Title

Preliminary validation of a structural magnetic resonance imaging metric for tracking dementia-related neurodegeneration and future decline

Authors

Kress, Gavin T
Popa, Emily S
Thompson, Paul M
et al.

Publication Date

2023

DOI

10.1016/j.nicl.2023.103458

Peer reviewed

NeuroImage: Clinical 39 (2023) 103458

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Preliminary validation of a structural magnetic resonance imaging metric for tracking dementia-related neurodegeneration and future decline

Gavin T. Kress [a,b,1], Emily S. Popa [a,1], Paul M. Thompson [c], Susan Y. Bookheimer [a,d], Sophia I. Thomopoulos [c], Christopher R.K. Ching [c], Hong Zheng [c], Daniel A. Hirsh [a], David A. Merrill [a,e,f], Stella E. Panos [a], Cyrus A. Raji [g], Prabha Siddarth [a,f], Jennifer E. Bramen [a,*]

[a] Pacific Brain Health Center, Pacific Neuroscience Institute Foundation, Santa Monica, CA 90404, USA
[b] Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA
[c] Imaging Genetics Center, Mark and Mary Stevens Neuroimaging & Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA 90292, USA
[d] David Geffen School of Medicine, University of California, Los Angeles, Westwood, CA 90095, USA
[e] Department of Translational Neurosciences and Neurotherapeutics, Providence Saint John's Cancer Institute, Santa Monica, CA 90404, USA
[f] UCLA Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Westwood, CA 90095, USA
[g] Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO, USA

ARTICLE INFO

Keywords:
Mild cognitive impairment
Alzheimer's disease
Dementia
Magnetic resonance imaging
Biomarker

ABSTRACT

Alzheimer's disease (AD) is a progressive neurodegenerative disease characterized by cognitive decline and atrophy in the medial temporal lobe (MTL) and subsequent brain regions. Structural magnetic resonance imaging (sMRI) has been widely used in research and clinical care for diagnosis and monitoring AD progression. However, atrophy patterns are complex and vary by patient. To address this issue, researchers have made efforts to develop more concise metrics that can summarize AD-specific atrophy. Many of these methods can be difficult to interpret clinically, hampering adoption.

In this study, we introduce a novel index which we call an "AD-NeuroScore," that uses a modified Euclidean-inspired distance function to calculate differences between regional brain volumes associated with cognitive decline. The index is adjusted for intracranial volume (ICV), age, sex, and scanner model. We validated AD-NeuroScore using 929 older adults from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, with a mean age of 72.7 years (SD = 6.3; 55.1–91.5) and cognitively normal (CN), mild cognitive impairment (MCI), or AD diagnoses.

Our validation results showed that AD-NeuroScore was significantly associated with diagnosis and disease severity scores (measured by MMSE, CDR-SB, and ADAS-11) at baseline. Furthermore, baseline AD-NeuroScore was associated with both changes in diagnosis and disease severity scores at all time points with available data. The performance of AD-NeuroScore was equivalent or superior to adjusted hippocampal volume (AHV), a widely used metric in AD research. Further, AD-NeuroScore typically performed as well as or sometimes better when compared to other existing sMRI-based metrics.

Abbreviations: AD, Alzheimer's Disease; ADAS-11, Alzheimer's Disease Assessment Scale-Cognitive Subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; AD-PS, Alzheimer's Disease Pattern Similarity; aMCI, Amnestic Mild Cognitive Impairment; AUC-ROC, Area Under the Receiver Operator Characteristic Curve; ANOVA, Analysis of Variance; CSF, Cerebrospinal Fluid; CDR-SB, Clinical Dementia Rating Scale Sum of Boxes; CN, Cognitively Normal; DTI, Diffusion Tensor Imaging; ENIGMA, Enhancing Neuro Imaging Genetics Meta-Analysis; FA, Fractional Anisotropy; FTD, Frontotemporal Lobe Dementia; GM, Grey Matter; ML, Machine Learning; MTL, Medial Temporal Lobe; MCI, Mild Cognitive Impairment; MMSE, Mini-Mental State Exam; MRDATS, MRI-Based Dementia of Alzheimer's Type Score; PHI, Protected Health Information; ROC, Receiver Operator Characteristic; ROI, Region of Interest; RVI, Regional Vulnerability Index; SD, Standard Deviation; SPARE-AD, Spatial Pattern of Abnormality for Recognition of Early Alzheimer's Disease; STAND, STructural Abnormality iNDex; sMRI, Structural Magnetic Resonance imaging; SuStaIn, Subtype and Stage Inference; UMI, Univariate Morphometry Index; VBM, Voxel-Based Morphometry; WM, White Matter; ZWE, Z-Weighted Euclidean.

* Corresponding author at: Pacific Brain Health Center and Pacific Neuroscience Institute, 1301 20th Street, Suite 250, Santa Monica, CA 90404, USA.
E-mail addresses: daniel@hirschtechsolutions.com (D.A. Hirsh), jbramen@pacificneuro.org (J.E. Bramen).

[1] Gavin Kress and Emily Popa shared equally to the joint first authorship.

https://doi.org/10.1016/j.nicl.2023.103458
Received 26 May 2023; Accepted 20 June 2023
Available online 26 June 2023

In conclusion, we have introduced a new metric, AD-NeuroScore, which shows promising results in detecting AD, benchmarking disease severity, and predicting disease progression. AD-NeuroScore differentiates itself from other metrics by being clinically practical and interpretable.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease characterized by initial cognitive decline and early atrophy of the medial temporal lobe (MTL) followed by atrophy of other brain regions. Neurodegeneration often precedes cognitive decline (Gómez-Isla et al., 1996), making structural magnetic resonance imaging (sMRI) an ideal clinical biomarker for detecting AD early and predicting future cognitive and functional decline, as it is a passive, widely accessible, and cost-effective measurement. Improving quantitative sMRI reporting could enhance its value in clinical and research settings, and detecting AD before irreversible neurodegeneration could improve the efficacy of available treatments (Coupé et al., 2015).

Extensive research efforts have been directed towards understanding AD-related neurodegeneration (Chen et al., 2007), with significant focus on the hippocampus and the MTL. While hippocampal or other MTL regions are often used as sMRI metrics due to their strong associations with cognition, diagnosis, and AD-etiology (Gosche et al., 2002; Jack et al., 2002; Csernansky et al., 2004; Jack et al., 1997), other parts of the brain also provide important information for differential diagnosis of mild cognitive impairment (MCI), AD, and frontotemporal lobe dementia (FTD) (Rabinovici et al., 2008). Over time, neurodegeneration can spread from the MTL (Rabinovici et al., 2008) to other brain regions (Gómez-Isla et al., 1996; Rabinovici et al., 2008; Thompson et al., 2003; Braak et al., 1997; Frisoni et al., 1999; Thompson et al., 2007; Laakso et al., 1996; Laakso et al., 1998; Dickerson et al., 2001; Vercelletto et al., 2002), including the parietal lobe (Lindeboom and Weinstein, 2004; Cabeza et al., 2008; Jacobs et al., 2012), posterior temporal, lateral occipital (Rabinovici et al., 2008), left frontal regions, as well as limbic structures including the thalamus, cingulate gyrus, and nucleus accumbens (Nie et al., 2017).

Medical imaging companies have implemented regional brain volume reports in patient care settings to help clinicians understand the overall pattern of neurodegeneration (Ahdidan et al., 2015; Brewer et al., 2009; Cavedo et al., 2022). However, assessing a patient's overall neurodegeneration from several numbers can be difficult, as the spatiotemporal pattern of atrophy varies widely among patients. Additionally, the current tools do not account for differences in scans performed at different facilities using different scanners, which can affect the validity of longitudinal follow-up reports. Therefore, a valid, harmonized summary score utilizing multiple brain atrophy-related features is needed to complement existing regional brain volume reports. Such a score could enhance clinical comparability and diagnostic sensitivity, while allowing for the integration of measures across research settings and facilitating *meta*-analyses.

Early sMRI-based, multivariate approaches primarily focused on classifying cognitively normal (CN) and AD (Diciotti et al., 2012; Vemuri et al., 2009). However, as the field has progressed, researchers have expanded their focus to include classification of more disease stages (i.e., CN, MCI, and AD) (Rallabandi et al., 2020; Popuri et al., 2020) and predicting which patients will stabilize or decline (Coupé et al., 2015; Popuri et al., 2020; Ezzati et al., 2019; Wang et al., 2021). Many of these multivariate scoring methods use machine learning (Diciotti et al., 2012; Vemuri et al., 2009; Casanova et al., 2018) and some combine sMRI with other imaging, biochemical, demographic, or clinical features to improve performance (Vemuri et al., 2009; Dukart et al., 2013; Salvatore et al., 2018). The STructural Abnormality iNDex (STAND)-score (Vemuri et al., 2008; Vemuri et al., 2008), AD Pattern Similarity (AD-PS) (Casanova et al., 2018; Casanova et al., 2013), Subtype and Stage Inference (SuStaIn) (Young et al., 20142014; Archetti et al., 2021; Young et al., 2021); MRI-Based Dementia of Alzheimer's Type Score (MRDATS) (Popuri et al., 2020), Spatial Pattern of Abnormality for Recognition of Early AD (SPARE-AD) (Davatzikos et al., 2009; Da et al., 2014), Univariate Morphometry Index (UMI) (Wang et al., 2021), and Regional Vulnerability Index (RVI) or the ENIGMA Dot Product (Kochunov et al., 2022) are examples of such metrics. These scores demonstrate that a multivariate summary approach offers promise for clinical diagnosis and research purposes.

However, the implementation of existing metrics in neurological practices is limited due to the requirement for quality control in radiology-based measures (Larson and Boland, 2019). Furthermore, these metrics lack both patient-level and model-level interpretability (Pinto et al., 2022). In contrast, narrowly focused region-of-interest (ROI)-based reports are commonly used in clinical practice because they allow for control over feature segmentation quality and are interpretable (Coupé et al., 2015; Ahdidan et al., 2015; Brewer et al., 2009; Casanova et al., 2013; Archetti et al., 2021; Davatzikos et al., 2009; Achterberg et al., 2014; Sørensen et al., 2016). However, it is important to note that existing interpretable ROI-based approaches generally demonstrate inferior performance compared to more complex metrics.

The objective of this study is to develop and validate a clinically translatable biomarker which we call an "AD-NeuroScore." This score summarizes AD-related neurodegeneration using only sMRI-measured regional volumes, which are currently integrated within clinical electronic health records (Ahdidan et al., 2015; Brewer et al., 2009; Cavedo et al., 2022) and can be visually inspected by a radiologist. We chose these features because simplicity (Hwang and Park, 2020), clinical availability (Eweje et al., 2022), quality controllability (Larson and Boland, 2019), and interpretability (Pinto et al., 2022) are crucial for clinical adoption. Our approach involves creating a vector of region of interest (ROI)-based features that are associated with cognitive decline and calculating a computationally inexpensive summary score. The relative contribution of atrophy within each feature is transparent and holds clinical significance. We believe that these advantages will facilitate widespread clinical deployment and research applications.

2. Methods

2.1. Participants

2.1.1. Data acquisition and demographics

CN individuals and patients with MCI or AD diagnosis were drawn from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) (Mueller et al., 2005; Mueller et al., 2005; Jack et al., 2008). The ADNI is a global research study launched in 2003, primarily aimed at evaluating biological markers to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org.

All ADNI studies were conducted according to the Good Clinical Practice guidelines, the Declaration of Helsinki, the U.S. 45 Code of Federal Regulations (CFR) Part 46 and 21 CFR Part 50 – Protection of Human Subjects, and 21 CFR Part 56 - Institutional Review Boards. Written informed consent and HIPAA authorizations were obtained from all participants or authorized representatives prior to the conducting of protocol-specific procedures. The ADNI protocol was approved by the Institutional Review Boards of all participating institutions, listed in File 1 in Supplementary Materials (Mukherji et al., 2021).

In this work, a total of 1,619 subjects with available 3 T, accelerated, T1-weighted MR images at baseline were collated from the ADNI-GO, ADNI-2, and ADNI-3 study phases; of these, 388 participants were

Table 1
Demographic Information. Full description of all cohort demographics at baseline. Total number of participants (n), sex (by percentage female), age, and education in years are reported for each cohort in full (All), as well as for CN, MCI, and AD diagnostic subdivisions. The CN Template Cohort was comprised of only CN individuals. Values for age and education are summarized in the form mean (±standard deviation), with age ranges also given in the form min–max.

Dataset	Clinical Measure	Group			
		CN	MCI	AD	All
ROI Selection	n	50	50	50	150
	% Female	62	48	42	51
	Age	68.6 (±6.5) 55.1–88.7	71.9 (±8.6) 55.2–88.7	74.9 (±7.3) 55.3–89.7	71.8 (±7.9) 55.1–89.9
	Education	17.2 (±1.9)	16.2 (±2.3)	15.7 (±2.4)	16.4 (±2.3)
Template	n	152			
	% Female	50			
	Age	71.7 (±5.9) 55.6–85.3			
	Education	16.9 (±2.2)			
Experimental	n	286	514	129	929
	% Female	64	45	43	50
	Age	72.5 (±6.3) 56.3–90.2	72.1 (±7.5) 55.1–91.5	74.7 (±8.3) 55.7–90.4	72.6 (±7.3) 55.1–91.5
	Education	16.5 (±2.5)	16.2 (±2.6)	15.7 (±2.6)	16.2 (±2.6)

discarded due to incomplete data used to calculate or validate the biomarker, including diagnosis (n = 42), sex (n = 45), age (n = 8), scanner manufacturer (n = 3), and at least one brain volume (n = 298). Acquisition methods are detailed by Chow et al. (Chow et al., 2015). The remaining cohort of 1,231 individuals was comprised of 488 CN, 564 MCI, and 179 mild AD. Participants were diagnosed at baseline and reassessed at each study visit (Petersen et al., 2010). The study sample was subdivided into the following three cohorts: region of interest (ROI) selection, cognitively normal template creation, and experimental (see Table 1 for a full description of all cohort demographics).

2.1.2. Region of interest selection cohort

To determine the regional volumes that were associated with cognitive impairment, a sample of 150 age- and sex-matched individuals from the overall participant pool were pseudo-randomly selected using the numpy random module (Harris et al., 2020) with equivalent proportions of each diagnostic category (see Table 1 for a full description of ROI selection cohort demographics).

2.1.3. Cognitively normal template cohort

A sample of cognitively normal, age- and sex-matched individuals were randomly selected from the remaining participant pool to generate a CN template. The purpose of this template was to represent the average, healthy older adult brain (see Table 1 for a full description of CN template demographics). The sample size for this cohort was set to 152 individuals, the same number used to construct the widely adopted MNI152 template (Mazziotta et al., 1995; Mazziotta et al., 1995).

2.1.4. Experimental cohort

The remaining 929 participants were grouped into an experimental cohort for sMRI metric extraction and testing (see Table 1 for a full description of experimental cohort demographics). Baseline analyses were conducted for all 929 participants of the experimental cohort.

Longitudinal analyses were conducted on all participants who had follow-up data available. These participants were classified as those who worsened in diagnosis (denoted as $Diagnosis_{decline}$) and those who did not change in diagnosis or, in rare cases, improved (denoted as $Diagnosis_{stable}$), based on their diagnosis at 12, 24, 36 and 48-month follow-up sessions. $Diagnosis_{decline}$ is thus composed of those who declined from CN to MCI or CN to AD ($CN_{decline}$) and from MCI to AD ($MCI_{decline}$); $Diagnosis_{stable}$ is the group who either did not change in diagnosis or improved (composed of MCI_{stable}, CN_{stable}, and AD_{stable}). There was no later stage diagnostic category than AD in this sample, and therefore all AD participants were AD_{stable}, with no $AD_{decline}$ group (see Fig. 1 for a flowchart detailing the construction of longitudinal cohorts).

2.2. Neurocognitive assessments

Participant scores on the Clinical Dementia Rating Scale Sum of Boxes (CDR-SB) (Morris, 1993; Hughes et al., 1982; O'Bryant et al., 2008; Balsis et al., 2015), Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-11) (Grochowalski et al., 2016), and Mini-Mental.

State Exam (MMSE) (Folstein et al., 1975) were collated for all participants in the experimental cohort who had scores available at each time point.

2.3. AD-NeuroScore

Cortical reconstruction and volumetric segmentation of the 3 T, accelerated, T1-weighted MR images was performed with version 7.1 of the Freesurfer image analysis suite, which is documented and freely available for download online (https://surfer.nmr.mgh.harvard.edu/). The details of these procedures are described in previous publications (Dale et al., 1999; Reuter et al., 2010; Segonne et al., 2007; Fischl et al., 1999; Ségonne et al., 2004; Jovicich et al., 2006; Fischl et al., 1999; Fischl et al., 2002; Fischl et al., 2001; Fischl et al., 2004; Fischl et al., 2004; Desikan et al., 2006). With these methods, eighty-four cortical and subcortical regional volumes were estimated at baseline for each subject. Additionally, results were reviewed using the ENIGMA structural imaging quality control protocols (https://enigma.usc.edu/) (Stein et al., 2012).

2.3.1. ROI selection

To determine the brain regions most sensitive to diagnosis at baseline, we first performed an Analysis of Variance (ANOVA) in our ROI selection cohort to test for an effect of diagnosis for each of the 84 regions (using a Bonferroni-corrected alpha of 0.05/84; see Supplementary File 2 for a complete list of the 84 cortical and subcortical regions tested). Forty-one regions were found to be significant. All these regions were consistent with the existing literature (Yin et al., 2013; Zanchi et al., 2017; Harper et al., 2017) and were used to compute AD-NeuroScore (Table 2).

2.3.2. Cognitively normal template

To generate a CN template vector representative of the healthy average older adult brain, each volume that was determined from the ROI selection process was averaged across all CN template cohort participants (see Table 2). Data harmonization procedures were applied to each region in the CN template vector (see Methods Section 2.3.3. Data Harmonization). This CN template vector had the same dimensions as vectors extracted from experimental cohort participants. These two vectors were used to compute the distance metrics evaluated in this work (see Methods Section 2.3.4. Calculating the Z-Weighted Euclidean Distance).

2.3.3. Data harmonization

To account for interindividual variations in head size, age, sex, and MRI acquisition features like scanner model and manufacturer, we adjusted ROI volumes using a w-score, a data harmonization approach

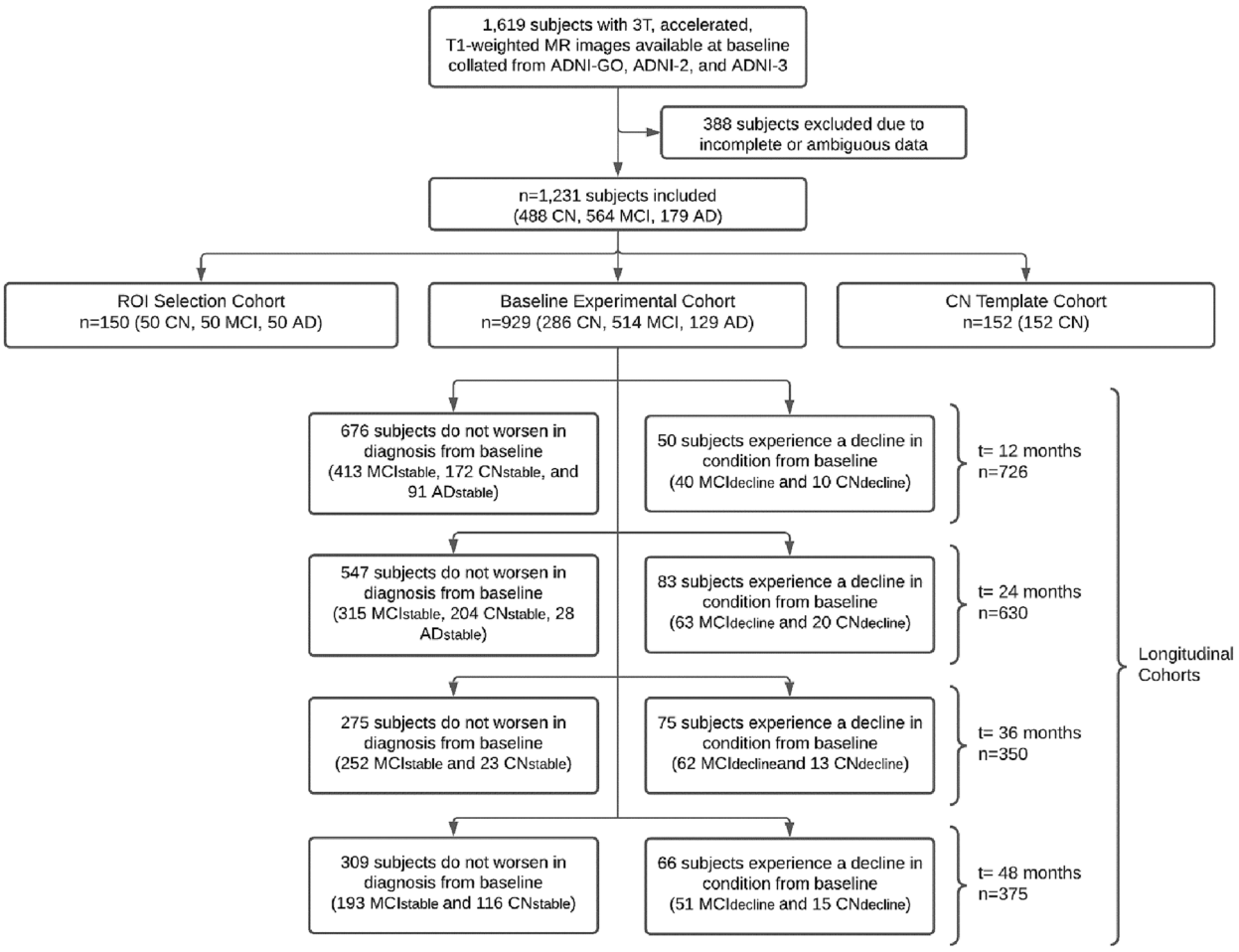

Fig. 1. Flowchart visualizing baseline allocation of participants into the 3 experimental cohorts (ROI Selection, CN Template, and Experimental) and longitudinal construction of the stable and decline groups at 12, 24, 36, and 48-months. Patients whose diagnosis progressed relative to baseline were classified into a Diagnosis (decline) group, while those who did not worsen were categorized as Diagnosis (stable) for each respective time point.

previously validated on sMRI data (Popuri et al., 2020; Ma et al., 2019). This approach uses a generalized linear model (GLM) framework, where the structural volume for a given region of each participant is modeled as the linear combination of all covariates as shown in Eq. (1):

$$V_i^r = \beta_o^r + \sum_c^{NC} \beta_c^r x_{c,i} + \varepsilon_i^r \quad (1)$$

Here, the volume V_i^r of region r for subject i, is modeled as the sum over NC covariates (age, sex, scanner model, and total ICV) of coefficient β_c^r, for covariate c, multiplied by the value of the covariate, $x_{c,i}$, plus the residual term, ε_i^r.

To compute the w-score, w_i^r, also known as the standardized residual, we took the z-transformation of the residual term from Eq. (1), ε_i^r, as shown in Eq. (2):

$$w_i^r = \frac{(\varepsilon_i^r - \mu_\varepsilon^r)}{\sigma_\varepsilon^r} \quad (2)$$

where μ_ε^r and σ_ε^r represent the mean and standard deviation, respectively, of the residual term for a given subject's regional brain volume.

2.3.4. Calculating the Z-Weighted Euclidean distance

To compute the difference between each experimental cohort participant and the CN template, we employed a novel modified Euclidean inspired distance function which we call the Z-weighted Euclidean distance (ZWE). A traditional Euclidean distance metric is calculated by treating the list of significant, harmonized regional volumes as a vector in n-dimensional space, where n is the number of regions, and computing the Euclidean distance between each participant and the CN template vectors. The ZWE distance function differed in that each region was multiplied by a weight resulting from each region's level of significance (z-score) as determined during ROI selection process as shown in Eq. (3):

$$ZWE = \left(\sum_{i=1}^{NR} \left[\left(w_i^T - w_i^S\right) z_i\right]^2\right)^{1/2} \quad (3)$$

where z_i refers to the average of the z-scores associated with the p-values across the three possible pairwise comparisons (CN vs MCI, CN vs AD, and MCI vs AD) for a given region, i, and NR denotes the total number of regions. The ZWE distance was computed between the harmonized residual of each subject, w_i^S, and the CN template, w_i^T. Thus, the more significant a region was in ROI-selection, the greater its contribution to the computed AD-NeuroScore. The concept of weighing multivariate sums by measures of significance has been used in a number of applications since its inception in the proposal of Hotelling's T^2 distance in 1947 (Hotelling et al., 1947; Gutman et al., 2013; Hua et al., 2013). The significant ROIs listed in Table 2 are visualized in Fig. 2, with a color scale indicating respective z_i weighting.

We investigated several other methods for computing the difference between each experimental cohort subject and the CN template. Each of these algorithms similarly involved computing a distance function between like objects constructed from the harmonized, significant regional brain volumes, which we determined in the ROI selection step. Generally, the distance functions fell into one of two categories (curves and

Table 2
Significant ROIs by Z-Score Ranking and CN Template Values. Resulting 41 significant regions of interest (ROIs) extracted by performing ANOVA for each of the 84 regions in the ROI selection cohort are reported in the above table along with corresponding z scores. Significance was established based on an alpha = 0.05, Bonferroni corrected. Structures are identified by the FreeSurfer version 7.1 ROI labels. Mean volumes and standard deviations (SD) of the CN template cohort are included for each respective region.

FreeSurfer ROI Label	Z	CN Template Mean Vol (±SD) (mm^3)
Left-Hippocampus	9.12	0.48 (±0.74)
Left-Amygdala	8.43	0.60 (±0.79)
Right-Amygdala	8.37	0.54 (±0.80)
Right-Hippocampus	7.82	0.45 (±0.71)
lh_middletemporal_volume	6.53	0.40 (±0.88)
lh_fusiform_volume	6.20	0.27 (±0.82)
lh_inferiorparietal_volume	6.11	0.33 (±0.90)
rh_middletemporal_volume	6.12	0.42 (±0.92)
lh_inferiortemporal_volume	5.88	0.36 (±0.85)
lh_precuneus_volume	5.81	0.26 (±0.91)
lh_rostralmiddlefrontal_volume	5.69	0.16 (±1.06)
rh_inferiortemporal_volume	5.54	0.27 (±0.87)
rh_precuneus_volume	5.50	0.34 (±0.93)
rh_inferiorparietal_volume	5.48	0.33 (±0.88)
rh_superiortemporal_volume	5.44	0.35 (±0.94)
rh_entorhinal_volume	5.39	0.12 (±0.77)
Left-Thalamus	5.17	0.33 (±1.01)
rh_lateralorbitofrontal_volume	5.02	0.14 (±0.91)
lh_superiortemporal_volume	5.01	0.28 (±0.87)
lh_insula_volume	4.86	0.19 (±0.86)
lh_bankssts_volume	4.75	0.12 (±0.93)
lh_lateralorbitofrontal_volume	4.67	0.26 (±0.89)
lh_isthmuscingulate_volume	4.66	0.20 (±1.04)
rh_parahippocampal_volume	4.64	0.21 (±0.91)
Right-Accumbens-area	4.58	0.33 (±0.88)
lh_entorhinal_volume	4.56	0.23 (±0.74)
rh_rostralmiddlefrontal_volume	4.50	0.18 (±1.04)
lh_superiorparietal_volume	4.44	0.14 (±0.87)
Right-Thalamus	4.42	0.31 (±1.02)
lh_superiorfrontal_volume	4.38	0.17 (±0.89)
rh_fusiform_volume	4.25	0.23 (±0.88)
rh_insula_volume	4.14	0.21 (±0.93)
lh_lateraloccipital_volume	4.08	0.31 (±0.87)
rh_superiorfrontal_volume	3.96	0.16 (±0.99)
lh_posteriorcingulate_volume	3.86	0.29 (±0.98)
rh_isthmuscingulate_volume	3.78	0.13 (±0.87)
rh_superiorparietal_volume	3.75	0.26 (±1.00)
rh_parsorbitalis_volume	3.70	0.19 (±0.92)
rh_posteriorcingulate_volume	3.60	0.24 (±0.90)
Left-Accumbens-area	3.60	0.45 (±1.04)
rh_lateraloccipital_volume	3.47	0.24 (±0.83)

points). For the distance functions that computed distances between curves, in addition to projecting the curve in one dimensional space, the k-false nearest neighbors method was implemented to embed each curve in n-dimensions (see Supplementary File 3 for further details). The functions which computed a distance between curves included the Fréchet distance and the Hausdorff distance. The Fréchet Distance is one method used to quantify the similarity between two curves or sets of points in space, with an emphasis on the location and ordering of points (Dumitrescu and Rote, 2004). Similarly, the Hausdorff distance measures how far two subspaces of a metric space are from each other; however, it does not account for the flow or order of points (Maiseli, 2021). The vector-based distance functions included the Euclidean and ZWE distance. Further details along with equations modeling these mathematical distances are included in Supplementary File S3.

Statistical analyses tested the performance of each distance metric based on sensitivity to diagnosis, disease severity, and progression (See Table S1–S3 in Supplementary Materials). All metrics evaluated were benchmarked using AHV, an NIA-AA diagnostic biomarker for Alzheimer's disease (Jack et al., 2018) (Supplemental Table S1). The optimal biomarker, the ZWE distance, was selected as our "AD-NeuroScore" (See Section 2.4 Validation Procedures).

2.4. Validation procedures

2.4.1. Baseline validation procedures

At baseline, we evaluated sensitivity to diagnosis of AD-NeuroScore by calculating the area under the Receiver Operating Characteristic curve (AUC-ROC) and the associated 95% confidence intervals, using a logistic regression model for each pairwise comparison of diagnostic groups (CN vs MCI, MCI vs AD, and CN vs AD). Specificity, sensitivity, and accuracy were evaluated at the optimal false positive rate of 0.145 determined by a Youden index analysis. These results were then converted to minimum sample size estimation for statistical power using a two-sided test and standard significance level of 0.05. Sensitivity to diagnosis was further assessed using pairwise, two-tailed t-tests for each possible diagnostic group comparison, using a Holm-Bonferroni-corrected alpha of 0.05. Results were then converted to z-scores and Cohen's *d* was calculated to indicate effect size.

Baseline associations of AD-NeuroScore and AHV with disease severity, operationalized as MMSE, ADAS-11, and CDR-SB scores, were tested using linear regression, both in the overall baseline experimental cohort and in each diagnostic sub-group. In addition to examining the significance of the slope using a Holm-Bonferroni-corrected alpha of 0.05, we also estimated the correlation coefficients and compared the performance of AD-NeuroScore and AHV by conducting a Fisher's z-test of the z-transformed correlation coefficients, using the Holm-Bonferroni method to adjust for the 3 comparisons at baseline.

2.4.2. Longitudinal validation procedures

To determine if AD-NeuroScore might be predictive of disease progression, we assessed the relationship between AD-NeuroScore at baseline and both the change in diagnosis and change in disease severity at 12, 24, 36, and 48 months. Logistic regression was used to examine whether baseline AD-NeuroScore was predictive of change in diagnosis ($Diagnosis_{decline}$ vs. $Diagnosis_{stable}$); AUC-ROC (and associated 95% CI) was used as a metric of the predictive ability. We also compared the baseline distribution of AD-NeuroScore between $Diagnosis_{decline}$ and $Diagnosis_{stable}$ groups using pairwise, two-tailed t-tests in the full experimental cohort and subsequently further stratified by baseline diagnosis (MCI or CN) to investigate if the ability to predict decline is driven by a specific patient population. Similar to the baseline validation procedures, the results for all comparisons were converted to z-scores, and Cohen's *d* was calculated to assess effect size.

Longitudinal association with disease severity was tested using linear regression between baseline metric scores and the change in the neuropsychological assessment scores (MMSE, ADAS-11, and CDR-SB) from baseline, at each respective longitudinal session. Significance was assessed using a Holm-Bonferroni-corrected alpha of 0.05. All absent longitudinal comparisons were excluded due to insufficient sample size.

2.5. Validation using clinically implemented ROI

To validate AD-NeuroScore in the context of the alternative imaging analysis tools frequently used by and accessible to clinicians, a vector of Neuroreader®-analogous ROI volum was approximated by transforming the FreeSurfer results into a less granular vector of regions closely matching the Neuroreader® atlas. 80 out of the 84 brain regions estimated by FreeSurfer were systematically mapped to create the 22 ROI structures in Neuroreader®. However, four regions (left and right nucleus accumbens and insula) estimated by FreeSurfer did not have corresponding ROIs in Neuroreader®, and therefore, they were excluded from this specific analysis. The calculation and validation methods utilized for computing and validating the FreeSurfer-based AD-NeuroScore were applied to the pseudo-Neuroreader® ROI vector. (Supplementary Table S4–S7).

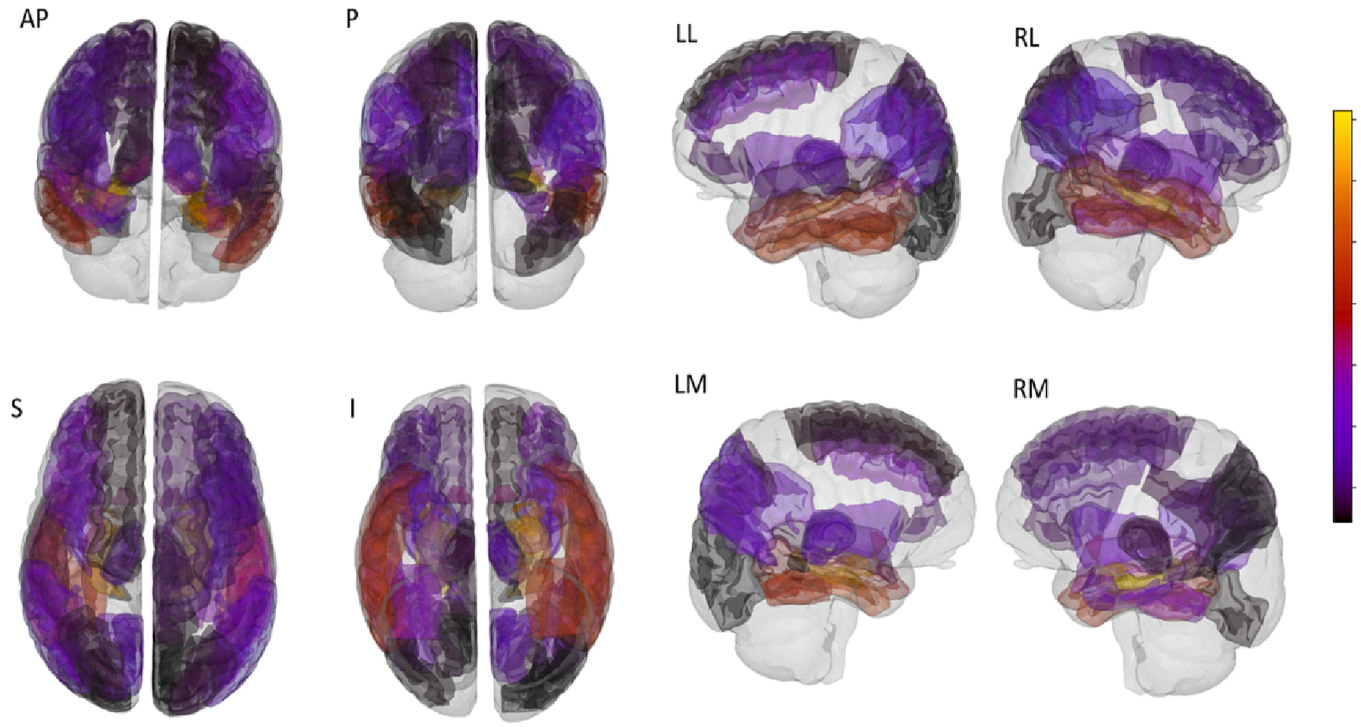

Fig. 2. The 41 significant regions of interest (ROIs) extracted by performing ANOVA for each of the 84 regions in the ROI selection cohort are visualized above on the Allen 500-μm Human Brain Atlas. The color scale depicts the z-score-based weighting of each ROI in AD-NeuroScore (ADNS), as described in Section 2.3.1. The Brainrender python library (https://github.com/brainglobe/brainrender) was used to create this figure. The following letters denote anatomical orientation: Anterior (A), Posterior (P), Superior (S), Inferior (I), Left Lateral Surface (LL), Left Medial Surface (LM), Right Lateral Surface (RL), Right Medial Surface (RM).

Table 3
Baseline Results for AD-NeuroScore Sensitivity to Diagnosis. Sensitivity to diagnosis assessed using pairwise, two-tailed t-tests performed for each possible diagnostic group comparison. Resulting z-scores, effect sizes (Cohen's *d*) with 95% confidence intervals (CI), AUC-ROC values with 95% CI, and minimum sample size estimation for statistical power are included. Results using AHV are included for benchmarking. Significant results from group comparisons are denoted by * to indicate $p < 0.05$, ** to indicate $p < 0.01$, and *** to indicate $p < 0.001$, Holm-Bonferroni corrected.

Group Comparison	Metric	AUC [95% CI]	Cohen's *d* [95% CI]	z-score	Min. Sample Estimate
CN vs MCI	AD-NeuroScore	0.66 [0.59, 0.73]	−0.56 [−0.59, −0.53]***	−7.46	110
	AHV	0.65 [0.58, 0.72]	0.55 [0.52, 0.58]***	−7.27	126
CN vs AD	AD-NeuroScore	0.91 [0.86, 0.96]	−2.06 [−2.24, −1.88]***	−16.37	14
	AHV	0.88 [0.81, 0.95]	1.72 [1.55, 1.89]***	−14.27	16
MCI vs AD	AD-NeuroScore	0.80 [0.71, 0.89]	−1.16 [−1.19, −1.14]***	−11.23	38
	AHV	0.76 [0.67, 0.84]	0.92 [0.90, 0.94]***	−9.04	52

2.6. Data availability

Full, open access to all de-identified ADNI imaging and clinical data is publicly and freely available to individuals who register with the ADNI and agree to the conditions in the "ADNI Data Use Agreement," upon approval of a request that includes the proposed analysis and the named lead investigator (contact via https://adni.loni.usc.edu/data-samples/access-data/). Additional details about the ADNI data acquisition and sharing policies can be found at https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf.

The code supporting the findings of this study is openly available in [repository name: "AD-NeuroScore"] at https://github.com/jbramen/AD-NeuroScore. The Python programming language (version 3.9) was used for all analyses (Python Software Foundation, https://www.python.org/).

3. Results

3.1. Baseline validation

3.1.1. Baseline sensitivity to diagnosis

AD-NeuroScore was significantly associated with diagnosis at baseline ($p < 0.001$ for all comparisons, Holm-Bonferroni corrected; Table 3). AD-NeuroScore performed best at distinguishing AD from other groups (CN and MCI) and least well at distinguishing CN and MCI participants. AD-NeuroScore performed as well as our benchmark, AHV ($p < 0.001$ for all comparisons, Holm-Bonferroni corrected), in this cross-sectional validation (Table 3). AUC-ROC values of AD-NeuroScore and AHV were similar across all group comparisons. Visual inspection of the overlaid AD-NeuroScore and AHV AUC-ROC curves for AD comparisons (CN vs AD and MCI vs AD) indicated that at low false positive rates, the AD-NeuroScore true positive rate tended to be higher (Fig. 3). Examining the distribution of AD-NeuroScore and AHV for each diagnostic group at baseline (Fig. 4), AD-NeuroScore qualitatively

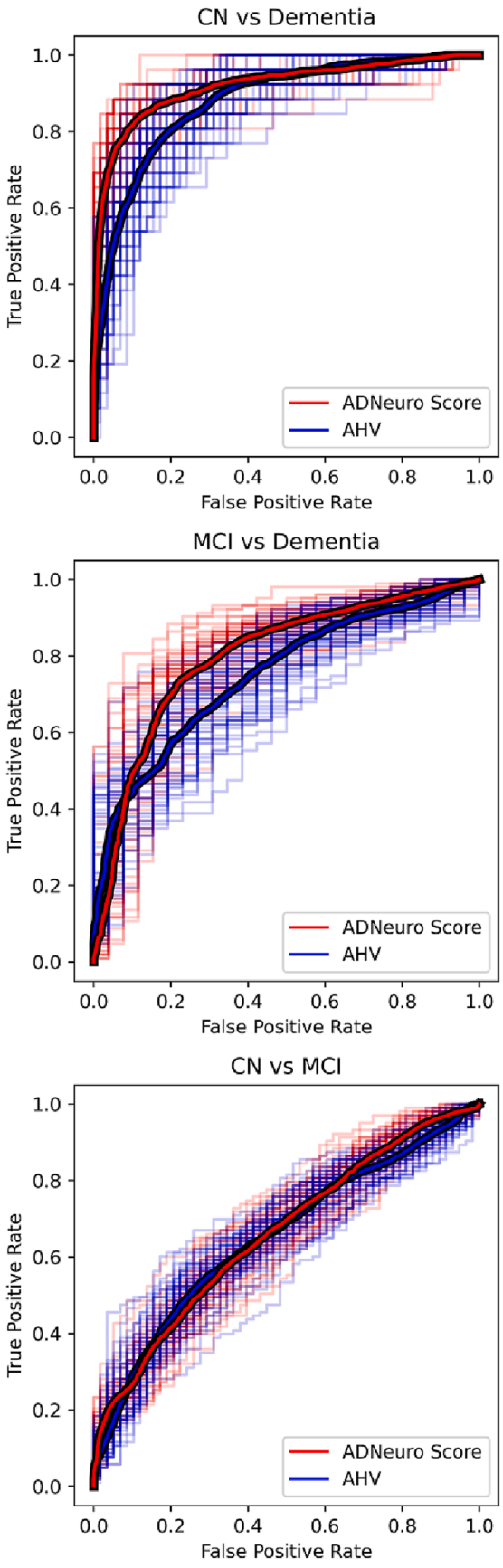

Fig. 3. Overlaid AUC-ROC curves visualizing baseline classification performance of AD-NeuroScore and AHV across the 3 diagnostic group comparisons.

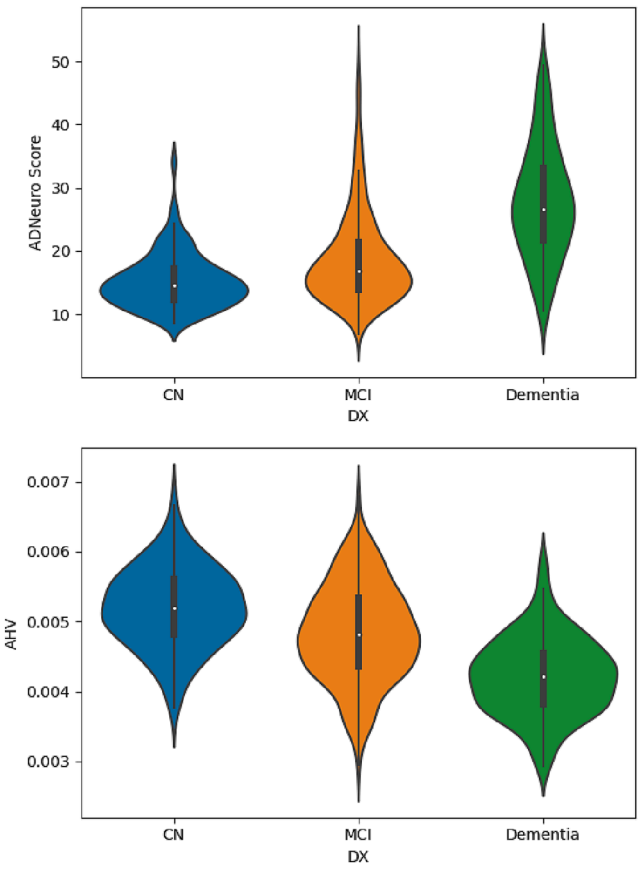

Fig. 4. Violin plots depicting the distribution of AD-NeuroScore and AHV for each diagnostic group at baseline.

demonstrated greater separation between diagnostic category medians, with more centrally concentrated distributions and longer tails in CN and MCI groups at the edges of the distribution.

3.1.2. Baseline association with disease severity (MMSE, ADAS-11, and CDR-SB)

In the overall baseline experimental cohort, we found that AD-NeuroScore was significantly associated with disease severity, as measured by MMSE, ADAS-11, and CDR-SB scores (Table 4; Fig. 5; all p-values < 0.001, Holm-Bonferroni corrected). In sub-analyses stratified by baseline diagnosis, we found significant associations between AD-NeuroScore and metrics of disease severity in participants with MCI (MMSE, ADAS-11, and CDR-SB) and AD (ADAS-11, and CDR-SB). Conversely, we found no significant association with disease severity in CN individuals.

AD-NeuroScore generally performed as well or better than our benchmark, AHV, in this cross-sectional validation using the overall baseline experimental cohort (Table 4; Supplementary Fig. S1). Results from z-tests conducted using Fisher z-transformed correlation coefficients revealed that associations between both CDR-SB and ADAS-11 and AD-NeuroScore were significantly stronger than with AHV ($p = 0.006$ for CDR-SB and $p = 0.024$ for ADAS-11). Sub-analyses stratified by baseline diagnosis revealed that AD-NeuroScore and AHV performed the most similarly in participants with MCI, and the most differently in participants with AD (AD-NeuroScore: $p < 0.01$ for ADAS-11 and $p < 0.05$ for CDR-SB, both Holm-Bonferroni corrected, AHV: NS for all).

3.2. Longitudinal validation

3.2.1. Metric sensitivity to progression

In the overall baseline experimental cohort, we found that baseline

Table 4
Relationship Between AD-NeuroScore and Disease Severity. Baseline rows include results from cross-sectional analysis of AD-NeuroScore and disease severity, operationalized as MMSE, ADAS-11, and CDR-SB scores. All other rows include results from longitudinal analysis of baseline AD-NeuroScore and change in disease severity scores at 12, 24, 36, and 48 months. Testing was performed with linear regression in the overall experimental cohort and in each diagnostic sub-group. Performance of AHV is included for benchmarking. Significant associations are indicated by * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$, Holm-Bonferroni corrected.

Class	Time	Metric	r			n		
			CDR-SB	MMSE	ADAS-11	CDR-SB	MMSE	ADAS-11
All	Baseline	AD-NeuroScore	0.52***	−0.34***	0.53***	926	926	926
		AHV	−0.43***	0.28***	−0.47***	926	926	926
	12	AD-NeuroScore	−0.27***	0.21***	−0.23***	525	697	721
		AHV	0.23***	−0.16***	0.19***	525	697	721
	24	AD-NeuroScore	−0.40***	0.27***	−0.39***	402	614	626
		AHV	0.32***	−0.19***	0.26***	402	614	626
	36	AD-NeuroScore	−0.38***	0.21***	−0.31***	285	332	345
		AHV	0.34***	−0.17**	0.26***	285	332	345
	48	AD-NeuroScore	−0.42***	0.26***	−0.31***	205	369	375
		AHV	0.35***	−0.19***	0.26***	205	369	375
CN	Baseline	AD-NeuroScore	−0.08	−0.01	0.10	286	286	286
		AHV	0.00	−0.01	−0.17	286	286	286
	12	AD-NeuroScore	−0.81*	−0.02	0.09	13	183	180
		AHV	0.53	−0.03	0.01	13	183	180
	24	AD-NeuroScore	−0.69*	0.02	−0.09	17	220	221
		AHV	0.45	−0.08	0.05	17	220	221
	36	AD-NeuroScore	–	−0.06	0.01	1	44	36
		AHV	–	−0.19	0.03	1	44	36
	48	AD-NeuroScore	−0.12	0.00	0.01	5	126	133
		AHV	0.52	−0.21	−0.01	5	126	133
MCI	Baseline	AD-NeuroScore	0.23***	−0.09*	0.33***	511	511	511
		AHV	−0.22***	0.11*	−0.35***	511	511	511
	12	AD-NeuroScore	−0.30***	0.21***	−0.19***	418	423	452
		AHV	0.23***	−0.13*	0.17**	418	423	452
	24	AD-NeuroScore	−0.39***	0.28***	−0.34***	357	357	377
		AHV	0.32***	−0.15*	0.25***	357	357	377
	36	AD-NeuroScore	−0.38***	0.24***	−0.31***	284	273	309
		AHV	0.35***	−0.16*	0.27***	284	273	309
	48	AD-NeuroScore	−0.42***	0.30***	−0.31***	199	226	242
		AHV	0.35***	−0.19*	0.28***	199	226	242
AD	Baseline	AD-NeuroScore	0.27*	−0.17	0.33**	129	129	129
		AHV	−0.08	0.08	−0.14	129	129	129
	12	AD-NeuroScore	0.09	0.04	0.00	94	91	89
		AHV	−0.08	−0.03	−0.22	94	91	89
	24	AD-NeuroScore	0.03	0.16	−0.36	28	37	28
		AHV	−0.34	−0.10	−0.20	28	37	28
	36	AD-NeuroScore	–	−0.13	–	0	15	0
		AHV	–	−0.09	–	0	15	0
	48	AD-NeuroScore	–	0.28	–	1	17	0
		AHV	–	−0.11	–	1	17	0

AD-NeuroScore differentiated $Diagnosis_{stable}$ from $Diagnosis_{decline}$ significantly at all timepoints (12, 24, 36, and 48-months; Table 5; $p < 0.001$, Holm-Bonferroni corrected, for all). Sub-analyses stratified by baseline diagnosis revealed that AD-NeuroScore's ability to predict decline was primarily driven by individuals with MCI at baseline ($p < 0.001$, Holm-Bonferroni corrected, for all). We found no significant effects in CN individuals.

We found that AD-NeuroScore performed as well as AHV in the overall experimental cohort (AHV: $p < 0.01$ at 12-months, $p < 0.001$ at all successive sessions, Holm-Bonferroni corrected). While there were no significant differences between AD-NeuroScore and AHV, qualitatively, AD-NeuroScore tended to do modestly better than AHV at differentiating $Diagnosis_{stable,}$ from $Diagnosis_{decline}$ participants at 12- and 24-months, and slightly worse at 36- and 48-months. Sub-analysis by baseline diagnosis revealed similar patterns; AD-NeuroScore performed equivalently to AHV in individuals with MCI. However, AHV demonstrated somewhat higher sensitivity to change in diagnosis in the CN group.

3.2.2. Longitudinal association with disease severity (MMSE, ADAS-11, and CDR-SB)

In the overall longitudinal cohort, we found that baseline AD-NeuroScore was also significantly associated with the change in disease severity scores (Table 4; $p < 0.001$, Holm-Bonferroni corrected, for all), as measured by the change in MMSE, ADAS-11, and CDR-SB scores from baseline to each time point (12, 24, 36, and 48-months). Sub-analysis by baseline diagnosis revealed that AD-NeuroScore's association with the change in disease severity scores was also primarily driven by individuals with MCI at baseline. We found significant associations between AD-NeuroScore and the change in MMSE, ADAS-11, and CDR-SB scores in participants with MCI at all time points ($p < 0.001$, Holm-Bonferroni corrected for all). Conversely, we found no significant association with the change in disease severity in AD individuals, and only two significant associations in CN individuals (CDR-SB at 12- and 24-months; $p < 0.05$, Holm-Bonferroni corrected, for both).

AD-NeuroScore generally performed as well or better than our benchmark, AHV, in this longitudinal validation (Table 4). As with AD-NeuroScore, correlations between AHV at baseline and change in disease

Fig. 5. Compares the features that are most important for the widespread clinical and investigational implementation of the algorithms discussed in this study. * Indicates differences in the tests used to compare metrics. See Table 6 for more information on performance of previous metrics relative to AD-NeuroScore.

severity scores were significant in the overall longitudinal cohort (Table 4; $p < 0.001$, Holm-Bonferroni corrected, for all timepoints). Results from z-tests conducted using Fisher z-transformed correlation coefficients revealed that correlations between AD-NeuroScore and 24-month change in ADAS-11 were significantly stronger than with AHV ($p = 0.003$, Holm-Bonferroni corrected). In the MCI group, AD-NeuroScore performed as well or somewhat better than AHV (AD-NeuroScore: $p < 0.001$ for all; AHV: $p < 0.05$ for MMSE across all time points, $p < 0.01$ for ADAS-11 at 12-months, and $p < 0.001$ for all other scores and time points, all Holm-Bonferroni corrected).

In the AD group, neither AD-NeuroScore nor AHV were significantly associated with change in disease severity scores. The AD sub-analysis was limited by sample size longitudinally. In the CN group, both AD-NeuroScore and AHV had generally weak correlations, none of which were significant aside from AD-NeuroScore and CDR-SB at 12- and 24-months (both $p < 0.05$, Holm-Bonferroni corrected).

3.3. Comparison of similar existing metrics

We also compared performance of AD-NeuroScore against other benchmarks, and we present the results in Table 6. This table presents an overview of the methods, features, comparisons, and performance of several sMRI-based benchmarks in addition to AD-NeuroScore. This comparative analysis demonstrates that in general, AD-NeuroScore performs as well as these other biomarkers.

3.4. Alternative atlas analysis

3.4.1. Baseline and longitudinal validation

Results from analyses repeated using the Neuroreader® ROI-based AD-NeuroScore were similar to those obtained with the Freesurfer-based AD-NeuroScore, including all cross-sectional and longitudinal analyses. Benchmarking the pseudo-Neuroreader® ROI-based AD-NeuroScore against AHV also produced results analogous to those of the original analysis (Supplementary Table S5–S7).

4. Discussion

We developed AD-NeuroScore, a Euclidean inspired sMRI-based distance metric, which demonstrated significant associations with diagnosis (CN, MCI, and AD) and disease severity (MMSE, ADAS-11, and CDR-SB scores) at baseline, and performed as well as or better than AHV. AD-NeuroScore was also significantly associated with changes in all three disease severity scores over a 48-month follow-up period, which was largely driven by the MCI group due to the limited sample size of CN and AD participants with available longitudinal data. AD-NeuroScore and AHV were both able to differentiate between participants who declined in cognitive status and those who remained stable in longitudinal analyses. Qualitatively, AD-NeuroScore performed slightly better than AHV at earlier time points (12- and 24-months) and slightly worse at later sessions (36- and 48-months).

ROIs included in AD-NeuroScore based on ROI selection were generally consistent with the literature (Gómez-Isla et al., 1996; Rabinovici et al., 2008; Thompson et al., 2003; Braak et al., 1997; Frisoni et al., 1999; Thompson et al., 2007; Laakso et al., 1996; Laakso et al., 1998; Dickerson et al., 2001; Vercelletto et al., 2002; Jacobs et al., 2012; Nie et al., 2017). However, we expected but did not find lateral ventricular volumes to be significant during ROI selection, even though ventricular dilation is frequently observed in patients with AD (Attier-Zmudka et al., ; Ott et al., 2010; Nestor et al., 2008). We did find a large overlap in total ventricular volumes of patients and controls, which is consistent with previous cross-sectional volumetric studies (Nestor et al., 2008; Giesel et al., 2006; Schott et al., 2005). Therefore, absolute ventricular change may be a more sensitive measure of AD-pathology than total ventricular volume.

In this study, we also evaluated the performance and approach of AD-NeuroScore in comparison to several similar metrics. Our findings indicate that AD-NeuroScore outperforms or is at least comparable to these scores. We compared AD-NeuroScore with SuStaIn (Young et al., 2021), UMI (Wang et al., 2021), MRDATS (Popuri et al., 2020), aSTAND-score (Vemuri et al., 2008; Vemuri et al., 2008), AD-PS (Casanova et al., 2018), and the SPARE-AD index (Davatzikos et al., 2009). SuStaIn (Young et al., 20142014; Young et al., 2021) utilizes multi-modal disease data from several data sources and acquisition

Table 5
Relationship Between AD-NeuroScore and Longitudinal Diagnosis Transitions. Sensitivity to diagnostic transition category assessed using pairwise, two-tailed t-tests performed between groups with a stable or declining diagnosis from baseline, at each respective time point. Resulting z-scores, effect sizes (Cohen's *d*) with 95% confidence intervals (CI)s, and AUC-ROC values with 95% CIs are included. Performance was evaluated both in the overall experimental cohort and in sub-groups based on starting diagnosis (CN or MCI). Results using AHV are included for benchmarking. Significant results from group comparisons are denoted by * to indicate $p < 0.05$, ** to indicate $p < 0.01$, and *** to indicate $p < 0.001$, Holm-Bonferroni corrected.

Comparison	time	Metric	AUC [95% CI]	Cohen's *d* [95% CI]	z-score	
Stable vs. Decline	All	12 months	ADNS	0.67 [0.53, 0.80]	0.64 [0.35, 0.93]***	−4.35
			AHV	0.65 [0.50, 0.80]	−0.53 [−0.81, −0.24]**	−3.57
		24 months	ADNS	0.71 [0.59, 0.82]	0.84 [0.61, 1.08]***	−7.00
			AHV	0.71 [0.61, 0.82]	−0.79 [−1.02, −0.55]***	−6.58
		36 months	ADNS	0.69 [0.53, 0.85]	0.72 [0.46, 0.98]***	−5.42
			AHV	0.72 [0.60, 0.85]	−0.79 [−1.05, −0.52]***	−5.88
		48 months	ADNS	0.67 [0.52, 0.83]	0.74 [0.47, 1.01]***	−5.36
			AHV	0.72 [0.62, 0.83]	−0.79 [−1.06, −0.51]***	−5.66
	CN	12 months	ADNS	0.35 [0.08, 0.62]	−0.06 [−0.70, 0.59]	−0.17
			AHV	0.72 [0.35, 1.09]	−0.59 [−1.23, 0.06]	−1.79
		24 months	ADNS	0.64 [0.41, 0.88]	0.55 [0.08, 1.01]	−2.32
			AHV	0.74 [0.53, 0.95]	−0.79 [−1.26, −0.33]**	−3.34
		36 months	ADNS	0.59 [0.21, 0.97]	0.34 [−0.36, 1.05]	−0.98
			AHV	0.78 [0.46, 1.09]	−0.84 [−1.57, −0.11]	−2.30
		48 months	ADNS	0.44 [0.14, 0.75]	0.11 [−0.44, 0.65]	−0.39
			AHV	0.72 [0.46, 0.98]	−0.65 [−1.20, −0.10]	−2.34
	MCI	12 months	ADNS	0.77 [0.65, 0.90]	1.08 [0.75, 1.42]***	−6.39
			AHV	0.68 [0.51, 0.85]	−0.69 [−1.02, −0.36]***	−4.14
		24 months	ADNS	0.76 [0.64, 0.88]	1.05 [0.77, 1.34]***	−7.36
			AHV	0.73 [0.59, 0.86]	−0.85 [−1.12, −0.57]***	−5.99
		36 months	ADNS	0.73 [0.62, 0.84]	0.87 [0.59, 1.16]***	−5.99
			AHV	0.73 [0.56, 0.90]	−0.86 [−1.14, −0.57]***	−5.87
		48 months	ADNS	0.73 [0.57, 0.89]	0.78 [0.46, 1.10]***	−4.83
			AHV	0.69 [0.51, 0.88]	−0.76 [−1.08, −0.44]***	−4.71

methods with an event-based model to subtype individuals based on their likelihood to stabilize or decline (Fonteijn et al., 2012). UMI is derived using a low-rank and sparse subspace decomposition algorithm to extract common group structure and impose regularization constraints from morphological and connectivity features to predict cognitive status decline (Wang et al., 2021). Additionally, MRDATS uses w-corrected brain ROI volumes in an ensemble-learning algorithm to arrive at a score between 0 and 1 to indicate CN to dementia progression (Popuri et al., 2020). The aSTAND-score uses a support vector machine classifier which takes in sMRI input features such as GM, WM, and CSF tissue densities and demographic information to assign a numerical value to classify a patient as CN or AD (Vemuri et al., 2008). AD-PS uses the same sMRI input features (Casanova et al., 2013) and high dimensional regularized logistic regression to classify a patient as CN or AD (Casanova et al., 2018). The SPARE-AD index (Davatzikos et al., 2009) uses high-dimensional pattern classification of volumetric atrophy and is highly predictive of cognitive status, with its rate of change particularly sensitive in comparisons between CN individuals and patients with MCI. AD-NeuroScore shares the most similarity with RVI or the ENIGMA Dot Product (Kochunov et al., 2022), which also weighs a multivariate sum by a metric describing the effect of a group status with harmonized phenotypes. However, RVI uses several metrics derived from multiple scan types, including gray matter thicknesses and fractional anisotropy values. Furthermore, it averages the hemispheric volumes, which may lead to loss of interpretable clinical information about laterality.

Although these metrics accurately classify patients, they are not widely used in clinical settings likely due to various barriers to translation. These barriers include the use of features that are from multiple image modalities (Hwang and Park, 2020), are not readily available in clinics (Eweje et al., 2022), or cannot be inspected by a radiologist to assure quality (Larson and Boland, 2019). Moreover, many metrics lack interpretability at both the patient and model levels (Pinto et al., 2022). Finally, lack of harmonization in data collected on different scanners significantly hinders the interpretability of inter-site and longitudinal follow-up in many existing approaches. Fig. 5 provides a visual comparison of the features that affect the clinical translatability of AD-NeuroScore and similar metrics as well as their relative performance. SPARE-AD is the most accurate, but least translatable score. RVI is the most translatable of the previous metrics summarized, but AD-NeuroScore is more so and significantly outperforms it in disease stage separation.

AD-NeuroScore was developed with the aim of enhancing patient-level interpretability of sMRI reports, recognizing the importance for clinicians to fully understand the patient's condition to make a diagnosis and treatment plan. An ideal metric should provide an easily understandable score, accompanied by a report that breaks down its clinically relevant features. Fig. 6 demonstrates how AD-NeuroScore can be broken down, allowing for quick assessment of the harmonized atrophy pattern that contributes to the total score, making it a useful tool for clinicians. In addition to facilitating diagnosis, this could be helpful in

Table 6
Comparison Between AD-NeuroScore and Similar Metrics. Each relevant biomarker and its performance are reported as well as the method, input features used, and AD-NeuroScore benchmark for the performance metric; *** indicate $p < 0.001$ for the statistical comparison between the performance metric of the related biomarker and corresponding performance metric of AD-NeuroScore.[a, b and c]

Biomarker	Method	Features	Comp.	Benchmark Performance	ADNS Performance
aSTAND	Linear support vector machine (ML)	sMRI: GM, WM, and CSF density; demographics	CN vs AD	Accuracy: 89% Sensitivity: 88% Specificity: 90%	Accuracy: 85% Sensitivity: 83% Specificity: 86%
AD-PS	High dimensional regularized logistic regression (ML)	sMRI: GM, WM, and CSF density	CN vs AD	AUC = 0.89 [0.79, 0.99]	AUC = 0.91 [0.86, 0.96]
			CN vs MCI	AUC = 0.70 [0.61, 0.79]	AUC = 0.66 [0.59, 0.73]
			12 MO: MCI stable vs decline	$p < 0.001$	$P < 0.001$
SuStaIn	Unsupervised learning algorithm (ML)	sMRI: regional volumes; MMSE; demographics, CSF: Aβ1-42 concentration	12 MO: MCI stable vs decline	AUC = 0.76	AUC = 0.77 [0.65, 0.90]
UMI	Low-rank and sparse subspace decomposition algorithm	sMRI: morphology; connectivity	18/24 MO: MCI stable vs decline	AUC = 0.749 (18 months)	AUC = 0.76 (24 months)
MRDATS	Ensemble-learning algorithm (ML)	sMRI: regional volumes (GM, and CSF); demographics; scanner information	CN vs AD	AUC = 0.964[a]	AUC = 0.91 [0.86, 0.96]
			0–3 YR MCI stable vs decline	AUC = 0.75	AUC = 0.75[b]
SPARE-AD	High-dimensional pattern classification using a Support Vector Machine (ML)	sMRI: regional volumes (GM, WM, and CSF)	CN vs MCI[c]	AUC = 0.89	AUC = 0.66 [0.59, 0.73]
			CN vs AD	AUC = 0.98	AUC = 0.91 [0.86, 0.96]
			12 MO: MCI stable vs decline	$p < 0.001$	$P < 0.001$
RVI	Correlation of harmonized phenotypes' effect sizes with their absolute magnitude	sMRI: regional volumes, GM thicknesses; DTI: FA; demographics	CN vs AD	Cortical thickness model: Cohen's d = 1.28*** Subcortical GM volume model: Cohen's d = 0.81*** WM model: Cohen's d = 0.90***	Cohen's d = 2.06***

DTI = diffusion tensor imaging; ML = machine learning; VBM = voxel-based morphometry.
[a] Statistical comparison was only conducted in training set.
[b] 0-3 year AUC value was approximated by averaging AUC at 12, 24, and 36 months.
[c] Based on rate of change of SPARE-AD.

targeting treatments.

Apart from its clinical utility, AD-NeuroScore has the potential to benefit researchers as well. By replacing multiple anatomical regions of interest (ROIs) with a single endpoint, it can enhance the sensitivity of research studies while encapsulating more information. Moreover, AD-NeuroScore could potentially provide an affordable participant screening tool to capture AD-specific neurodegeneration (Jack et al., 2016), particularly in retrospective studies lacking amyloid or tau biomarker data.

There are several important avenues for future research that could enhance the clinical and research utility of AD-NeuroScore. For instance, the value of AD-NeuroScore in the differential diagnosis of AD and related dementias could be explored, including its validation in data that includes other dementias such as FTD, and its association with biomarker data. Furthermore, developing norms and cutoffs and examining the sensitivity of these ranges in combination with modeling disease progression are important in deploying AD-NeuroScore. To this end, examining the sensitivity of these ranges in combination with modeling disease progression through Cox univariate hazard ratios represents an important future direction of this research. Additionally, incorporating the rate of change of AD-NeuroScore, a feature inspired by SPARE-AD (Davatzikos et al., 2009), could be explored in future studies to improve its performance.

However, there are some limitations that need to be addressed in future research. The sample size of patients with AD was much smaller than the CN and MCI cohorts, and this number decreased longitudinally, rendering some later tests underpowered. The AD subset only included patients with mild AD dementia, leaving more progressed patients understudied with regards to the metric. Furthermore, repeating this study in a real-world clinical sample that better represents the AD patient class at various stages of disease advancement would strengthen the findings presented in this work.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The Pacific Neuroscience Institute Foundation (PNIF) is a public charity that holds a patent for AD-NeuroScore. JB, EP, DM, and SP, are employees of PNIF. GK, SB, and PS served as consultants/advisors for PNIF. PT received partial grant support from Biogen, Inc., for research unrelated to this manuscript. ST, CC, and HZ have nothing to disclose.

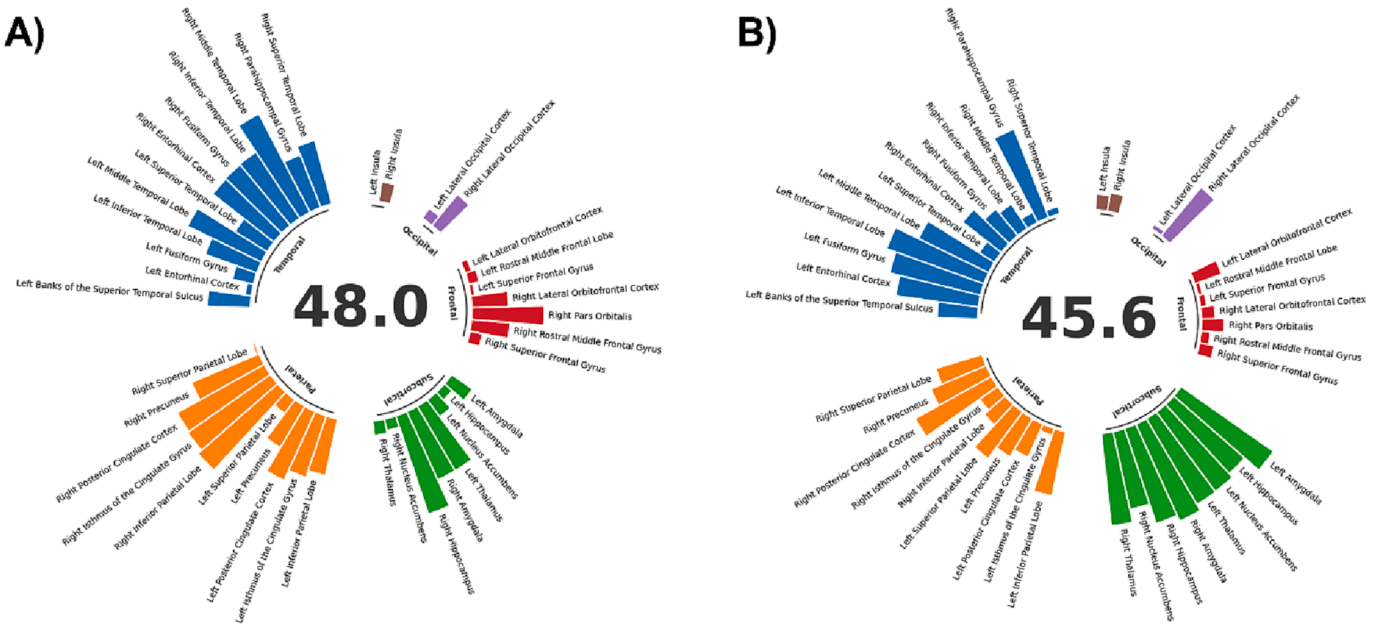

Fig. 6. Comparison of AD-NeuroScore breakdown for two patients with AD. Patient A shows relatively greater cortical involvement, especially within parietal and temporal cortices and with some frontal involvement, while Patient B exhibits relatively greater subcortical involvement. The center of the images represents the AD-NeuroScore value, and the bar heights depict the unweighted term in the score summation that represents the harmonized atrophy of the corresponding region. Different parts of the brain are coded by color.

Data availability

Open access to all imaging/clinical data is freely available with the ADNI. The code is openly available in [repository: "AD-NeuroScore"] at https://github.com/jbramen/AD-NeuroScore.

Acknowledgements

We thank the participants and families volunteering for this research. We thank the Pacific Neuroscience Institute and Foundation staff and leadership, including the CEO and Founder, Dan Kelly, MD, Vice President, Melissa Coleman, Director of Research Administration and Operations Melanie Lampa, Director of Marketing, Zara Jethani, Executive Assistant, Danielle Wozniak, and Accounting Manager, Bersabel Belay and the Pacific Brain Health Clinic's Practice Manager, Brenda Smith for their support. This work is supported by the Pacific Neuroscience Institute Foundation, including the generous support of the Barbara and John McLoughlin Family as well as the Cary and Will Singleton Family. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is 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. Additionally, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2023.103458.

References

Achterberg, H.C., van der Lijn, F., den Heijer, T., Vernooij, M.W., Ikram, M.A., Niessen, W.J., de Bruijne, M., 2014. Hippocampal shape is predictive for the development of dementia in a normal, elderly population. Hum. Brain Mapp. 35 (5), 2359–2371.

Ahdidan, J., Raji, C.A., DeYoe, E.A., Mathis, J., Noe, K.Ø., Rimestad, J., Kjeldsen, T.K., Mosegaard, J., Becker, J.T., Lopez, O., 2015. Quantitative Neuroimaging Software for Clinical Assessment of Hippocampal Volumes on MR Imaging. J. Alzheimers Dis. 49 (3), 723–732.

Archetti, D., Young, A.L., Oxtoby, N.P., Ferreira, D., Mårtensson, G., Westman, E., Alexander, D.C., Frisoni, G.B., Redolfi, A., 2021. Inter-cohort validation of sustain model for alzheimer's disease. Front Big Data 4, 30.

Attier-Zmudka J, Sérot JM, Valluy J, Saffarini M, Macaret AS, Diouf M, Dao S, Douadi Y, Piotr Malinowski K, Balédent O (2019) Decreased Cerebrospinal Fluid Flow Is Associated With Cognitive Deficit in Elderly Patients. *Front Aging Neurosci* **11**.

Balsis, S., Benge, J.F., Lowe, D.A., Geraci, L., Doody, R.S., 2015. How Do Scores on the ADAS-Cog, MMSE, and CDR-SOB Correspond? Clin. Neuropsychol. 29 (7), 1002–1009.

Braak H, Braak E, Braak (H, Braak E (1997) Staging of Alzheimer-Related Cortical Destruction. *Int Psychogeriatr* **9**, 257–261.

Brewer, J.B., Magda, S., Airriess, C., Smith, M.E., 2009. Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in alzheimer disease. Am. J. Neuroradiol. 30 (3), 578–580.

Cabeza, R., Ciaramelli, E., Olson, I.R., Moscovitch, M., 2008. Parietal cortex and episodic memory: an attentional account. Nat. Rev. Neurosci. 9, 613.

Casanova, R., Hsu, F.-C., Sink, K.M., Rapp, S.R., Williamson, J.D., Resnick, S.M., Espeland, M.A., Ginsberg, S.D., 2013. Alzheimer's disease risk assessment using large-scale machine learning methods. PLoS One 8 (11), e77949.

Casanova, R., Barnard, R.T., Gaussoin, S.A., Saldana, S., Hayden, K.M., Manson, J.A.E., Wallace, R.B., Rapp, S.R., Resnick, S.M., Espeland, M.A., Chen, J.C., 2018. Using high-dimensional machine learning methods to estimate an anatomical risk factor for Alzheimer's disease across imaging databases. Neuroimage 183, 401–411.

Cavedo, E., Tran, P., Thoprakarn, U., Martini, J.-B., Movschin, A., Delmaire, C., Gariel, F., Heidelberg, D., Pyatigorskaya, N., Ströer, S., Krolak-Salmon, P., Cotton, F., dos Santos, C.L., Dormont, D., 2022. Validation of an automatic tool for the rapid measurement of brain atrophy and white matter hyperintensity: QyScore®. Eur. Radiol. 32 (5), 2949–2961.

Chen, K., Reiman, E.M., Alexander, G.E., Caselli, R.J., Gerkin, R., Bandy, D., Domb, A., Osborne, D., Fox, N., Crum, W.R., Saunders, A.M., Hardy, J., 2007. Correlations between apolipoprotein E epsilon4 gene dose and whole brain atrophy rates. Am. J. Psychiatry 164, 916–921.

Chow, N., Hwang, K.S., Hurtz, S., Green, A.E., Somme, J.H., Thompson, P.M., Elashoff, D. A., Jack, C.R., Weiner, M., Apostolova, L.G., 2015. Comparing 3T and 1.5T MRI for mapping hippocampal atrophy in the Alzheimer's disease neuroimaging initiative. Am. J. Neuroradiol. 36 (4), 653–660.

Coupé, P., Fonov, V.S., Bernard, C., Zandifar, A., Eskildsen, S.F., Helmer, C., Manjón, J. V., Amieva, H., Dartigues, J.-F., Allard, M., Catheline, G., Collins, D.L., 2015. Detection of Alzheimer's disease signature in MR images seven years before conversion to dementia: toward an early individual prognosis. Hum. Brain Mapp. 36 (12), 4758–4770.

Csernansky, J.G., Wang, L., Joshi, S.C., Tilak Ratnanather, J., Miller, M.I., 2004. Computational anatomy and neuropsychiatric disease: probabilistic assessment of variation and statistical inference of group difference, hemispheric asymmetry, and time-dependent change. Neuroimage 23, S56–S68.

Da, X., Toledo, J.B., Zee, J., Wolk, D.A., Xie, S.X., Ou, Y., Shacklett, A., Parmpi, P., Shaw, L., Trojanowski, J.Q., Davatzikos, C., 2014. Integration and relative value of biomarkers for prediction of MCI to AD progression: Spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. Neuroimage Clin. 4, 164–173.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I segmentation and surface reconstruction. Neuroimage 9 (2), 179–194.

Davatzikos, C., Xu, F., An, Y., Fan, Y., Resnick, S.M., 2009. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain 132 (8), 2026–2035.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31 (3), 968–980.

Diciotti, S., Ginestroni, A., Bessi, V., Giannelli, M., Tessa, C., Bracco, L., Mascalchi, M., Toschi, N., 2012. Identification of mild Alzheimer's disease through automated classification of structural MRI features. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 2012, 428–431.

Dickerson, B.C., Goncharova, I., Sullivan, M.P., Forchetti, C., Wilson, R.S., Bennett, D.A., Beckett, L.A., DeToledo-Morrell, L., 2001. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. Neurobiol. Aging 22, 747–754.

Dukart, J., Mueller, K., Barthel, H., Villringer, A., Sabri, O., Schroeter, M.L., 2013. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. Psychiatry Res. 212 (3), 230–236.

Dumitrescu A, Rote G (2004) *On the Fréchet distance of a set of curves.*

Eweje, F.R., Byun, S., Chandra, R., Hu, F., Kamel, I., Zhang, P., Jiao, Z., Bai, H.X., 2022. Translatability Analysis of National Institutes of Health-Funded Biomedical Research That Applies Artificial Intelligence. JAMA Netw. Open 5 (1), e2144742.

Ezzati, A., Zammit, A.R., Harvey, D.J., Habeck, C., Hall, C.B., Lipton, R.B., 2019. Optimizing machine learning methods to improve predictive models of Alzheimer's disease. J. Alzheimers Dis. 71 (3), 1027–1036.

Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9 (2), 195–207.

Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8 (4), 272–284.

Fischl, B., Liu, A., Dale, A., ~m., 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Med. Imag. 20, 70–80.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.

Fischl, B., Salat, D.H., van der Kouwe, A.J.W., Makris, N., Ségonne, F., Quinn, B.T., Dale, A.M., 2004. Sequence-independent segmentation of magnetic resonance images. Neuroimage 23, S69–S84.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A. M., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12 (3), 189–198.

Fonteijn, H.M., Modat, M., Clarkson, M.J., Barnes, J., Lehmann, M., Hobbs, N.Z., Scahill, R.I., Tabrizi, S.J., Ourselin, S., Fox, N.C., Alexander, D.C., 2012. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. Neuroimage 60 (3), 1880–1889.

Frisoni, G.B., Laakso, M.P., Beltramello, A., Geroldi, C., Bianchetti, A., Soininen, H., Trabucchi, M., 1999. Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease. Neurology 52 (1), 91.

Giesel FL, Hahn HK, Thomann PA, Widjaja E, Wignall E, von Tengg-Kobligk H, Pantel J, Griffiths PD, Peitgen HO, Schroder J, Essig M Temporal Horn Index and Volume of Medial Temporal Lobe Atrophy Using a New Semiautomated Method for Rapid and Precise Assessment.

Gómez-Isla, T., Price, J.L., McKeel Jr., D.W., Morris, J.C., Growdon, J.H., Hyman, B.T., 1996. Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. J. Neurosci. 16 (14), 4491–4500.

Gosche, K.M., Mortimer, J.A., Smith, C.D., Markesbery, W.R., Snowdon, D.A., 2002. Hippocampal volume as an index of Alzheimer neuropathology. Neurology 58, 1476–1482.

Grochowalski, J.H., Liu, Y., Siedlecki, K.L., 2016. Examining the reliability of ADAS-Cog change scores. Aging Neuropsychol. Cogn. 23 (5), 513–529.

Gutman, B.A., Hua, X., Rajagopalan, P., Chou, Y.-Y., Wang, Y., Yanovsky, I., Toga, A.W., Jack, C.R., Weiner, M.W., Thompson, P.M., 2013. Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features. Neuroimage 70, 386–401.

Harper, L., Bouwman, F., Burton, E.J., Barkhof, F., Scheltens, P., O'Brien, J.T., Fox, N.C., Ridgway, G.R., Schott, J.M., 2017. Patterns of atrophy in pathologically confirmed dementias: a voxelwise analysis. J. Neurol. Neurosurg. Psychiatry 88 (11), 908–916.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585 (7825), 357–362.

Hotelling, H., 1947. Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights. In: Eisenhart, C., Hastay, M.W., Wallis, W.A. (Eds.), Techniques of Statistical Analysis. McGraw Hill, New York, pp. 111–184.

Hua, X., Hibar, D.P., Ching, C.R.K., Boyle, C.P., Rajagopalan, P., Gutman, B.A., Leow, A. D., Toga, A.W., Jack, C.R., Harvey, D., Weiner, M.W., Thompson, P.M., 2013. Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer's disease clinical trials. Neuroimage 66, 648–661.

Hughes, C.P., Berg, L., Danziger, W., Coben, L.A., Martin, R.L., 1982. A new clinical scale for the staging of dementia. Br. J. Psychiatry 140 (6), 566–572.

Hwang, E.J., Park, C.M., 2020. Clinical implementation of deep learning in thoracic radiology: potential applications and challenges. Korean J. Radiol. 21 (5), 511.

Jack, C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G. E., Ivnik, R.J., Kokmen, E., 1997. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology 49 (3), 786–794.

Jack, C.R., Dickson, D.W., Parisi, J.E., Xu, Y.C., Cha, R.H., O'Brien, P.C., Edland, S.D., Smith, G.E., Boeve, B.F., Tangalos, E.G., Kokmen, E., Petersen, R.C., 2002. Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. Neurology 58 (5), 750–757.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Gunnar Krueger, Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27 (4), 685–691.

Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Feldman, H.H., Frisoni, G.B., Hampel, H., Jagust, W.J., Johnson, K.A., Knopman, D.S., Petersen, R.C., Scheltens, P., Sperling, R.A., Dubois, B., 2016. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology 87 (5), 539–547.

Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J.L., Montine, T., Phelps, C., Rankin, K.P., Rowe, C.C., Scheltens, P., Siemers, E., Snyder, H.M., Sperling, R., Elliott, C., Masliah, E., Ryan, L., Silverberg, N., 2018. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 14 (4), 535–562.

Jacobs, H.I.L., Van Boxtel, M.P.J., Jolles, J., Verhey, F.R.J., Uylings, H.B.M., 2012. Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings. Neurosci. Biobehav. Rev. 36 (1), 297–309.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. Neuroimage 30 (2), 436–443.

Kochunov, P., Ma, Y., Hatch, K.S., Schmaal, L., Jahanshad, N., Thompson, P.M., Adhikari, B.M., Bruce, H., Chiappelli, J., Van der Vaart, A., Goldwaser, E.L., Sotiras, A., Ma, T., Chen, S., Nichols, T.E., Hong, L.E., 2022. Separating clinical and subclinical depression by big data informed structural vulnerability index and its impact on cognition: ENIGMA Dot Product. Pac. Symp. Biocomput. 27, 133.

Laakso, M.P., Partanen, K., Riekkinen, P., Lehtovirta, M., Helkala, E.-L., Hallikainen, M., Hanninen, T., Vainio, P., Soininen, H., 1996. Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia: an MRI study. Neurology 46 (3), 678–681.

Laakso, M.P., Soininen, H., Partanen, K., Lehtovirta, M., Hallikainen, M., Hänninen, T., Helkala, E.-L., Vainio, P., Riekkinen, P.J., 1998. MRI of the hippocampus in Alzheimer's disease: sensitivity, specificity, and analysis of the incorrectly classified subjects. Neurobiol. Aging 19 (1), 23–31.

Larson, D.B., Boland, G.W., 2019. Imaging Quality Control in the Era of Artificial Intelligence. J. Am. Coll. Radiol. 16 (9), 1259–1266.
Lindeboom, J., Weinstein, H., 2004. Neuropsychology of cognitive ageing, minimal cognitive impairment, Alzheimer's disease, and vascular cognitive impairment. Eur. J. Pharmacol. 490 (1-3), 83–86.
Ma, D.a., Popuri, K., Bhalla, M., Sangha, O., Lu, D., Cao, J., Jacova, C., Wang, L., Beg, M. F., 2019. Quantitative assessment of field strength, total intracranial volume, sex, and age effects on the goodness of harmonization for volumetric analysis on the ADNI database. Hum. Brain Mapp. 40 (5), 1507–1527.
Maiseli, B.J., 2021. Hausdorff Distance with Outliers and Noise Resilience Capabilities. SN Comput Sci 2 (5).
Mazziotta, J.C., Toga, A.W., Evans, A.C., Fox, P.T., Lancaster, J.L., 1995. Digital brain atlases. Trends Neurosci. 18 (5), 210–211.
Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). Neuroimage 2 (2), 89–101.
Morris, J.C., 1993. The clinical dementia rating (CDR). Neurology 43 (11), 2412-2412-a.
Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement. 1 (1), 55–66.
Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin. N. Am. 15 (4), 869–877.
Mukherji D, Mukherji M, Mukherji N, Initiative ADN (2021) Early Detection of Alzheimer's Disease with Low-Cost Neuropsychological Tests: A Novel Predict-Diagnose Approach using Recurrent Neural Networks. *medRxiv* 2021.01.17.21249822.
Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R., 2008. Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. Brain 131 (9), 2443–2454.
Nie X, Sun Y, Wan S, Zhao H, Liu R, Li X, Wu S, Nedelska Z, Hort J, Qing Z, Xu Y, Zhang B (2017) Subregional Structural Alterations in Hippocampus and Nucleus Accumbens Correlate with the Clinical Impairment in Patients with Alzheimer's Disease Clinical Spectrum: Parallel Combining Volume and Vertex-Based Approach. *Front Neurol* **8**.
O'Bryant, S.E., Waring, S.C., Cullum, C.M., Hall, J., Lacritz, L., Massman, P.J., Lupo, P.J., Reisch, J.S., Doody, R., 2008. Staging dementia using clinical dementia rating scale sum of boxes scores: A texas Alzheimer's research consortium study. Arch. Neurol. 65, 1091.
Ott, B.R., Cohen, R.A., Gongvatana, A., Okonkwo, O.C., Johanson, C.E., Stopa, E.G., Donahue, J.E., Silverberg, G.D., 2010. Brain ventricular volume and cerebrospinal fluid biomarkers of Alzheimer's disease. J. Alzheimers Dis. 20 (2), 647–657.
Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q., Weiner, M.W., 2010. Alzheimer's Disease neuroimaging initiative (ADNI): clinical characterization. Neurology 74 (3), 201–209.
Pinto, M.F., Leal, A., Lopes, F., Pais, J., Dourado, A., Sales, F., Martins, P., Teixeira, C.A., 2022. On the clinical acceptance of black-box systems for EEG seizure prediction. Epilepsia Open 7.
Popuri, K., Ma, D., Wang, L., Beg, M.F., 2020. Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: Independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. Hum. Brain Mapp. 41, 4127–4147.
Rabinovici, G.D., Seeley, W.W., Kim, E.J., Gorno-Tempini, M.L., Rascovsky, K., Pagliaro, T.A., Allison, S.C., Halabi, C., Kramer, J.H., Johnson, J.K., Weiner, M.W., Forman, M.S., Trojanowski, J.Q., DeArmond, S.J., Miller, B.L., Rosen, H.J., 2008. Distinct MRI atrophy patterns in autopsy-proven Alzheimer's disease and frontotemporal lobar degeneration. Am. J. Alzheimers Dis. Other Demen. 22 (6), 474–488.
Rallabandi, V.P.S., Tulpule, K., Gattu, M., 2020. Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis. Inform Med Unlocked 18, 100305.
Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: A robust approach. Neuroimage 53 (4), 1181–1196.
Salvatore, C., Cerasa, A., Castiglioni, I., 2018. MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia 24 months before probable diagnosis. Front. Aging Neurosci. 10, 135.
Schott, J.M., Price, S.L., Frost, C., Whitwell, J.L., Rossor, M.N., Fox, N.C., 2005. Measuring atrophy in Alzheimer disease. Neurology 65, 119–124.
Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22 (3), 1060–1075.
Segonne, F., Pacheco, J., Fischl, B., 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans. Med. Imaging 26 (4), 518–529.
Sørensen, L., Igel, C., Liv Hansen, N., Osler, M., Lauritzen, M., Rostrup, E., Nielsen, M., 2016. Early detection of Alzheimer's disease using MRI hippocampal texture. Hum. Brain Mapp. 37, 1148–1161.
Stein, J.L., Medland, S.E., Vasquez, A.A., Hibar, D.P., Senstad, R.E., Winkler, A.M., Toro, R., Appel, K., Bartecek, R., Bergmann, Ø., Bernard, M., Brown, A.A., Cannon, D.M., Chakravarty, M.M., Christoforou, A., Domin, M., Grimm, O., Hollinshead, M., Holmes, A.J., Homuth, G., Hottenga, J.-J., Langan, C., Lopez, L.M., Hansell, N.K., Hwang, K.S., Kim, S., Laje, G., Lee, P.H., Liu, X., Loth, E., Lourdusamy, A., Mattingsdal, M., Mohnke, S., Maniega, S.M., Nho, K., Nugent, A.C., O'Brien, C., Papmeyer, M., Pütz, B., Ramasamy, A., Rasmussen, J., Rijpkema, M., Risacher, S.L., Roddey, J.C., Rose, E.J., Ryten, M., Shen, L.i., Sprooten, E., Strengman, E., Teumer, A., Trabzuni, D., Turner, J., van Eijk, K., van Erp, T.G.M., van Tol, M.-J., Wittfeld, K., Wolf, C., Woudstra, S., Aleman, A., Alhusaini, S., Almasy, L., Binder, E.B., Brohawn, D.G., Cantor, R.M., Carless, M.A., Corvin, A., Czisch, M., Curran, J.E., Davies, G., de Almeida, M.A.A., Delanty, N., Depondt, C., Duggirala, R., Dyer, T.D., Erk, S., Fagerness, J., Fox, P.T., Freimer, N.B., Gill, M., Göring, H.H.H., Hagler, D.J., Hoehn, D., Holsboer, F., Hoogman, M., Hosten, N., Jahanshad, N., Johnson, M.P., Kasperaviciute, D., Kent, J.W., Kochunov, P., Lancaster, J.L., Lawrie, S.M., Liewald, D.C., Mandl, R., Matarin, M., Mattheisen, M., Meisenzahl, E., Melle, I., Moses, E.K., Mühleisen, T.W., Nauck, M., Nöthen, M.M., Olvera, R.L., Pandolfo, M., Pike, G.B., Puls, R., Reinvang, I., Rentería, M.E., Rietschel, M., Roffman, J.L., Royle, N.A., Rujescu, D., Savitz, J., Schnack, H.G., Schnell, K., Seiferth, N., Smith, C., Steen, V.M., Valdés Hernández, M.C., Van den Heuvel, M., van der Wee, N.J., Van Haren, N.E.M., Veltman, J.A., Völzke, H., Walker, R., Westlye, L.T., Whelan, C.D., Agartz, I., Boomsma, D.I., Cavalleri, G.L., Dale, A.M., Djurovic, S., Drevets, W.C., Hagoort, P., Hall, J., Heinz, A., Jack, C.R., Foroud, T.M., Le Hellard, S., Macciardi, F., Montgomery, G.W., Poline, J.B., Porteous, D.J., Sisodiya, S.M., Starr, J.M., Sussmann, J., Toga, A.W., Veltman, D.J., Walter, H., Weiner, M.W., Bis, J.C., Ikram, M.A., Smith, A.V., Gudnason, V., Tzourio, C., Vernooij, M.W., Launer, L.J., DeCarli, C., Seshadri, S., Andreassen, O.A., Apostolova, L.G., Bastin, M.E., Blangero, J., Brunner, H.G., Buckner, R.L., Cichon, S., Coppola, G., de Zubicaray, G.I., Deary, I.J., Donohoe, G., de Geus, E.J.C., Espeseth, T., Fernández, G., Glahn, D.C., Grabe, H.J., Hardy, J., Hulshoff Pol, H.E., Jenkinson, M., Kahn, R.S., McDonald, C., McIntosh, A.M., McMahon, F.J., McMahon, K.L., Meyer-Lindenberg, A., Morris, D.W., Müller-Myhsok, B., Nichols, T. E., Ophoff, R.A., Paus, T., Pausova, Z., Penninx, B.W., Potkin, S.G., Sämann, P.G., Saykin, A.J., Schumann, G., Smoller, J.W., Wardlaw, J.M., Weale, M.E., Martin, N.G., Franke, B., Wright, M.J., Thompson, P.M., 2012. Identification of common variants associated with human hippocampal and intracranial volumes. Nat. Genet. 44 (5), 552–561.
Thompson, P.M., Hayashi, K.M., de Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W., 2003. Dynamics of gray matter loss in Alzheimer's disease. J. Neurosci. 23 (3), 994–1005.
Thompson, P.M., Hayashi, K.M., Dutton, R.A., Chiang, M.-C., Leow, A.D., Sowell, E.R., De zubicaray, G., Becker, J.T., Lopez, O.L., Aizenstein, H.J., Toga, A.W., 2007. Tracking Alzheimer's disease. Ann. N. Y. Acad. Sci. 1097 (1), 183–214.
Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack, C.R., 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 39 (3), 1186–1197.
Vemuri, P., Whitwell, J.L., Kantarci, K., Josephs, K.A., Parisi, J.E., Shiung, M.S., Knopman, D.S., Boeve, B.F., Petersen, R.C., Dickson, D.W., Jack, C.R., 2008. Antemortem MRI based structural abnormality index (STAND)-scores correlate with postmortem braak neurofibrillary tangle stage. Neuroimage 42 (2), 559–567.
Vemuri, P., Wiste, H.J., Weigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C., Jack, C.R., 2009. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. Neurology 73 (4), 294–301.
Vercelletto, M., Martinez, F., Lanier, S., Magne, C., Jaulin, P., Bourin, M., 2002. How to define treatment success using cholinesterase inhibitors. Int. J. Geriatr. Psychiatry 17, 388–390.
Wang, G., Dong, Q., Wu, J., Su, Y.i., Chen, K., Su, Q., Zhang, X., Hao, J., Yao, T., Liu, L.i., Zhang, C., Caselli, R.J., Reiman, E.M., Wang, Y., 2021. Developing univariate neurodegeneration biomarkers with low-rank and sparse subspace decomposition* HHS Public Access. Med. Image Anal. 67, 101877.
Yin, C., Li, S., Zhao, W., Feng, J., 2013. Brain imaging of mild cognitive impairment and Alzheimer's disease. Neural Regen. Res. 8, 435–444.
Young, A.L., Oxtoby, N.P., Daga, P., Cash, D.M., Fox, N.C., Ourselin, S., Schott, J.M., Alexander, D.C., 2014. A data-driven model of biomarker changes in sporadic Alzheimer's disease. Brain 137 (9), 2564–2577.
Young, A.L., Vogel, J.W., Aksman, L.M., Wijeratne, P.A., Eshaghi, A., Oxtoby, N.P., Williams, S.C.R., Alexander, D.C., 2021. Ordinal SuStaIn: subtype and stage inference for clinical scores, visual ratings, and other ordinal data. Front Artif Intell 4, 111.
Zanchi, D., Giannakopoulos, P., Borgwardt, S., Rodriguez, C., Haller, S., 2017. Hippocampal and amygdala gray matter loss in elderly controls with subtle cognitive decline. Front. Aging Neurosci. 9.