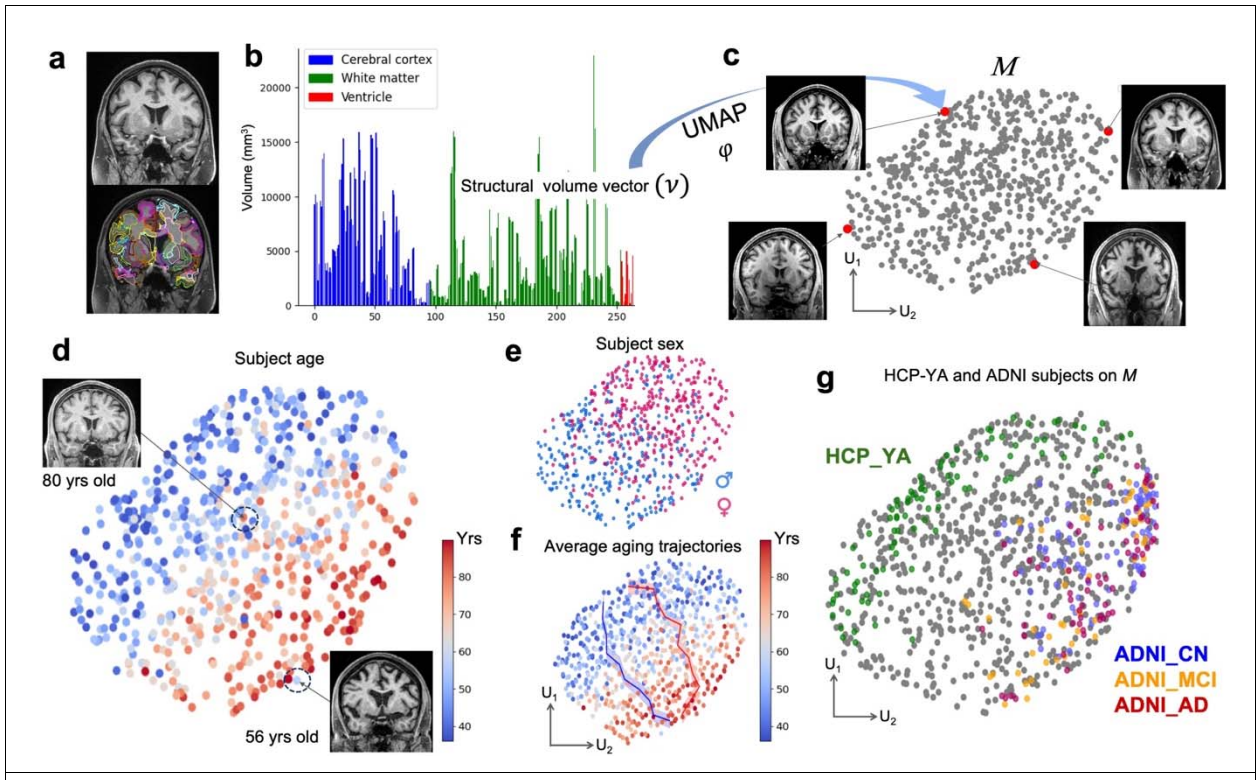

Figure 1: Mapping of high-dimensional regional volume data from the HCP-A dataset onto a 2D manifold using UMAP. **a-b:** T_1-weighted MRI images from each HCP-A subject were segmented into 267 regions, whose volumes formed a volume vector $\boldsymbol{v}$. **c:** The volume vectors of 690 HCP-A subjects were projected using UMAP onto a 2D manifold $\boldsymbol{M}$, visualized as a point cloud where each point represents one subject. **d**: Distribution of chronological age across individual subjects on $\boldsymbol{M}$. The insets show two selected outliers whose chronological ages differed markedly from their neighbors' but who shared similar brain morphology (e.g., ventricle size) as shown in panel **c**. **f:** Average aging trajectories of male and female subjects on $\boldsymbol{M}$ with 95% confidence intervals. **g:** Distribution of external validation datasets on $\boldsymbol{M}$: HCP-YA subjects (green) and ADNI subjects, color-coded by diagnostic status (red for AD, orange for mild cognitive impairment (MCI), and blue for cognitively normal (CN)).

Estimation of brain ages on the manifold

The observed age-related gradient motivated us to treat age as a smooth function on the aging manifold and estimate a subject's brain age. For a subject located in a neighborhood on the manifold that contained other subjects with known chronological ages, we applied Watson kernel regression in the corresponding local coordinates—a Euclidean space homeomorphic to the neighborhood—to estimate that subject's brain age (the insert in **Fig. 2a**). We optimized the

neighborhood size and found that approximately 50 subjects per neighborhood yielded the best performance. Using leave-one-out validation, we observed a strong correlation between the estimated and chronological age (R^2=0.74, p<0.0001) (**Fig. 2a**), with a MAE of 5.8 years. This performance was comparable to previously published deep learning networks trained on similar amounts of data[11], though higher than MAEs reported by networks trained on 10,000 or more subjects[30,31]. Brain ages estimated using a 3D manifold yielded a comparable performance (R^2=0.73, p<0.0001, MAE = 6.1 years). The distribution of estimation error was relatively uniform across the manifold (**Fig. 2b**) and depended on the subject density and age variations within each neighborhood (**Fig. 2c-d**).

Reducing the density of subjects on the manifold increased estimation errors. For example, when using a randomly selected subset of 600 subjects, the average MAE of brain age estimation increased to 6.1 years and continued to increase as subject density was further reduced (**Fig. 2e**). Furthermore, brain age estimation for 200 randomly selected HCP-YA subjects yielded an MAE of 10.2 years, likely because HCP-YA subject ages fell outside the age range of the HCP-A training dataset. However, using an extended manifold constructed from both HCP-A and HCP-YA datasets (**Extended Data Fig. 3c**), the MAE for brain age estimation of an independent set of HCP-YA subjects was reduced to 5.5 years.

The method can be extended to incorporate additional MRI markers. We constructed a 2D manifold using fractional anisotropy (FA) and mean diffusivity (MD) data—two widely used diffusion MRI markers—from white matter regions (**Fig. 2f**). Brain age estimation based on white matter FA and MD from 600 subjects yielded an MAE of 8.3 years and R^2 of 0.61 (**Fig. 2g**). Combining regional volumes and white matter FA and MD values (**Fig. 2h-i**) improved MAE to 6.2 years (R^2 of 0.71 with 600 subjects on the manifold).

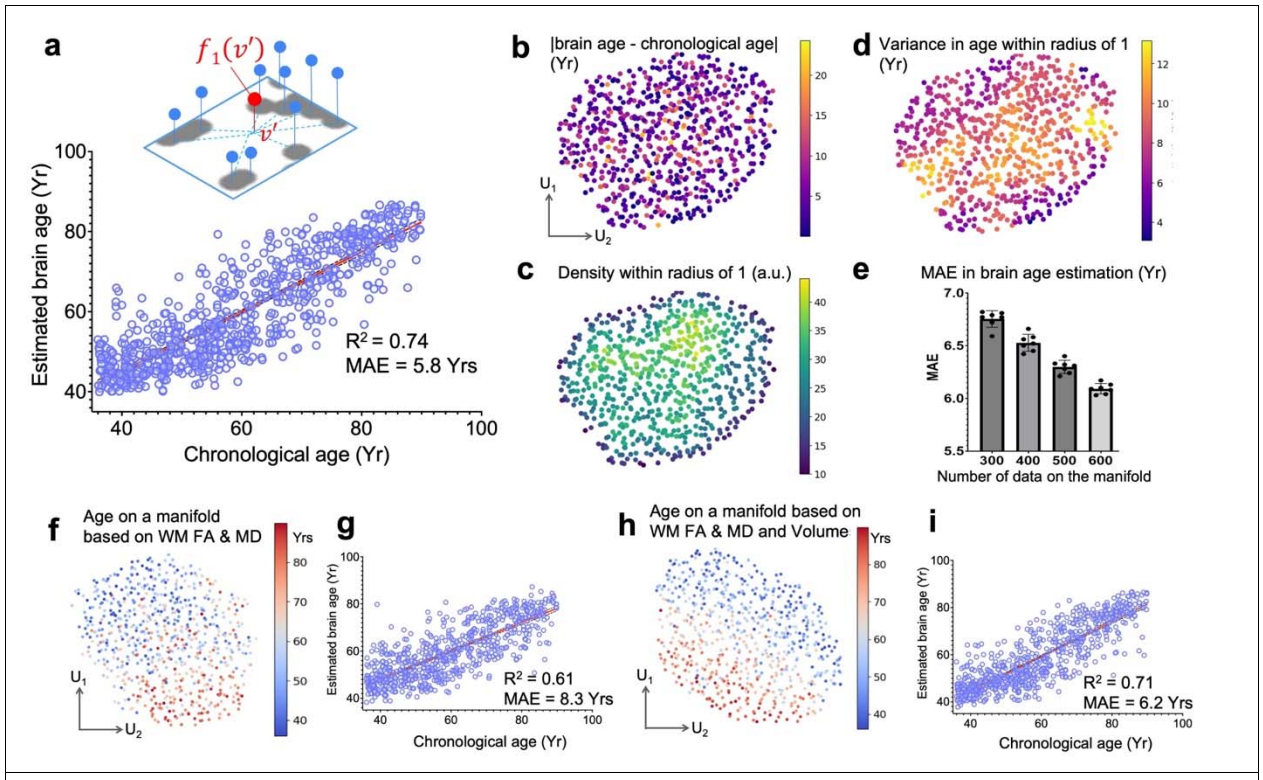

Figure 2: Brain age estimation on the aging manifold. **a:** Correlation between estimated brain age and chronological age for HCP-A subjects. The inset shows a schematic of regression for a subject with a regional volume vector (***v'***, red) based on existing data points (blue) within local coordinates corresponding to a neighborhood on the manifold. **b:** Absolute differences between the estimated brain age and chronological age across the manifold. **c:** Number of subjects within a unit sphere for each point on the manifold (local density). **d:** Variance of chronological ages within each subject's neighborhood. **e:** MAEs of brain age estimation as a function of the number of subjects included on the manifold. **f:** 2D manifold constructed using white matter fractional anisotropy (FA) and mean diffusivity (MD). **g:** Brain age estimation based on white matter FA and MD showing strong correlation with chronological age. **h-i:** 2D manifold constructed using both regional volumes and white matter FA/MD (**h**), and corresponding brain age estimation performance (**i**).

The manifold encoded regional volume information

We also examined how regional volume information was encoded by positions on the manifold by treating volumes as smooth functions across the manifold. Whole brain volume varied gradually across the manifold (**Extended data Fig. 4a**), indicating that subjects with similar volumes clustered together. Leave-one-out validation revealed a strong correlation between estimated and actual whole brain volumes (**Extended data Fig. 4b**, R^2=0.99).

Individual structures showed lower but still strong correlations (**Extended data Fig. 4c-g**), potentially due to the regional volumes being noisier than whole brain volume. A 3D manifold further improved correlation strength and volume estimation accuracy (**Extended Data Fig. 5**), suggesting it captured additional structural information beyond the 2D representation.

Regional volumes that contribute to brain age estimation

Understanding how individual regional volumes contribute to estimated brain age is crucial for biological interpretability. The smooth gradients observed across the manifold (**Fig. 1d & Extended Data Fig. 4**) suggested that subjects within each neighborhood shared similar structural volumes and chronological ages, forming a relatively homogeneous subgroup of the HCP-A dataset. This local homogeneity provided the necessary conditions for approximating the relationship between regional volumes and age using a multivariable linear function within each neighborhood. To address collinearity among structural volumes (**Extended Data Fig. 6**), we employed partial least squares regression (PLSR) to estimate local relationships and ranked input markers based on variable importance in projection (VIP) scores (**Fig. 3a**). Seventy-two regions consistently maintained VIP scores above 1 across the manifold (**Fig. 3b-d**), where VIP > 1 is typically considered indicative of high impact. Among these, frontal sulcus and cerebellar gray matter exhibited the highest average VIP scores, followed by parietal sulcus, frontal lateral ventricles, superior temporal gyrus, and medial frontal gyrus (**Fig. 3e-h**). Brain age estimation using a new manifold constructed from these 72 key structures achieved an MAE of 6.1 years (**Extended Data Fig. 7a-b**), matching the performance of all 267 structures and confirming that these regions contained equivalent information for brain age estimation. Incorporating cognitive function data further reduced the MAE to 5.9 years (**Extended Data Fig. 7c-d**).

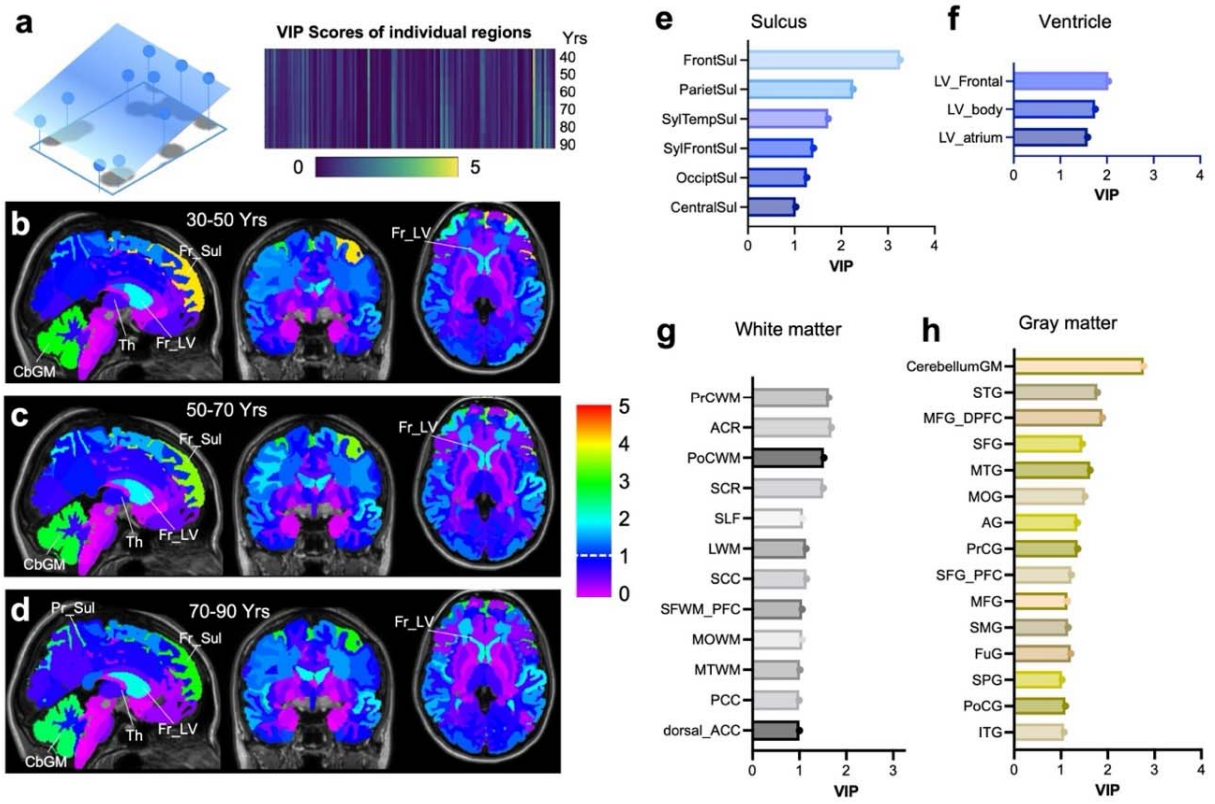

Figure 3: Local analysis of individual brain regions' contributions to brain age estimation. **a:** Schematic illustrating how the brain age function is locally approximated using multivariable linear functions, from which VIP scores are calculated for each brain structure. **b-d:** Maps of regional VIP scores for subjects in three age groups: 30-50 years, 50-70 years, and 70-90 years. **e-h:** Average VIP scores of the 72 high-impact structures organized into four anatomical categories, with scores shown for both left and right hemispheres. Brain region abbreviations are provided in **Supplementary Table 1**.

VIP scores of individual regions varied with age, reflecting their changing contributions to brain age estimation. The frontal lateral ventricle (LV) VIP score increased with age (**Fig. 4a**), consistent with its relatively stable volume in subjects aged 40-50 years followed by accelerated expansion at later ages (**Fig. 4b**). The frontal LV volume increased more rapidly in males than females, explaining males' higher VIP scores for this region. In contrast, two white matter structures—superior corona radiata (SCR) and anterior corona radiata (ACR)—showed steady age-related volume declines with relatively constant VIP scores. Interestingly, the thalamus

maintained a VIP score near one despite significant age-related changes, likely because its volume was highly correlated with other brain regions, including ACR and SCR (**Extended Data Fig. 8**). High correlations between regional volumes reduce individual VIP scores because correlated regions provide redundant rather than unique information for age prediction. Similarly, hippocampal volume had a lower VIP score due to strong correlations with middle temporal gyrus (MTG), ACR, and SCR volumes. A separate PLSR analysis revealed that 99% of hippocampal volume variation could be predicted from the 72 key structures, rendering it redundant for age estimation in this framework.

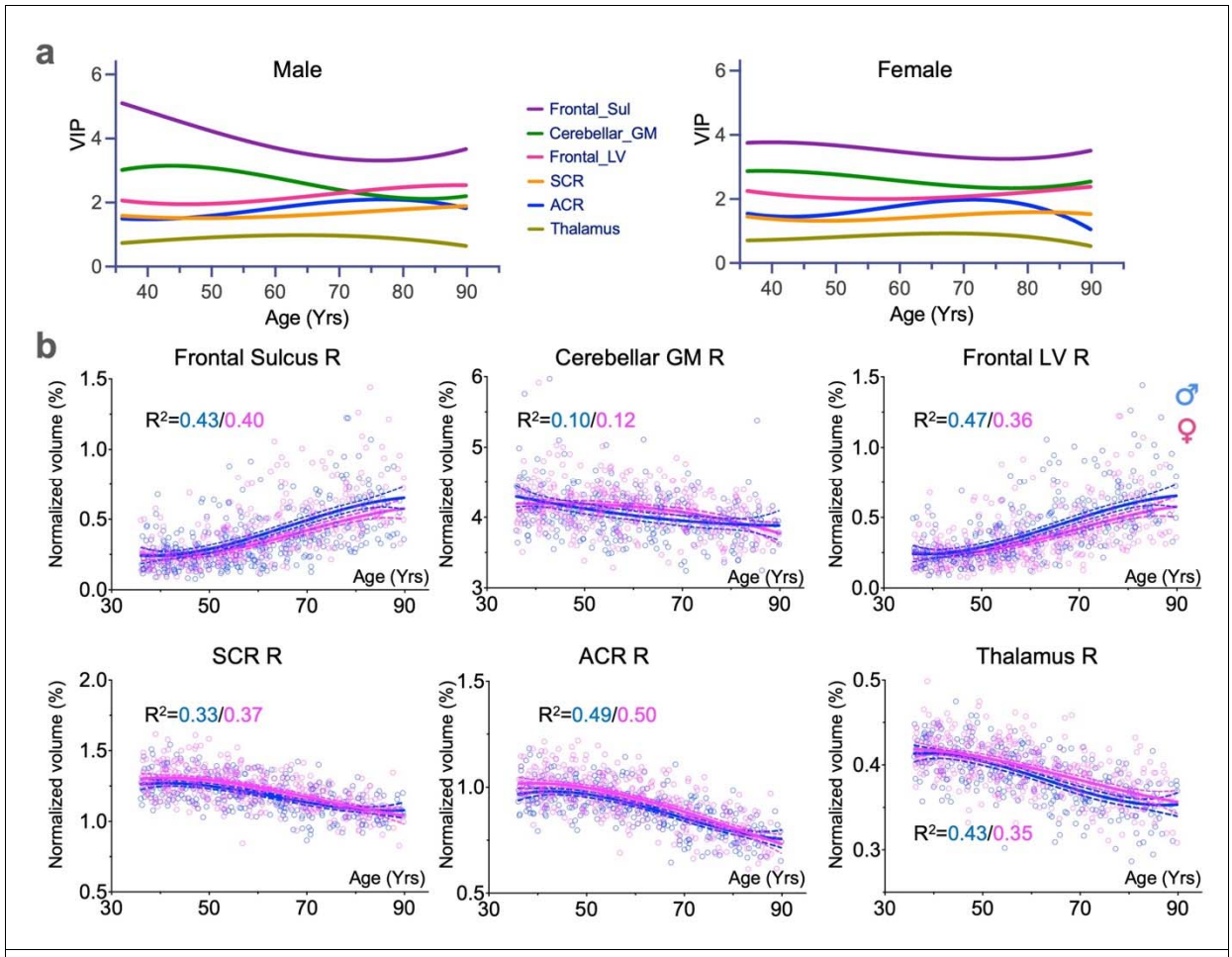

Figure 4: Age-dependent VIP scores in the HCP-A dataset. **a:** The age-related VIP score curves for selected brain regions in male and female HCP-A subjects. **b:** Scatter plots showing volume changes with age for six representative brain regions. Volumes of male (blue) and female (pink) subjects were

plotted separately.

Aging patterns on the manifold

The local aging patterns from individual neighborhoods can be aggregated to reveal broader trends. We applied K-means clustering on the coefficients of local PLSR of age, which were the regression weights and reflected how regional volume changes affect estimated brain age. We experimented with different numbers of clusters, finding that five clusters provided the most interpretable and stable results (**Fig. 5a**). The HCP-A subjects in each cluster occupied distinct and compact regions on the manifold (**Fig. 5b**), indicating that neighboring subjects shared similar broader aging patterns in regional volumes. Cluster 1 was associated with younger subjects, clusters 4 and 5 with older subjects, and clusters 2 and 3 with subjects spanning a wide age range.

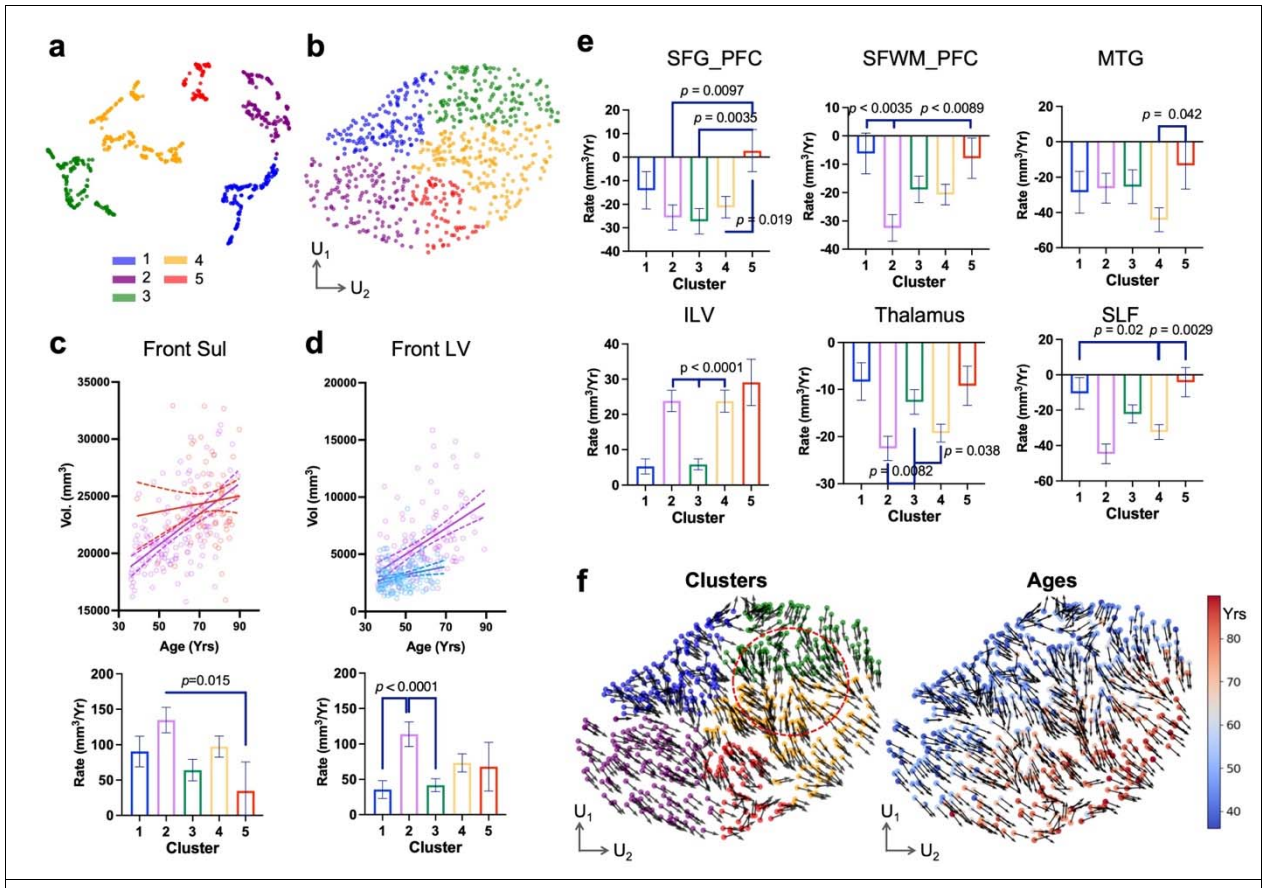

Figure 5: Distinct aging patterns on the manifold. **a:** Five clusters of HCP-A subjects based on locally estimated rates of volumetric changes in 72 brain structures with VIP scores greater than one. **b:** Distribution of the five clusters on the aging manifold. **c:** Frontal sulcus volume versus age for clusters 2 and 5, with linear regression fits showing estimated volumetric change rates (dashed lines indicate 95% confidence intervals). Cluster 2 exhibited significantly higher atrophy rates than cluster 5 (bottom panel). **d:** Frontal lateral ventricle volume versus age, demonstrating that cluster 2 showed significantly faster ventricular expansion than clusters 1 and 3. **e:** Comparison of volumetric change rates across key brain regions, revealing significant differences among the five clusters. **f:** Aging gradient vector field across the manifold with individual subjects colored by cluster assignment (left) and age (right).

For each cluster, we computed age-related volumetric change rates for high-VIP structures, exemplified by the frontal sulcus and frontal LV measurements (**Fig. 5c-d**). The five clusters exhibited distinct patterns of age-related volume changes. Cluster 1 (blue) and cluster 3 (green) showed moderate changes across most regions, though cluster 3 had notable changes in the prefrontal cortex (SFG_PFC). Cluster 5 (red) also exhibited moderate changes except in the inferior lateral ventricle (ILV). In contrast, clusters 2 and 4 demonstrated more pronounced

regional specificity: cluster 2 (purple) showed accelerated frontal atrophy with faster ventricular expansion (frontal LV, inferior LV), sulcal widening, and atrophy of the superior frontal white matter (SFWM) and superior longitudinal fasciculus (SLF) (**Fig. 5e-f**). Cluster 4 (yellow) primarily showed changes in the MTG and inferior LV. Both clusters 2 and 4 showed faster thalamic atrophy than the other clusters. We also visualized aging gradients—volume changes corresponding to the greatest age increase—as vector fields across the manifold based on local PLSR results (**Fig. 5f**). Most aging vectors aligned with the age gradient (**Fig. 1d**), but a bifurcation occurred between clusters 3 and 4 (red circle in **Fig. 5f**), suggesting divergent aging pathways.

Discussion

This study introduces a manifold-based framework for assessing individual brain aging based on regional volume measurements. This aging manifold serves as a low-dimensional map of plausible regional volume configurations in the aging human brain, where age-related volumetric changes manifest as trajectories across the manifold[22]. The approach enables computation of both overall brain age estimates and identification of key regional contributors that drive these aging patterns, thereby improving interpretability. The framework is adaptable to diverse MRI and non-MRI data and can reveal broad-scale aging patterns.

Brain aging is inherently heterogeneous: no single region adequately captures its complexity[32–36]. This challenge is further compounded by strong interregional correlations[37–39], which create redundancy in neuroimaging data and make dimension reduction essential for extracting latent variables that govern aging dynamics. Conventional methods such as PCA and independent component analysis have been previously applied[13,34,38,40–42], but their linear

assumptions primarily capture global structure and may overlook subtle, region-specific[27] or nonlinear patterns across subgroups[43,44]. Modern dimension reduction methods, including *t*-distributed stochastic neighbor embedding (*t*-SNE)[45] and UMAP[26], better preserve local topological relationships and can characterize nonlinear brain aging trajectories without linear constraints[46]. Several groups have demonstrated the utility of these methods in uncovering local structures within MRI datasets[22,23,47–50]. Building on this foundation, our aging manifold (**Fig. 1**) was derived from the HCP-A dataset, embedding high-dimensional regional volume measurements into 2D or 3D maps (**Extended Data Figures 4 & 5**). For visualization clarity, we focused on the 2D manifold in this study, though the methodology can be readily extended to 3D or higher-dimensional representations as needed.

Unlike the distinct clusters often observed in gene expression classification or similar datasets[51], the manifolds derived from HCP-A structural MRI data as well as those reported previously[22,23] are relatively compact and continuous, with no apparent discrete clusters (**Fig. 1c**). This likely reflects the gradual nature of age-related brain changes and the multifactorial influences of genetic, environmental, and demographic variables on brain morphology[52–54]. While HCP-A provides high quality MRI data from healthy aging participants, the manifold could be substantially improved by incorporating larger datasets (e.g. UK Biobank[55]) or harmonized multiple datasets[18,19]. Such integration would (1) increase subject density in local manifold regions to improve inference accuracy, and (2) expand representation of under-sampled populations, particularly those over 90 years of age. A notable advantage of this framework is its ability to guide strategic data collection by identifying sparsely populated regions, it can predict the specific demographic and morphometric characteristics (e.g., brain volume as in **Extended Data Figure 4**) needed to optimize manifold coverage. Additionally, incorporating pathological

cases from datasets such as ADNI would further extend the manifold beyond normal aging to encompass disease-related trajectories.

Although the manifold encodes topological rather than quantitative relationships, these relationships can be used for inference, as demonstrated by our brain age estimation analyses. Chronological age is the strongest known risk factor for late-onset AD[56,57], yet brain age more closely predicts disease likelihood[58–60]. Deep learning-based brain age models, however, often suffer from limited interpretability, which hinders generalizability. To address this, methods such as layer-wise relevance propagation[61] and occlusion sensitivity analysis[62], have been used to identify regional contributions to brain age estimation. While informative, these methods are challenged by the high dimensionality of MRI data, including computational burden, multiple comparisons, spatial correlations, and artifacts from unrealistic occlusion. Manifold-based brain age estimation addresses many of these issues by operating in a low-dimensional space grounded in realistic aging data. Beyond providing estimates, the manifold offers additional advantages: by locating individual subjects within the manifold, we can assess extrapolation risk (e.g., proximity to sparsely sampled regions or manifold boundaries) and derive confidence intervals for predictions. This will ensure that inferences are grounded in biologically plausible aging patterns. It is important to note, however, that manifold representations are not unique – they depend on both the dimension reduction algorithm and the underlying dataset characteristics – highlighting the importance of optimizing manifold construction as a key direction for future research.

Local analysis of individual neighborhoods on the manifold revealed region-specific volume changes that both confirm established findings and provide new insights into brain aging patterns. For example, the strong impact of cerebellar volume on brain age (**Fig. 3h**) is consistent with recent work by Raz et al.[36] and Bernard et al.[63], who demonstrated significant cerebellar atrophy

in advancing aging – challenging earlier views that the cerebellum is largely spared during normal aging. Similarly, the pronounced contribution of frontal sulcus volume reflects frontal lobe atrophy (**Fig. 3e**), supporting the "last-in, first-out" hypothesis described by Fjell et al.[1,64], in which late-developing brain regions show earlier decline. Our observation of lateral ventricle enlargement (**Fig. 3g**) with age represents one of the frequently reported neuroimaging biomarkers of brain aging. White matter changes observed in this study provide further evidence of selective vulnerability: we found atrophy in the ACR and SCR, aligning with previous structural and diffusion MRI studies[65,66] that identified these projection fiber tracts as particularly vulnerable to age-related degeneration. Such vulnerability supports the disconnection hypothesis[67], which posits that age-related disruption of frontal-subcortical pathways contributes to cognitive decline, particularly in executive function and processing speed.

Our findings also align with prior data-driven reports that identified varying degrees of frontal, temporal, and subcortical atrophy as central markers for characterizing both heterogeneous normal aging and AD-related trajectories[18,19]. Although the relatively small sample size of our normal aging dataset limited sensitivity to subtle aging patterns, the results nonetheless echoed earlier studies and revealed several interesting aging patterns. For example, cluster 2 (**Fig. 5**), characterized by marked frontal atrophy, included a higher proportion of male subjects – consistent with previous reports showing stronger frontal atrophy in males than in females[68]. In contrast, clusters 3 and 4, colocalized with many of the female subjects, showed moderate and then more severe atrophy, especially in the temporal regions. This pattern may reflect hormonal changes associated with menopause, as demonstrated before[69].

One limitation of this study is its reliance solely on regional volume measurements for manifold construction, brain age estimation, and subsequent analyses. Incorporating additional

MRI markers (e.g., cortical thickness, intensity, microstructure, and connectivity) or non-MRI biomarkers would likely enhance both the aging manifold's representational capacity and brain age estimation accuracy[8,70]. Interestingly, our attempts to integrate both structural and diffusion MRI yielded inferior results compared to using structural MRI alone. This unexpected finding may stem from residual mismatches between T_1-weighted and diffusion MRI modalities, as well as partial volume effects that disproportionately impact diffusion metrics in complex tissue boundaries[71,72].

The local analysis assumes that structural and age similarities within a neighborhood are sufficient to establish homogeneity for aging analysis. While making the neighborhood smaller may better satisfy the homogeneity condition, this approach reduces the number of subjects per neighborhood and consequently diminishes the statistical power. Maintaining adequate statistical power requires large population datasets, yet subjects may still vary substantially in other critical factors—including genetics, health history, and lifestyle—that influence aging patterns.

In conclusion, this study introduces the aging manifold as a novel computational framework for evaluating individual brain aging. By enhancing interpretability, the methodology allows researchers to identify the specific factors that drive individual differences in brain aging patterns. This transparent framework not only advances our ability to model the complex, multidimensional nature of brain aging but also yields clinically relevant insights into the heterogeneous pathways through which aging manifests across diverse populations.

Methods

Dataset

This study included 725 participants from the publicly available HCP-Aging database, Lifespan 2.0 release (https://www.humanconnectome.org/study/hcp-lifespan-aging/document/hcp-aging-20-release). Among the participants, 35 subjects were later excluded due to poor MRI data segmentation results as described later. The remaining 690 subjects ranged in age from 36 to 100 years (mean = 60.0 years, SD = 15.0 years) and was approximately balanced by sex (Male/Female =319/371). All participants were free of major neurological or psychiatric conditions (e.g., stroke, brain tumor, clinically diagnosed dementia). Detailed demographic information of HCP-Aging can be found in prior publications[24]. The study protocol was approved by the institutional review board, and all participants provided written informed consent. The HCP-Aging data represents normal aging without known pathological conditions that affect cognitive abilities.

Cognitive Function Assessment

Cognitive measures from the HCP-Aging dataset (https://www.humanconnectome.org/storage/app/media/documentation/LS2.0/LS_2.0_Release_Appendix_2.pdf) were incorporated into the UMAP embedding alongside regional brain volume features. The cognitive variables included Fluid Cognition Composite scores, Crystallized Cognition Composite scores, and Total Cognition Composite scores, all derived from the NIH Toolbox Cognition Battery[73,74]. Cognitive measures were obtained from the NIH Toolbox in the HCP-Aging dataset, including the Fluid, Crystallized, and Total Cognition Composites. The Fluid Composite assesses processing speed,

working memory, episodic memory, and executive function, while the Crystallized Composite reflects premorbid intellectual abilities such as vocabulary knowledge and oral reading recognition. To capture both raw performance and age-normed scaling, we included six variables: nih_fluidcogcomp_unadjusted, nih_fluidcogcomp_np, nih_eccogcomp_unadjusted, nih_earlychildcogcomp_np, nih_totalcogcomp_unadjusted, and nih_totalcogcomp_np.

MRI

We only used T_1-weighted and multi-shell diffusion MRI data from the HCP-Aging dataset in this study. MRI data were acquired across four sites using Siemens Prisma 3 Tesla scanners equipped with 32-channel head coils, all running the same software version (E11C). A harmonized, state-of-the-art imaging protocol was implemented across sites, with standardized procedures, centralized training, and ongoing quality control analyses ensuring high inter-site consistency. T_1-weighted MRI data was obtained using a magnetization-prepared rapid acquisition gradient echo sequence with the following parameters: voxel size = 0.8 × 0.8 × 0.8 mm^3, echo times (TE) = 1.8/3.6/5.4/7.2 ms, repetition time (TR) = 2,500 ms, inversion time (TI) = 1,000 ms, flip angle = 8°, number of echoes = 4. Diffusion MRI data were acquired using a multiband 2D spin echo EPI sequence with the following parameters: TR/TE = 3,230 ms/89.20 ms, voxel size = 1.5 × 1.5 × 1.5 mm^3, multiband acceleration factor = 4, b-values = 1,500 and 3,000 s/mm^2 (92∼93 directions per shell) with 28 non-diffusion-weighted (b = 0 s/mm^2) images. Maps of fractional anisotropy (FA) and mean diffusivity (MD) were computed from diffusion MRI data using MRtrix 3.0 (www.mrtrix.org).

Image segmentation and regional volumes

T_1-weighted brain images were segmented using MRIcloud[75] into 286 regions according to the Johns Hopkins adult brain atlas[76]. The quality of segmentation was visually inspected, and data

from 35 subjects were excluded due to poor segmentation quality, and data from the remaining 690 subject were used in subsequent analysis. Among the 286 regions, the volumes of non-brain skull, optic tract, and bone marrow were excluded in this study. We also excluded the bilateral fimbria and rostral anterior cingulate white matter, which were not properly segmented in some subjects. We used the remaining 267 regional values for analysis. To align structural and diffusion data, T_1-weighted images were registered to the mean diffusion-weighted image using the ANTs Symmetric Normalization (SyN) algorithm[77].

UMAP

We computed the aging manifold using our in-house Python code built upon the open-source UMAP library (version 0.5.8; https://umap-learn.readthedocs.io)[26]. A 267x690 matrix of subjects' regional brain volumes was projected onto a low-dimensional manifolds using UMAP while preserving the local neighborhood structure of the high-dimensional volumetric data. The hyperparameters were tuned to favor settings that yielded an approximately uniform density of sample points within the boundary. After embedding a set of feature vectors in a low-dimensional space, UMAP also supports the online embedding of novel biomarker vectors from new subjects.

Estimation of brain age and volume

A linear regression age prediction model was built using the low-dimensional manifold as the feature space. We used a distance-weighted kernel regression model with a kernel size (k) of 5, chosen to balance bias and variance: smaller ks produced unstable predictions due to noise, while larger ks overly smoothed local aging patterns.

To evaluate the performance of age prediction, we employed a repeated leave-one-out strategy. In each run, *n* subjects were randomly selected to be placed on the manifold as samples, and the remainder *m* subjects were used for testing. The sample/test splits were stratified by age and sex, and random seeds were controlled for reproducibility. This procedure was run multiple times, and out-of-sample predictions were aggregated across runs to obtain final estimates. Model performance was quantified using mean absolute error and R^2. The same method was also applied to estimate regional volumes. All analyses were performed in Python 3.11 using the scikit-learn and matplotlib libraries.

Partial least squares regression (PLSR)

PLSR projects regional volumes and chronological ages into respective latent spaces via linear transformations that maximize the covariance between the projected volumes and ages. For each subject on the manifold, we identified 50 nearest neighbors using the Ball Tree search. The neighborhood size from 25 to 100 were tested, and 50 was chosen to balance the need to avoid overfitting while preserving regional specificity. PLSR was performed with the number of latent variables set to 2~5. The VIP score quantifies each region's contribution to explaining variation in age, and regions with VIP scores above 1 were considered important. If a region's VIP score exceeded 1 over the entire manifold, we classified it as a globally important feature.

Clustering of the local PLSR results

The regression coefficients (βs) from local PLSR quantify regional volumes' association with age. Clustering of local βs on the manifold was performed using UMAP with the following parameters: n_components = 2, n_neighbors = 15, min_dist = 0.1, and random_state = 2. We then applied k-means clustering (n_clusters = 3~6, random_state = 30) to partition the embedding into discrete groups, and n_clusters = 5 produced more robust than others.

Generation of the aging gradient vector field

Within each neighborhood on the manifold, we performed linear regression to find a linear approximation of chronological age as a function of the two UMAP coordinates. The local gradient vectors from the regression indicated the direction of volume changes that corresponded to the fastest aging. These gradient vectors were normalized to the same length to visualize directions rather than magnitudes, which may not be directly comparable on different parts of the manifold.

Acknowledgments

Funding: National Institutes of Health (U24NS135568, R01HD074593, P41EB017183)

Author contributions: Conceptualization: ZL, CJS, JZ; Methodology: ZL, CJS, CL, JL, JZ; Investigation: ZL, CJS, CL, JL, JZ, HR, YG; Visualization: ZL, CJS, JZ; Supervision: JZ; Writing—original draft: ZL, CJS, JZ; Writing—review & editing: ZL, CJS, JZ, HR, YG

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are available from HCP and ADNI websites.

References

1. Fjell, A. M. *et al.* What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Prog. Neurobiol.* **117**, 20–40 (2014).
2. Roe, J. M. *et al.* Brain change trajectories in healthy adults correlate with Alzheimer's related genetic variation and memory decline across life. *Nat. Commun.* **15**, 10651 (2024).
3. Leuba, G. *et al.* Differential damage in the frontal cortex with aging, sporadic and familial Alzheimer's disease. *Brain Res. Bull.* **80**, 196–202 (2009).
4. MacDonald, M. E. & Pike, G. B. MRI of healthy brain aging: A review. *NMR Biomed.* **34**, e4564 (2021).
5. Walhovd, K. B. *et al.* Effects of age on volumes of cortex, white matter and subcortical structures. *Neurobiol. Aging* **26**, 1261–1270; discussion 1275-1278 (2005).
6. Nyberg, L. *et al.* Individual differences in brain aging: heterogeneity in cortico-hippocampal but not caudate atrophy rates. *Cereb. Cortex* **33**, 5075–5081 (2023).
7. Verdi, S., Marquand, A. F., Schott, J. M. & Cole, J. H. Beyond the average patient: how neuroimaging models can address heterogeneity in dementia. *Brain* **144**, 2946–2953 (2021).
8. Cole, J. H. *et al.* Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage* **163**, 115–124 (2017).
9. Cole, J. H. *et al.* Brain age predicts mortality. *Mol. Psychiatry* **23**, 1385–1392 (2018).
10. Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C. & Mechelli, A. Machine learning for brain age prediction: Introduction to methods and clinical applications. *EBioMedicine* **72**, 103600 (2021).
11. Dörfel, R. P. *et al.* Prediction of brain age using structural magnetic resonance imaging: A comparison of accuracy and test–retest reliability of publicly available software packages. *Hum. Brain Mapp.* **44**, 6139–6148 (2023).
12. Vinke, E. J. *et al.* Trajectories of imaging markers in brain aging: the Rotterdam Study. *Neurobiol. Aging* **71**, 32–40 (2018).

13. Duan, H. *et al.* Population clustering of structural brain aging and its association with brain development. Preprint at https://doi.org/10.7554/eLife.94970.2 (2024).
14. Fujita, S. *et al.* Characterization of Brain Volume Changes in Aging Individuals With Normal Cognition Using Serial Magnetic Resonance Imaging. *JAMA Netw. Open* **6**, e2318153 (2023).
15. Sørensen, Ø., Walhovd, K. B. & Fjell, A. M. A recipe for accurate estimation of lifespan brain trajectories, distinguishing longitudinal and cohort effects. *NeuroImage* **226**, 117596 (2021).
16. Bethlehem, R. a. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
17. Alexander-Bloch, A., Giedd, J. N. & Bullmore, E. Imaging structural co-variance between human brain regions. *Nat. Rev. Neurosci.* **14**, 322–336 (2013).
18. Zhang, X. *et al.* Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proc. Natl. Acad. Sci.* **113**, (2016).
19. Yang, Z. *et al.* Brain aging patterns in a large and diverse cohort of 49,482 individuals. *Nat. Med.* **30**, 3015–3026 (2024).
20. Brahma, P. P., Wu, D. & She, Y. Why Deep Learning Works: A Manifold Disentanglement Perspective. *IEEE Trans. Neural Netw. Learn. Syst.* **27**, 1997–2008 (2016).
21. Fefferman, C., Mitter, S. & Narayanan, H. Testing the manifold hypothesis. *J. Am. Math. Soc.* **29**, 983–1049 (2016).
22. Lamontagne-Caron, R., Desrosiers, P., Potvin, O., Doyon, N. & Duchesne, S. Predicting cognitive decline in a low-dimensional representation of brain morphology. *Sci. Rep.* **13**, 16793 (2023).
23. van der Haar, D., Moustafa, A., Warren, S. L., Alashwal, H. & van Zyl, T. An Alzheimer's disease category progression sub-grouping analysis using manifold learning on ADNI. *Sci. Rep.* **13**, 10483 (2023).
24. Bookheimer, S. Y. *et al.* The Lifespan Human Connectome Project in Aging: An overview. *NeuroImage* **185**, 335–348 (2019).

25. Cole, J. H. & Franke, K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends Neurosci.* **40**, 681–690 (2017).
26. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://doi.org/10.48550/ARXIV.1802.03426 (2018).
27. Xifra-Porxas, A., Ghosh, A., Mitsis, G. D. & Boudrias, M.-H. Estimating brain age from structural MRI and MEG data: Insights from dimensionality reduction techniques. *NeuroImage* **231**, 117822 (2021).
28. Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80**, 62–79 (2013).
29. Jack, C. R. *et al.* The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging JMRI* **27**, 685–691 (2008).
30. 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 J. Neurol.* **143**, 2312–2324 (2020).
31. Yin, C. *et al.* Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment. *Proc. Natl. Acad. Sci.* **120**, e2214634120 (2023).
32. Liem, F. *et al.* Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage* **148**, 179–188 (2017).
33. Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E. & Miller, K. L. Estimation of brain age delta from brain imaging. *NeuroImage* **200**, 528–539 (2019).
34. Groves, A. R. *et al.* Benefits of multi-modal fusion analysis on a large-scale dataset: Life-span patterns of inter-subject variability in cortical morphometry and white matter microstructure. *NeuroImage* **63**, 365–380 (2012).
35. Raz, N. *et al.* Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers. *Cereb. Cortex* **15**, 1676–1689 (2005).

36. Raz, N., Ghisletta, P., Rodrigue, K. M., Kennedy, K. M. & Lindenberger, U. Trajectories of brain aging in middle-aged and older adults: Regional and individual differences. *NeuroImage* **51**, 501–511 (2010).
37. Salat, D. H. Thinning of the Cerebral Cortex in Aging. *Cereb. Cortex* **14**, 721–730 (2004).
38. Douaud, G. *et al.* A common brain network links development, aging, and vulnerability to disease. *Proc. Natl. Acad. Sci.* **111**, 17648–17653 (2014).
39. Cox, S. R. *et al.* Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nat. Commun.* **7**, 13629 (2016).
40. Liu, X., Tyler, L. K., Cam-CAN, Rowe, J. B. & Tsvetanov, K. A. Multimodal fusion analysis of functional, cerebrovascular and structural neuroimaging in healthy aging subjects. *Hum. Brain Mapp.* **43**, 5490–5508 (2022).
41. Franke, K., Ziegler, G., Klöppel, S., Gaser, C., & Alzheimer's Disease Neuroimaging Initiative. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *NeuroImage* **50**, 883–892 (2010).
42. Doan, N. T. *et al.* Distinguishing early and late brain aging from the Alzheimer's disease spectrum: consistent morphological patterns across independent samples. *NeuroImage* **158**, 282–295 (2017).
43. Antal, B. B. *et al.* Brain aging shows nonlinear transitions, suggesting a midlife "critical window" for metabolic intervention. *Proc. Natl. Acad. Sci.* **122**, e2416433122 (2025).
44. Fjell, A. M. *et al.* Critical ages in the life course of the adult brain: nonlinear subcortical aging. *Neurobiol. Aging* **34**, 2239–2247 (2013).
45. Hinton, G. E. & Roweis, S. Stochastic Neighbor Embedding. in *Advances in Neural Information Processing Systems* (eds Becker, S., Thrun, S. & Obermayer, K.) vol. 15 (MIT Press, 2002).
46. Kobak, D. & Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **10**, 5416 (2019).

47. Park, H. ISOMAP induced manifold embedding and its application to Alzheimer's disease and mild cognitive impairment. *Neurosci. Lett.* **513**, 141–145 (2012).
48. Liu, S. *et al.* Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. *Sci. Rep.* **12**, 17106 (2022).
49. Dadu, A. *et al.* Application of Aligned-UMAP to longitudinal biomedical studies. *Patterns* **4**, 100741 (2023).
50. Chen, A. A. *et al.* Deconfounded Dimension Reduction via Partial Embeddings. Preprint at https://doi.org/10.1101/2023.01.10.523448 (2023).
51. Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S. & Trapnell, C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nat. Commun.* **11**, 1537 (2020).
52. Manuello, J. *et al.* The effects of genetic and modifiable risk factors on brain regions vulnerable to ageing and disease. *Nat. Commun.* **15**, 2576 (2024).
53. Kremen, W. S. *et al.* Genetic and environmental influences on the size of specific brain regions in midlife: The VETSA MRI study. *NeuroImage* **49**, 1213–1223 (2010).
54. Jansen, A. G., Mous, S. E., White, T., Posthuma, D. & Polderman, T. J. C. What twin studies tell us about the heritability of brain development, morphology, and function: a review. *Neuropsychol. Rev.* **25**, 27–46 (2015).
55. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
56. 2023 Alzheimer's disease facts and figures. *Alzheimers Dement.* **19**, 1598–1695 (2023).
57. Fang, M. *et al.* Lifetime risk and projected burden of dementia. *Nat. Med.* (2025) doi:10.1038/s41591-024-03340-9.
58. Gaser, C. *et al.* BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease. *PLoS ONE* **8**, e67346 (2013).

59. Löwe, L. C., Gaser, C., Franke, K., & for the Alzheimer's Disease Neuroimaging Initiative. The Effect of the APOE Genotype on Individual BrainAGE in Normal Aging, Mild Cognitive Impairment, and Alzheimer's Disease. *PLOS ONE* **11**, e0157514 (2016).
60. Morris, J. C. Is Alzheimer's disease inevitable with age? *J. Clin. Invest.* **104**, 1171–1173 (1999).
61. Hofmann, S. M. *et al.* Towards the interpretability of deep learning models for multi-modal neuroimaging: Finding structural changes of the ageing brain. *NeuroImage* **261**, 119504 (2022).
62. He, L. *et al.* Network Occlusion Sensitivity Analysis Identifies Regional Contributions to Brain Age Prediction. Preprint at https://doi.org/10.1101/2022.10.31.514506 (2022).
63. Bernard, J. A. & Seidler, R. D. Moving forward: age effects on the cerebellum underlie cognitive and motor declines. *Neurosci. Biobehav. Rev.* **42**, 193–207 (2014).
64. Fjell, A. M. *et al.* High consistency of regional cortical thinning in aging across multiple samples. *Cereb. Cortex N. Y. N 1991* **19**, 2001–2012 (2009).
65. Ly, M. *et al.* Midlife measurements of white matter microstructure predict subsequent regional white matter atrophy in healthy adults. *Hum. Brain Mapp.* **35**, 2044–2054 (2014).
66. Giorgio, A. *et al.* Age-related changes in grey and white matter structure throughout adulthood. *NeuroImage* **51**, 943–951 (2010).
67. Madden, D. J. *et al.* Influence of structural and functional brain connectivity on age-related differences in fluid cognition. *Neurobiol. Aging* **96**, 205–222 (2020).
68. Xu, J. *et al.* Gender effects on age-related changes in brain structure. *AJNR Am. J. Neuroradiol.* **21**, 112–118 (2000).
69. Steventon, J. J. *et al.* Menopause age, reproductive span and hormone therapy duration predict the volume of medial temporal lobe brain structures in postmenopausal women. *Psychoneuroendocrinology* **158**, 106393 (2023).

70. Franke, K. & Gaser, C. Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? *Front. Neurol.* **10**, 789 (2019).
71. Metzler-Baddeley, C., O'Sullivan, M. J., Bells, S., Pasternak, O. & Jones, D. K. How and how not to correct for CSF-contamination in diffusion MRI. *NeuroImage* **59**, 1394–1403 (2012).
72. Vos, S. B., Jones, D. K., Viergever, M. A. & Leemans, A. Partial volume effect as a hidden covariate in DTI analyses. *NeuroImage* **55**, 1566–1576 (2011).
73. Weintraub, S. *et al.* The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: validation in an adult sample. *J. Int. Neuropsychol. Soc. JINS* **20**, 567–578 (2014).
74. Gershon, R. C. *et al.* NIH Toolbox for Assessment of Neurological and Behavioral Function. *Neurology* **80**, (2013).
75. Mori, S. *et al.* MRICloud: Delivering High-Throughput MRI Neuroinformatics as Cloud-Based Software as a Service. *Comput. Sci. Eng.* **18**, 21–35 (2016).
76. Oishi, K. *et al.* Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. *NeuroImage* **46**, 486–499 (2009).
77. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).