A Machine Learning Model to Harmonize Volumetric Brain MRI Data for Quantitative Neuroradiologic Assessment of Alzheimer Disease

Damiano Archetti, MSc • Vikram Venkatraghavan, PhD* • Béla Weiss, PhD • Pierrick Bourgeat, PhD • Tibor Auer, MD, PhD • Zoltán Vidnyánszky, PhD • Stanley Durrleman, PhD • Wiesje M. van der Flier, PhD • Frederik Barkhof, MD, PhD • Daniel C. Alexander, PhD • Andre Altmann, PhD • Alberto Redolfi, PhD • Betty M. Tijms, PhD • Neil P. Oxtoby, PhD • for the Australian Imaging, Biomarker and Lifestyle Study[1] • for the Alzheimer's Disease Neuroimaging Initiative[2] • for the E-DADS (Early Detection of Alzheimer's Disease Subtypes) Consortium*

From the Laboratory of Neuroinformatics, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Via Pilastroni 4, Brescia 25125, Italy (D.A., A.R.); Alzheimer Centre Amsterdam, Neurology, Vrije Universiteit, Amsterdam UMC, location VUmc, Amsterdam, the Netherlands (V.V., W.M.v.d.F., B.M.T.); Amsterdam Neuroscience, Neurodegeneration, Amsterdam, the Netherlands (V.V., W.M.v.d.F., B.M.T.); Brain Imaging Centre, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary (B.W., T.A., Z.V.); Biomatics and Applied Artificial Intelligence Institute, John von Neumann Faculty of Informatics, Óbuda University, Budapest, Hungary (B.W.); The Australian e-Health Research Centre, CSIRO Health and Biosecurity, Brisbane, Australia (P.B.); School of Psychology, University of Surrey, Guildford, United Kingdom (T.A.); Sorbonne Université, Institut du Cerveau-Paris Brain Institute–ICM, CNRS, Inria, Inserm, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France (S.D.); Department of Epidemiology and Data Science, Vrije Universiteit, Amsterdam UMC, location VUmc, Amsterdam, the Netherlands (W.M.v.d.F.); Department of Radiology and Nuclear Medicine, Amsterdam UMC, Vrije Universiteit, Amsterdam, the Netherlands (F.B.); Queen Square Institute of Neurology, University College London, United Kingdom (F.B.); and UCL Hawkes Institute, Department of Medical Physics and Biomedical Engineering and Department of Computer Science, University College London, London, United Kingdom (F.B., D.C.A., A.A., N.P.O.). Received January 23, 2024; revision requested March 11; revision received October 12; accepted November 15. **Address correspondence to** D.A. (email: *darchetti@fatebenefratelli.eu*).

* D.A. and V.V. contributed equally to this work.

[1] Data used in the preparation of this article were obtained from the Australian Imaging, Biomarker and Lifestyle (AIBL) Flagship Study of Ageing. Unless named, the AIBL researchers contributed data but did not participate in analysis or the writing of this report. AIBL researchers are listed at *https://aibl.org.au*.

[2] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database *(adni.loni.usc.edu)*. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or the 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*.

Supported by the Early Detection of Alzheimer's Disease Subtypes project, a European Union Joint Programme—Neurodegenerative Disease Research (JPND) project *(www.jpnd.eu)*. The project was also supported under the aegis of JPND through the following funding organizations: Medical Research Council, United Kingdom (MR/T046422/1); ZonMW, Netherlands (733051106); Agence Nationale de la Recherche, France (ANR-19-JPW2–000); Italian Ministry of Health (MoH Grant/Award Numbers: JPNDJPco-fuND2-"EDADS"No.825664 and in part through the IRCCS Research Program - Ricerca Corrente 2022-2024, Linea n. 2 "Piattaforme elettroniche per analisi di immagini cerebrali"), Italy; National Health and Medical Research Council, Australia (1191535); and National Research, Development and Innovation Office, Hungary (2019–2.1.7-ERA-NET-2020–00008). This work used the Dutch national e-infrastructure with the support of the SURF Cooperative (grant no. EINF-5353). B.W. was supported by the Consolidator Researcher program of the Óbuda University, Budapest, Hungary. B.W. and Z.V. were supported by project no. RRF-2.3.1-21-2022-00015, which was implemented with support provided by the European Union. F.B. was supported by the National Institute for Health and Care Research Biomedical Research Centre at University College London Hospitals. N.P.O. was supported by a UK Research and Innovation Future Leaders Fellowship (UK Medical Research Council MR/S03546X/1).

Conflicts of interest are listed at the end of this article.

See also commentary by Haller in this issue.

Radiology: Artificial Intelligence 2025; 7(1):e240030 • **https://doi.org/10.1148/ryai.240030** • **Content codes:** AI NR MR

Purpose: To extend a previously developed machine learning algorithm for harmonizing brain volumetric data of individuals undergoing neuroradiologic assessment of Alzheimer disease not encountered during model training.

Materials and Methods: Neuroharmony is a recently developed method that uses image quality metrics as predictors to remove scanner-related effects in brain-volumetric data using random forest regression. To account for the interactions between Alzheimer disease pathology and image quality metrics during harmonization, the authors developed a multiclass extension of Neuroharmony for individuals with and without cognitive impairment. Cross-validation experiments were performed to benchmark performance against other available strategies using data from 20 864 participants with and without cognitive impairment, spanning 11 prospective and retrospective cohorts and 43 scanners. Evaluation metrics assessed the ability to remove scanner-related variations in brain volumes (marker concordance between scanner pairs) while retaining the ability to delineate different diagnostic groups (preserving disease-related signal).

Results: For each strategy, marker concordances between scanners were significantly better ($P < .001$) compared with preharmonized data. The proposed multiclass model achieved significantly higher concordance (mean, 0.75 ± 0.09 [SD]) than the Neuroharmony model trained on individuals without cognitive impairment (mean, 0.70 ± 0.11) and preserved disease-related signal (ΔAUC [area under the receiver operating characteristic curve] = -0.006 ± 0.027) better than the Neuroharmony model trained on individuals with and without cognitive impairment that did not use the proposed extension (ΔAUC = -0.091 ± 0.036). The marker concordance was better in scanners seen during training (concordance > 0.97) than unseen (concordance < 0.79), independent of cognitive status.

Conclusion: In a large-scale multicenter dataset, the proposed multiclass Neuroharmony model outperformed other available strategies for harmonizing brain volumetric data from unseen scanners in a clinical setting.

Supplemental material is available for this article.

Abbreviations

AD = Alzheimer disease, ADNI = Alzheimer's Disease Neuroimaging Initiative, AUC = area under the receiver operating characteristic curve, FDR = false discovery rate, HC = healthy control, MCI = mild cognitive impairment, MRIQC = MRI Quality Control, SCD = subjective cognitive decline, UKBB = UK Biobank

Summary

A multiclass Neuroharmony model was developed and evaluated against other approaches for harmonizing volumetric data in a clinical setting using a large, multicenter brain MRI dataset of individuals undergoing neuroradiologic assessment of Alzheimer disease.

Key Points

- The proposed multiclass Neuroharmony model, trained on 20 864 participants, achieved state-of-the-art performance in harmonizing brain volumetric data from new MRI scanners.
- The proposed multiclass Neuroharmony model preserved disease signal on volumetric features better than the other tested approaches (ΔAUC = -0.006 ± 0.027).
- The multiclass Neuroharmony model performed better for harmonizing MRI-derived volumetric data in the clinical setting than other available approaches for harmonizing data from previously unseen scanners.

Keywords

Image Postprocessing, MR Imaging, Dementia, Random Forest

Structural MRI scans, such as T1-weighted MRI, are routinely acquired in memory clinics for diagnosing Alzheimer disease (AD) (1), performing clinical phenotyping (2), and differentiating AD from other types of dementias (3). In current clinical practice, radiologists primarily assess global and regional brain atrophy through visual examination of MRI. However, visual examinations are subjective and prone to intrarater and interrater variability. Quantitative imaging markers, such as brain volumetric data, are becoming increasingly popular due to their potential to improve diagnostic confidence (4). Quantitative imaging markers can be used for objective assessment in the radiologic workflow either by using automated digital tools based on normative modeling (3) or using the latest advances in artificial intelligence, including brain-age estimation (5) and data-driven subtyping (6).

However, differences in MRI acquisition protocols and scanners affect consistency and reproducibility of brain volumetry (7) and are a major impediment for the clinical translation of automated tools. To tackle this problem, many data harmonization tools have emerged in recent years (8). Such algorithms can harmonize either original scans (eg, DeepHarmony) (9) or derivatives extracted from the scans (eg, ComBat) (10). Some of these algorithms have been shown to harmonize patient data affected by a neurodegenerative disease (11,12) while preserving disease-related signature. However, such harmonization techniques typically work only for the scanner models they have been trained on and, in some instances, require the same individuals to be scanned with different scanners (13). Harmonizing volumetric data from MRI scanners not encountered during initial model training requires additional training with a substantial number of images from these scanners (14). This requirement poses a challenge for the deployment of such models for clinical use.

Neuroharmony (15) is a recently developed harmonization approach that can harmonize volumetric data from images acquired using new and unseen MRI scanners. The Neuroharmony model is trained to predict the volumetric corrections estimated by ComBat harmonization in the training phase. When trained on large enough samples, the model generalizes well for predictions of harmonized volumes in previously unseen scanners. It works under the assumption that the corrections needed to harmonize data from multiple scanners can be predicted from image quality metrics computed from the scans. Although the original Neuroharmony study indicated that harmonization works for healthy individuals (15), harmonizing data from patients with neurodegenerative diseases remains an open problem. This issue is a limitation because disease pathology in patients may affect the image quality metrics, and such effects remain unaccounted for in a Neuroharmony model trained on healthy controls (HCs).

In this article, we proposed an extension of the Neuroharmony model to account for interactions between disease pathology and image quality metrics to remove scanner-related effects (multiclass model of Neuroharmony). We systematically compared the performances of the proposed multiclass model in harmonizing data with two other approaches: the original Neuroharmony model trained only on individuals without cognitive impairment (normative model of Neuroharmony) and the original Neuroharmony model trained on individuals with and without cognitive impairment that did not use our proposed multiclass extension (inclusive model of Neuroharmony). We used data from 11 cohorts across three continents to evaluate these approaches. Last, we identified key challenges for clinical implementation of the best multicentric harmonization strategy identified in our experiments for enabling quantitative neuroradiologic assessment of AD.

Materials and Methods

Study Participants and Data

T1-weighted three-dimensional MRI data of HCs and individuals with subjective cognitive decline (SCD), mild cognitive impairment (MCI), and AD from 11 prospective and retrospective data cohorts were included in our analysis. The cohorts considered for this study were the Amsterdam Dementia Cohort (16); Alzheimer's Disease Neuroimaging Initiative (ADNI) (17); Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing (18); Alzheimer's Repository Without Borders (19); European DTI Study on Dementia (20); Hungarian Longitudinal Study of Healthy Brain Aging (21); Italian Alzheimer's Disease Neuroimaging Initiative (22); National Alzheimer's Coordination Center (23); Open Access Series of Imaging Studies (versions 1 and 2) (24); European Alzheimer's Disease Neuroimaging Initiative (also known as PharmaCog) (25); and UK Biobank (UKBB) (26). Detailed information about each cohort is summarized in Table S1. The clinical diagnoses of participants in these cohorts were made based on international consensus criteria; further details can be found in the respective studies cited above. Each study was approved by the respective institutional ethical committees, with informed consent obtained from each participant.

Minimum inclusion criteria included the availability of a T1-weighted three-dimensional MRI scan along with age, sex, and scanner information and a clinical diagnosis of HCs, SCD, MCI, or AD. All datasets were organized according to the Brain Imaging Data Structure standard (27) to ensure interoperability and data anonymization. An overview of the scanners used in this study is shown in Table 1, and the scanning parameters are summarized in Table S2.

Table 1: Scanners Considered in This Study and Their Characteristics

		No. of Scans by Sex	
Manufacturer and Scanner Model	Magnetic Field (T)	Female	Male
Canon			
Titan	3.0	252	329
GE			
Discovery MR750	3.0	290	372
Discovery MR750w	3.0	8	16
Genesis Signa	1.5	6	3
Signa Excite	1.5	181	197
Signa PET/MR	3.0	15	16
Signa HDx	1.5	10	19
Signa HDx	3.0	44	55
Signa HDxt	1.5	225	261
Signa HDxt	3.0	463	535
Signa Premier	3.0	6	8
Philips			
Achieva	1.5	4	7
Achieva	3.0	179	116
Achieva dStream	3.0	13	10
Eclipse	1.5	28	13
Gemini	3.0	312	214
Gyroscan NT	1.0	127	68
Ingenia	3.0	18	33
Ingenuity	3.0	298	339
Intera	1.0	275	161
Intera	1.5	19	42
Intera	3.0	27	27
Intera Achieva	1.5	1	4
Intera Gyroscan	1.5	11	16
Siemens			
Allegra	3.0	50	34
Avanto	1.5	150	153
Biograph	3.0	0	5
Espree	1.5	3	4
Magnetom Expert	1.0	397	416
Magnetom Impact	1.0	7	2
Magnetom Vida	3.0	6	14
Magnetom Vision	1.5	20	7
Prisma	3.0	131	122
Prisma fit	3.0	122	84
RCNS	3.0	68	48
Skyra	3.0	6146	5035
Sonata	1.5	189	226
Sonata Vision	1.5	3	2
Symphony	1.5	90	66
Trio	3.0	41	21
Trio Tim	3.0	457	380
Verio	3.0	186	137
Vision	1.5	233	126

Image Processing

Cortical reconstruction and volumetric segmentation were performed with the cross-sectional pipeline of FreeSurfer, version 7.1.1 (28), to extract volumes of 68 cortical regions in the Desikan-Killiany atlas and 14 subcortical brain regions as well as total cerebrospinal fluid volume, total gray matter volume, and total brain volume with and without ventricles. Figure S1 lists all features derived from FreeSurfer. Image quality metrics were estimated using the MRI Quality Control (MRIQC) tool, version 0.16.1 (29). Automatic quality control of the FreeSurfer segmentations was performed using the Euler number, where outliers, defined as 1.5 × IQR below the first quartile (30), for each scanner were excluded from our experiments.

To ensure reproducibility of our results across different computing environments (31), Docker containers for both FreeSurfer and the MRIQC tool were prepared by one author (N.P.O.) and shared with coauthors (D.A., V.V., B.W., P.B.) to process MRI scans from their local cohort (no images were shared; blinding was not necessary). These authors each have 5–15 years of MRI processing experience. The containers have been made available online to benefit the community (see the Data and Code Availability section).

Multiclass Neuroharmony Model

In the training phase, volumetric data from all individuals in the training set were harmonized using ComBat (10) with empirical Bayes optimization to remove scanner-related batch effects. While training, we imposed constraints that preserved the effects of age, sex, and cognitive status. Cognitive status was dichotomized based on the clinical diagnosis as either no cognitive impairment (HCs and SCD) or cognitive impairment (MCI and AD). Subsequently, a random forest regressor was trained with MRIQC-derived image quality metrics to predict the corrections needed to harmonize the volumes as predicted by ComBat. Additionally, to preserve disease-related signal during harmonization, we used the synthetic minority oversampling technique (32) to avoid class imbalance (no cognitive impairment vs cognitive impairment) before training the random forest regressor. This ensured that image quality metric values with and without neurodegeneration were equally distributed. The use of dichotomized cognitive status instead of clinical diagnosis ensured that in the test phase, a full clinical diagnosis was not required to predict the harmonized volumes. The hyperparameters for the random forest regressor were chosen to be the same as the ones used in the original Neuroharmony article (15).

Model Comparisons

The performance of the proposed multiclass extension of Neuroharmony was compared with two other harmonization strategies that are generalizable to external datasets.

Normative model.— In the training phase, volumetric data from only individuals without cognitive impairment were harmonized using ComBat harmonization using the aforementioned strategy while preserving the effects of age and sex. Subsequently, a random forest regressor was trained to predict the corrections needed to harmonize the volumes, as predicted by ComBat using MRIQC-derived image quality metrics.

Inclusive model.— The training strategy remained the same as for the normative model, but volumetric data of individuals with and without cognitive impairment were used.

Measures for Model Evaluation

We used two measures for model evaluation to assess how well each method removed unwanted scanner-related noise while retaining disease-related signal. First, we defined "marker concordance" (details below) as a statistical measure of similarity between brain-volumetric data from different scanners. Increased marker concordance after harmonization shows that a method successfully reduces scanner-related variance (see the Statistical Analysis section for details). Second, we used classification performance (HCs vs AD) to assess the amount of disease-related signal. The best performing harmonization model will return the best classification performance.

We used area under the receiver operating characteristic curve (AUC) to quantify the amount of disease-related signal that was retained in the volumetric measures after harmonization. The receiver operating characteristic curve for distinguishing HC participants from individuals with AD was computed independently for each volumetric measure with logistic regression. A reference measure for AUC was also computed for the nonharmonized data.

Cross-Validation Experiments

We performed two experiments in a cross-validation framework. Experiment 1 assessed concordance of the three harmonization strategies by performing cross-validation at the scanner level. Experiment 2 performed cross-validation at the participant level using the best performing scanner-level harmonization models.

Experiment 1.— To investigate the generalizability of the model to unseen scanners (not included in the training set), we performed fivefold cross-validation across the 43 available scanners. In each fold, 80% of the scanners were used for training the models, and the remaining 20% of the scanners were used for evaluation. To evaluate the bias introduced by using single-scanner data from the large UKBB cohort, we repeated this experiment for increasing portions of UKBB participants such that when the UKBB data were included in the training data, the proportions included were 10%, 33%, 67%, and 100%. However, in the cross-validation folds when UKBB cohort data were not used for training, we always used 100% of the cohort.

To investigate if this approach can be used for harmonizing cortical thickness measures, we selected the two best performing approaches from the above analysis and repeated our experiment on cortical thickness measures obtained from 68 brain regions defined by the Desikan-Killiany atlas.

Experiment 2.— We selected the two best performing models from experiment 1 and performed a stratified fivefold cross-validation across participants, stratified based on the dichotomized cognitive status. Different from experiment 1, the scanner was not used to define folds for cross-validation in experiment 2 in order to test the generalizability to new participants in seen scanners as opposed to unseen scanners tested in experiment 1. For this experiment, the proportion of the UKBB participants included was also decided based on experiment 1. To provide a reference measure, we compared the accuracies obtained with the corresponding accuracies obtained in experiment 1.

Statistical Analysis

To compute marker concordance, we compared the distributions of each volumetric measure for each pair of scanners by means of the Kolmogorov-Smirnov test with the null hypothesis that the distributions between any pair of scanners were the same. This comparison was done independently within each diagnostic group and after correcting for the confounding effects of age and sex by regressing out their effects estimated in individuals without cognitive impairment. Marker concordance was calculated as the proportion of such comparisons where there was no evidence that distributions were different between each pair of scanners across all brain regions, after controlling for multiple testing via false discovery rate (FDR ≥ 0.05) based on the *P* values of Kolmogorov-Smirnov tests. For statistical validity, we excluded scanners with fewer than 10 participants of the same diagnostic group from this evaluation.

AUCs of classification tasks for each harmonization strategy were compared with the AUCs in the case of nonharmonized data separately for each feature by means of a DeLong test.

The nonparametric McNemar χ^2 test was used to compare concordances across harmonization strategies. To control for multiple hypothesis testing, resulting *P* values were used to estimate the FDR. Statistical analyses were performed using the stattsmodel package (version 0.13.2) implemented in Python version 3.10.9.

Data and Code Availability

Amsterdam Dementia Cohort data can be made available to academic researchers upon reasonable request. ADNI and Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing data are managed by the Laboratory of Neuroimaging at the University of Southern California and are available to the

Table 2: Participant Demographics

	No. of Participants			Sex*		Diagnosis*				
Data Cohort	Processed	Considered after Removing Outliers	Age (y)*	Female	Male	HC	SCD	MCI	AD	No. of Unique Scanners*
ADC	4086	3722	63.9 ± 9.2	1717 (46.1)	2005 (53.9)	0	1355 (36.4)	805 (21.6)	1562 (42.0)	12
ADNI	2044	1830	72.2 ± 7.06	889 (48.6)	941 (51.4)	687 (37.5)	0	851 (46.5)	292 (16.0)	27
AIBL	557	524	72.7 ± 6.5	299 (57.1)	225 (42.9)	388 (74.0)	0	83 (15.8)	53 (10.1)	3
ARWiBo	913	831	56.3 ± 16.2	529 (63.7)	302 (36.3)	603 (72.6)	16 (1.9)	116 (14.0)	96 (11.6)	7
EDSD	416	384	70.4 ± 7.3	197 (51.3)	187 (48.7)	143 (37.2)	0	119 (31.0)	122 (31.8)	8
HuBA	121	116	62.4 ± 6.9	68 (58.6)	48 (41.4)	116 (100)	0	0	0	1
I-ADNI	179	172	72.2 ± 8.0	106 (61.6)	66 (38.4)	2 (1.2)	5 (2.9)	35 (20.3)	130 (75.6)	4
NACC	1861	1731	71.9 ± 9.8	910 (52.6)	821 (47.4)	0	0	949 (54.8)	782 (45.2)	22
OASIS	373	359	73.2 ± 10.7	233 (64.9)	126 (35.1)	211 (58.8)	0	111 (30.9)	37 (10.3)	1
PharmaCog	141	137	69.0 ± 7.3	80 (58.4)	57 (41.6)	0	0	137 (100)	0	7
UKBB	12 259	11 058	63.5 ± 7.6	6083 (55.0)	4975 (45.0)	11 058 (100)	0	0	0	1
Total	22 950	20 864	65.3 ± 9.4	11 111 (53.3)	9753 (46.7)	13 208 (63.3)	1376 (6.6)	3206 (15.4)	3074 (14.7)	43

Note.—Data are numbers of participants with percentages in parentheses or means ± SDs. AD = Alzheimer disease, ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's Disease Neuroimaging Initiative, AIBL = Australian Imaging, Biomarker and Lifestyle, ARWiBo = Alzheimer's Repository Without Borders, EDSD = European DTI Study on Dementia, HC = healthy control, HuBA = Hungarian Longitudinal Study of Healthy Brain Aging, I-ADNI = Italian Alzheimer's Disease Neuroimaging Initiative, MCI = mild cognitive impairment, NACC = National Alzheimer's Coordination Center, OASIS = Open Access Series of Imaging Studies, PharmaCog = European Alzheimer's Disease Neuroimaging Initiative, SCD = subjective cognitive decline, UKBB = UK Biobank.

* Values were calculated after removing the outliers, as described in the Image Processing section.

general scientific community for download *(http://ida.loni.usc.edu/)*. Alzheimer's Repository Without Borders, European DTI Study on Dementia, Italian Alzheimer's Disease Neuroimaging Initiative, Open Access Series of Imaging Studies, and PharmaCog data are available for all researchers on the NeuGRID2 platform *(https://www.neugrid2.eu/*, *https://doi.org/10.17616/R31NJN1E)*. Hungarian Longitudinal Study of Healthy Brain Aging data can be made available upon reasonable request. National Alzheimer's Coordination Center data are available through the National Alzheimer's Coordinating Center platform *(https://naccdata.org/*). UKBB data are available at the UK Biobank platform *(https://www.ukbiobank.ac.uk/)*.

Docker container source code for FreeSurfer and the MRIQC tool is available on GitHub *(https://github.com/E-DADS/freesurfer*, *https://github.com/E-DADS/mriqc)*. Multiclass Neuroharmony harmonization algorithm is available on GitHub *(https://github.com/88vikram/Multiclass-Neuroharmony)*. Trained model files for harmonization using multiclass Neuroharmony are available for all researchers on the NeuGRID2 platform *(https://www.neugrid2.eu/index.php/edads_harmonization)*.

Results

Participants

Table 2 shows descriptive statistics for the combined study sample used in our experiments, which consisted of volumetric data that passed quality control from 20 864 participants (mean age, 65.3 years ± 9.4 [SD]; 11 111 [53.3%] women, 9753 [46.7%] men) from 43 scanners across 11 cohorts. A total of 2086 individuals were excluded based on a low Euler number. Figure 1 shows age distributions by scanner and cognitive group.

Model Evaluation

Figure 2 shows the first result of experiment 1: marker concordance under cross-validation, independently for each diagnostic group and with increasing proportions of the UKBB dataset. Reference concordances for nonharmonized data are also shown for each diagnostic group for comparison. As expected, concordances for each harmonization strategy were significantly higher than the nonharmonized data for all the diagnostic groups (FDR < 0.001; *P* < .001). The use of the inclusive and multiclass models significantly improved the concordance with respect to the normative model for the diagnostic categories of MCI and AD (FDR < 0.001; *P* < .001). For diagnostic groups of HCs and SCD, the concordance of the multiclass model was significantly higher than the normative model with 100% of UKBB included (11 058 of 11 058) (HC: FDR = 0.01; *P* = .009; SCD: FDR = 0.02; *P* = .02). There was no evidence of a difference in concordance for HCs and participants with SCD between the inclusive model and normative model (FDR = 0.23; *P* = .21).

Figure 3 shows the second result of experiment 1: the AUCs for classifying HCs versus participants with AD, which were computed independently for each brain regional volume in the test set. Removing scanner-related differences decreased AUC for all harmonization approaches, potentially due to the significant imbalance (*P* < .0001) in the number of HCs and participants with AD in the different scanners (Table S3). For the normative

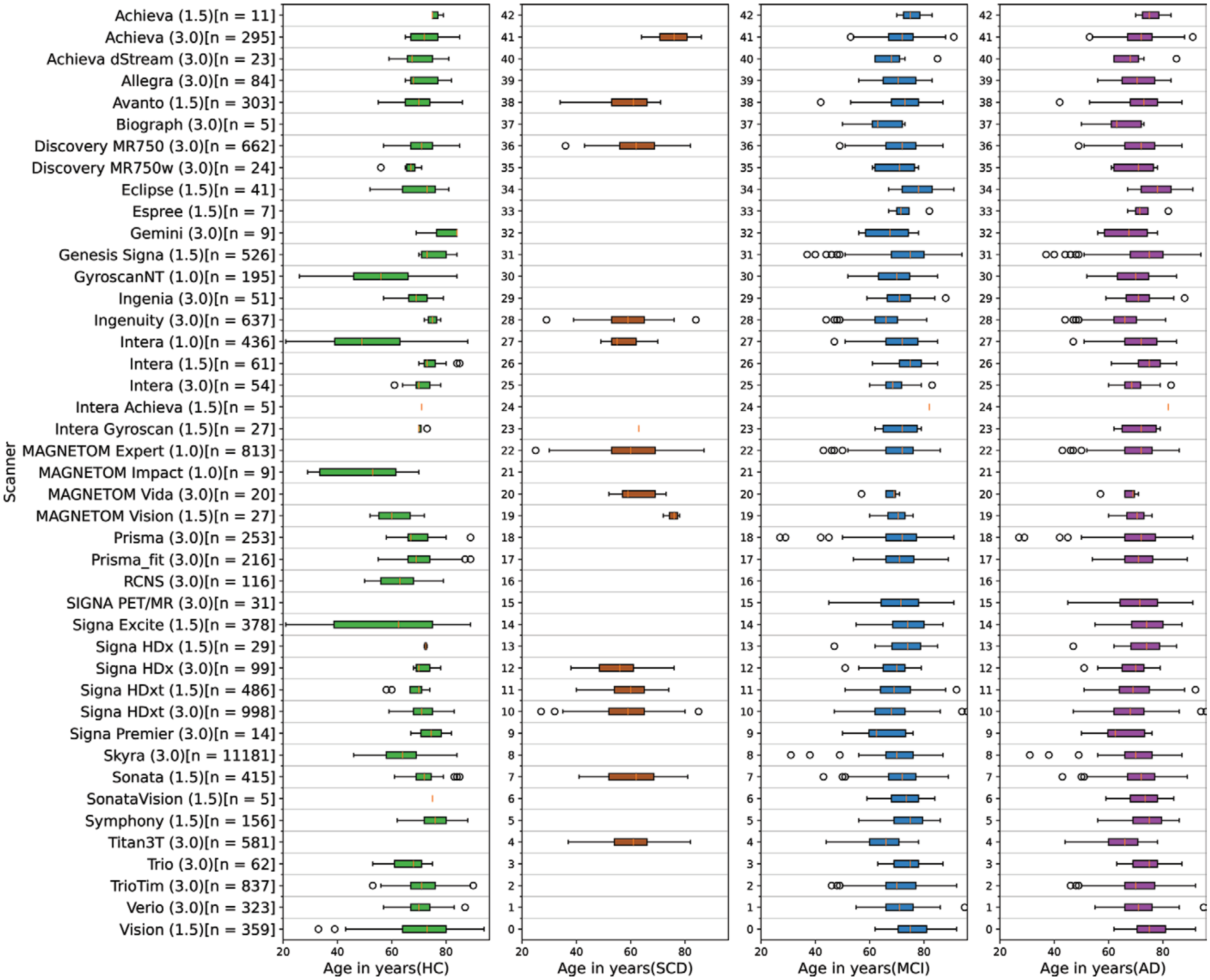

Figure 1: Box plots show age distributions by diagnosis for each scanner in the training cohort. Boxes represent individuals between the first and third quartiles, orange lines inside the boxes represent the medians, whiskers represent individuals above the third quartile and below the first quartile, and circles indicate age outliers. AD = Alzheimer disease, HC = healthy control, MCI = mild cognitive impairment, SCD = subjective cognitive decline.

model, the AUC was significantly lower than the preharmonized data for 45 volumetric features (ΔAUC = -0.013 ± 0.023). For the inclusive model, the AUC was significantly lower than the preharmonized data for 82 features (ΔAUC = -0.091 ± 0.036). For the multiclass model, the AUC was significantly lower than the preharmonized AUC for 40 features (ΔAUC = -0.006 ± 0.027), indicating relative loss of disease-related signal when using the inclusive model harmonization strategy. Across all brain regions, the best AUCs were achieved for the amygdalae and hippocampi in all harmonization scenarios (Fig S1).

Based on marker concordance and AUC, the two best models were the normative and the multiclass Neuroharmony models when trained with 100% UKBB data. For harmonizing cortical thickness measures, our proposed multiclass model achieved significantly higher marker concordance than the normative model (Fig S2).

Harmonization in Seen versus Unseen MRI Scanners

Figure 4 shows the results of experiment 2: marker concordance for seen versus unseen scanners during model training for both the normative model and multiclass model. Figure S3 shows these results for each brain volume individually. Marker concordance of the multiclass model was significantly higher than the normative model for unseen scanners for all diagnostic categories (HC: FDR = 0.01; *P* = .009; SCD: FDR = 0.02; *P* = .02; MCI: FDR < 0.001; *P* < .001; AD: FDR < 0.001; *P* < .001). For seen scanners, the multiclass model harmonization strategy significantly outperformed the normative model for the diagnostic groups of HCs, MCI, and AD (FDR < 0.001; *P* < .001) but significantly underperformed for SCD (FDR = 0.02; *P* = .02). Marker concordance using the multiclass model in a seen scanner (concordance > 0.97) was better for all diagnostic groups than in unseen scanners (concordance < 0.79).

Discussion

We introduced a novel extension of the Neuroharmony harmonization model (15) to train a generalizable machine learning model for harmonizing multicentric brain volumetric data for quantitative assessment of AD. The data for these evaluation

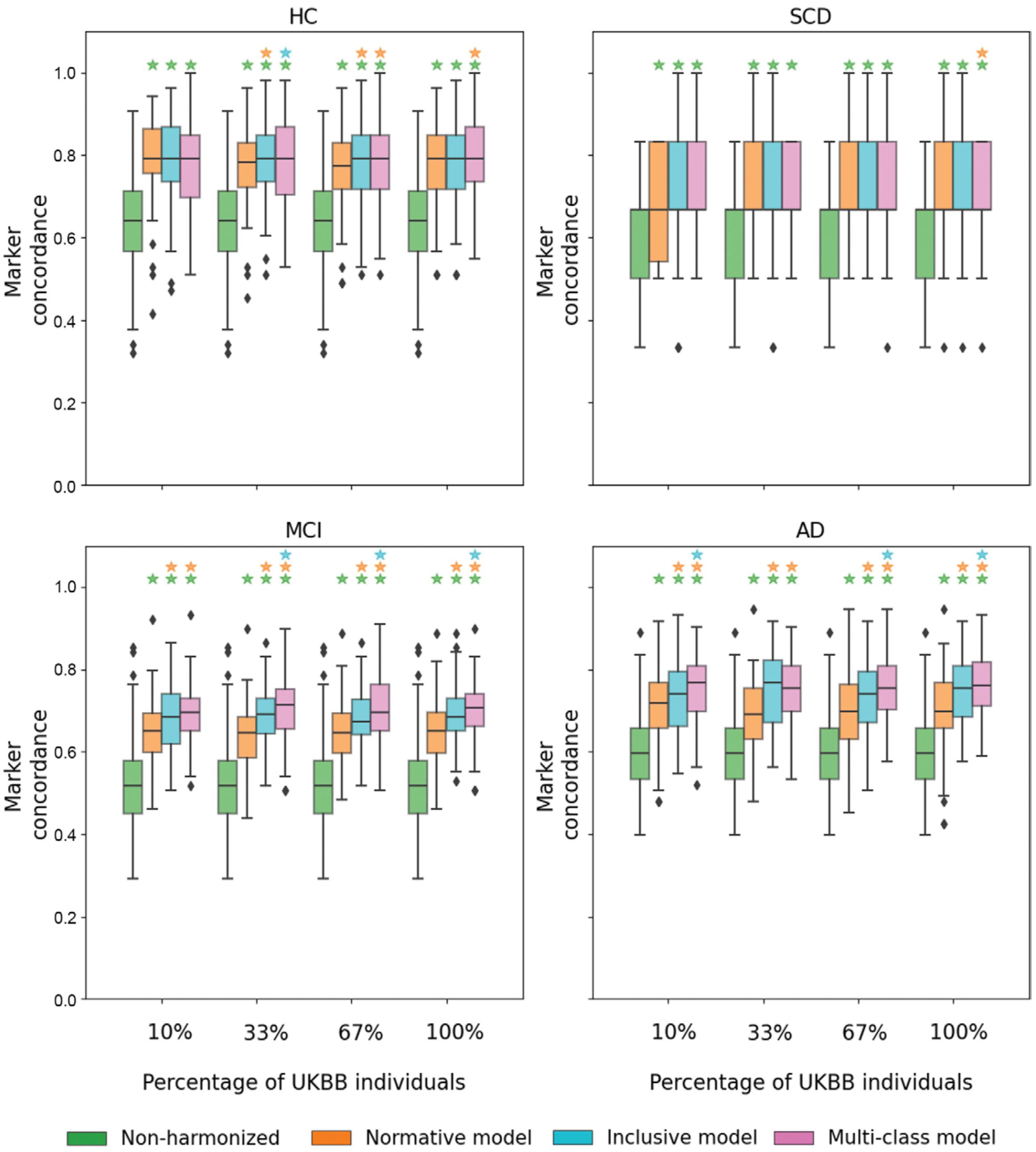

Figure 2: Experiment 1: Box plots of marker concordance for brain volumes on unseen scanners using different harmonization strategies. Concordance for nonharmonized data is also shown as a reference measure for comparison. In each diagnostic class, colored stars on top of the bars indicate statistically significant differences (false discovery rate < 0.05) between the model where the bar is located and the model indicated by the color of the star. Boxes represent individuals between the first and third quartiles, black lines inside the boxes represent the medians, whiskers represent individuals above the third quartile and below the first quartile, and diamonds indicate concordance outliers. AD = Alzheimer disease, HC = healthy control, MCI = mild cognitive impairment, SCD = subjective cognitive decline, UKBB = UK Biobank.

experiments were derived from T1-weighted three-dimensional MRI scans acquired with 43 different scanners from 20 864 participants spanning 11 cohorts. Our experiments showed that the multiclass model, which accounts for the interaction between disease pathology and image quality metrics to remove scanner-related effects, significantly improved marker concordance between scanner pairs for participants in unseen scanners as compared with normative modeling for all diagnostic groups (HC: FDR = 0.01; SCD: FDR = 0.02; MCI: FDR < 0.001; AD: FDR < 0.001). For seen scanners, it improved the marker concordance for all diagnostic groups except SCD, potentially due to the lower sample size of the SCD group or uncertainty in the etiology of this diagnostic category. Additionally, we showed that the multiclass model of Neuroharmony preserves disease-related signal during harmonization better than the other tested approaches that represent state-of-the-art methods. The newly introduced multiclass model would be helpful in harmonizing volumetric data while using automated tools in clinics and research where there could be data from new scanners not included in training.

However, we note that the AUC was slightly reduced compared with nonharmonized data for some brain regions, implying that multiclass Neuroharmony can remove some disease-related signal in the presence of diagnostic class imbalance across scanners.

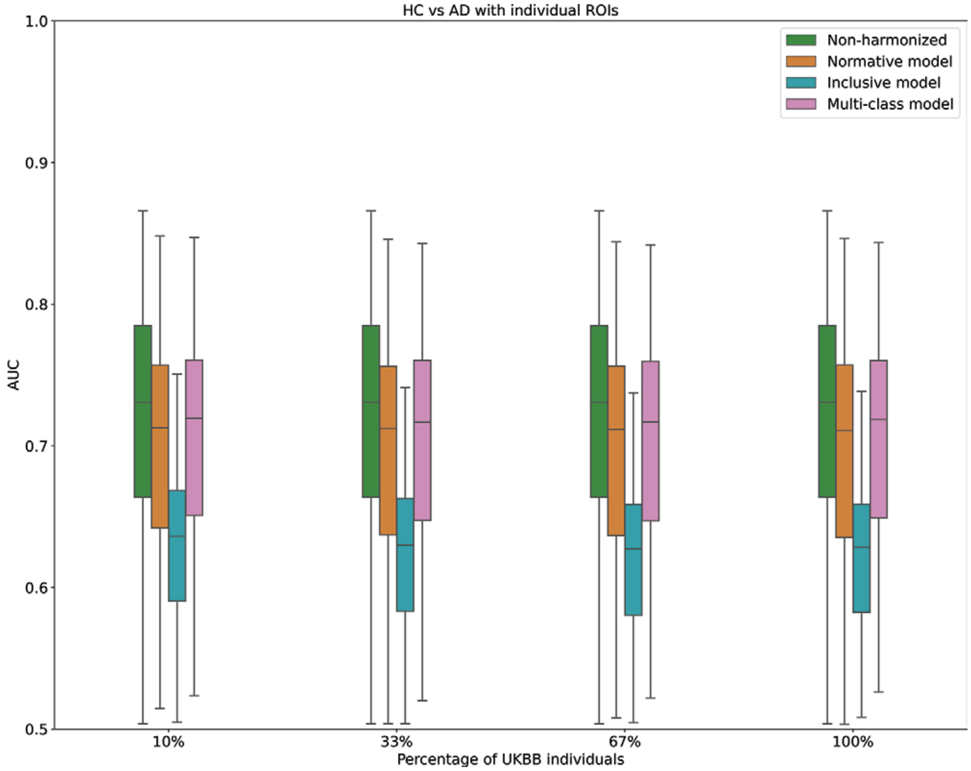

Figure 3: Experiment 1: Box plots of area under the receiver operating characteristic curves (AUCs) for distinguishing healthy controls from participants with Alzheimer disease (AD) in the test set based on the 86 brain regions of interest (ROIs) considered before and after harmonization. Boxes represent individuals between the first and third quartiles, black lines inside the boxes represent the medians, and whiskers represent individuals above the third quartile and below the first quartile. HC = healthy control, UKBB = UK Biobank.

Future work should explore model-based mechanisms for disentangling such associations to preserve disease-related signal.

Harmonization of marker data from unseen scanners remains a challenge: Marker concordance for both normative and multiclass models in unseen scanners was lower than in seen scanners. Although this leaves room for further method improvements to harmonization strategies for unseen scanners, it would also be useful to investigate if the achieved harmonization performance is sufficient for the generalizability of machine learning approaches such as classification, subtyping (33), and brain aging.

The different number of participants used to train the respective models could potentially bias the results against the model that uses a smaller dataset for training (normative model). However, we believe that this setting is a realistic and fair comparison because normative modeling always discards data from individuals with cognitive impairment. Through our modifications to the Neuroharmony model, we provided a way to include individuals with and without cognitive impairment in the training data, and our experiments showed improved harmonization in both seen and unseen scanners while preserving disease-related signal.

The harmonization performance obtained with the normative model in our experiments was lower than reported in the original Neuroharmony article (15). This difference may be due to removal of sex and age variability in the original Neuroharmony method. We preserved these effects, retaining this biologic variability, which we would argue is important for both research studies and clinical implementation.

There are challenges in the clinical implementation of the harmonization strategy. Although the multiclass model outperformed the normative model in terms of marker concordance, the implementation of the model in memory clinics might require additional work to include cognitive status of a patient during regular radiologic work-up. Machine learning models could potentially be used to overcome this limitation, as it has been shown in recent studies that classifying cognitive impairment from HC or SCD can be done with high accuracy using MRI (34). To avoid a circular dependency between the two tasks, developing multitask machine learning models to jointly harmonize and predict cognitive status is an important avenue of future work. Also, for broader use in memory clinics, the harmonization algorithm should be validated on other segmentation algorithms beyond FreeSurfer.

Although the current work was focused on the AD spectrum, we expect that our new method will be valuable for impaired cognition in general (eg, vascular dementia, frontotemporal dementia, dementia with Lewy bodies). We expect the approach to also be applicable for patients with psychiatric disorders, but further work would be needed for patients with other neurologic conditions—especially those in which the brain is affected by large lesions and other major structural modifications.

Some limitations of the original Neuroharmony model (15) apply to this work as well. The harmonization performance for an individual in the test set depends on the contrast-to-noise ratio in the T1-weighted three-dimensional MRI, and the pipeline cannot guarantee effective harmonization if the ratio is outside the range seen in our training data and might lead to incorrect harmonization. Second, the harmonization performance based on marker concordance across scanner pairs is a surrogate measure

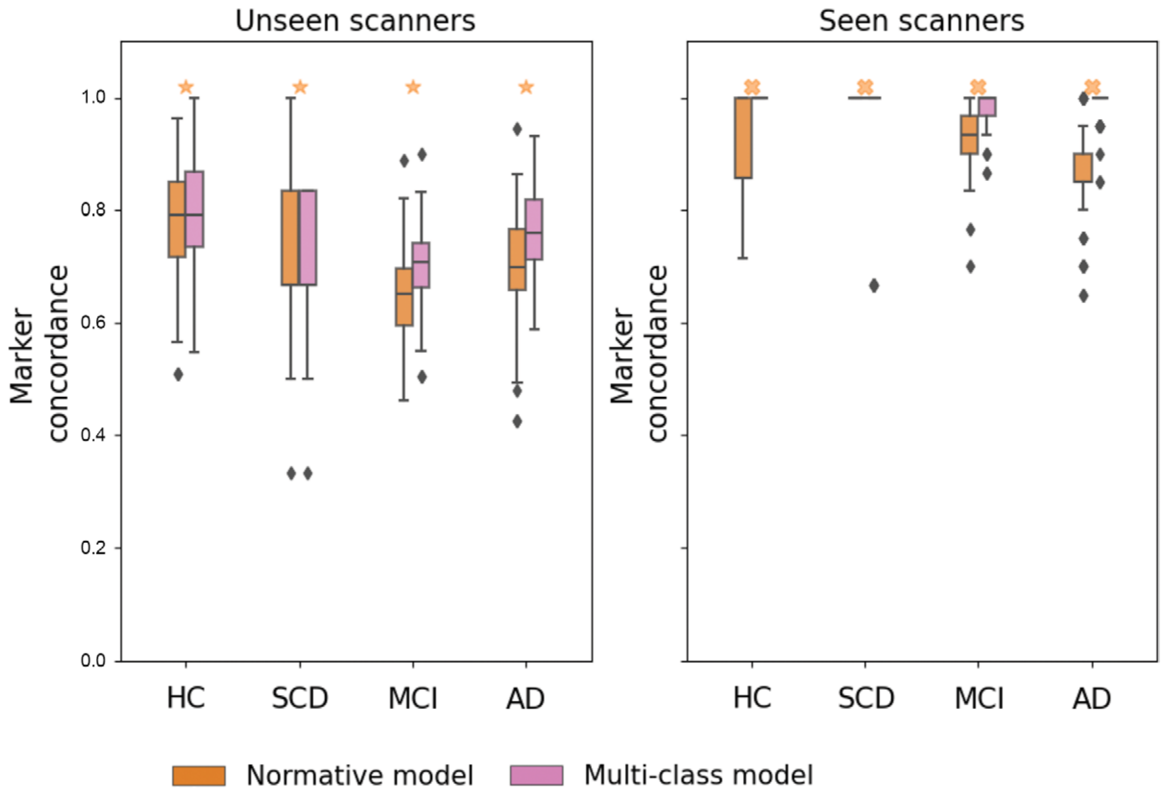

Figure 4: Experiment 2: Box plots of marker concordance for brain volumes on unseen versus seen scanners using a normative model and multiclass model. For each diagnostic class, crosses on top of bars indicate statistically significant differences (false discovery rate < 0.05) between marker concordances of normative and multiclass model in seen scanners, whereas stars on top of bars indicate statistically significant differences between marker concordances of normative and multiclass model in unseen scanners. Boxes represent individuals between the first and third quartiles, black lines inside the boxes represent the medians, whiskers represent individuals above the third quartile and below the first quartile, and diamonds indicate concordance outliers. AD = Alzheimer disease, HC = healthy control, MCI = mild cognitive impairment, SCD = subjective cognitive decline.

to measure consistency in the absence of a reference standard. A potential limitation of this study is the lack of a study to assess within-participant variability across scanners (ie, when a group of participants, including all diagnostic classes, are scanned across multiple scanners). This study would allow for evaluation of the ability of the model to remove scanner effects at the individual level, but such a study would face considerable ethical issues related to repeatedly scanning patients. Another potential issue of the present work is the definition of marker concordance, which may not be statistically robust as it implies that the failure of rejection of the null hypothesis (ie, failure to state that marker distributions are significantly different) corresponds to the null hypothesis being true (ie, marker distributions are similar), and consistency of future works may benefit from more apt definitions of marker concordance. An important limitation of this study, as with most research studies in this field, is that the imaging data used predominantly came from the developed Western countries of the European Union, United States, United Kingdom, and Australia. A more generalizable and inclusive model for harmonization would require data from nations in South America, Asia, and Africa. This model would include low-field-strength scanners that are predominantly used in these regions as well as more diverse biologic variation in the training data. Large global consortia such as the UNITED consortium (35) could potentially help in getting access to such diverse neuroimaging data. Further developing Neuroharmony for distributed or federated learning for harmonizing imaging data can also facilitate inclusion from underrepresented countries.

In summary, we have generalized the Neuroharmony model to harmonize FreeSurfer-based MRI marker data from multiple scanners and sites while retaining disease signal that could otherwise be removed by the harmonization procedure. When evaluated on brain MRI marker data from participants along the AD spectrum, our new model outperformed the other approaches we tested on both seen and unseen scanners. Further validation using different processing pipelines and evaluation criteria would be essential for clinical use of the model in applications related to cognitive decline, such as memory clinics and clinical trials of new interventions for neurodegenerative diseases.

Acknowledgments: The authors would like to thank the developers of the original Neuroharmony algorithm for making the code available and all the research participants and their families for donating their time and data for scientific research. Data collection and sharing for ADNI 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 provides funds to support ADNI clinical sites in Canada. Private sector contributions were facilitated by the Foundation for the National Institutes of Health *(www.fnih.org)*. The grantee organization was the Northern California Institute for Research and Education, and the study was coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data were disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing *(https://aibl.org.au)* is a consortium between Austin Health, CSIRO, Edith Cowan University, the Florey Institute (The University of Melbourne), and the National Ageing Research Institute. The study received partial financial support from the Alzheimer's Association (U.S.), the Alzheimer's Drug Discovery Foundation, an anonymous foundation, the Science and Industry Endowment Fund, the Dementia Collaborative Research Centres, the Victorian Government's Operational Infrastructure Support program, the Australian Alzheimer's Research Foundation, the National Health and Medical Research Council, and The Yulgilbar Foundation. Numerous commercial interactions supported data collection and analyses. In-kind support also was provided by Sir Charles Gairdner Hospital; Cogstate, Ltd; Hollywood Private Hospital; The University of Melbourne; and St Vincent's Hospital. The Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing team wishes to thank all clinicians who referred patients with Alzheimer disease and/or mild cognitive impairment to the study. We also thank all those who took part as participants in the study for their commitment and dedication to helping advance research into the early detection and causation of Alzheimer disease. We thank all the investigators within the Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing who contributed to the design and implementation of the resource and/or provided data but did not actively participate in the development, analysis, interpretation, or writing of this current study. The National Alzheimer's Coordination Center database is funded by National Institutes of Health National Institute on Aging grant U24 AG072122. National Alzheimer's Coordination Center data are contributed by National Institute on Aging–funded Aging and Disability Resource Centers: P30 AG062429 (principal investigator [PI] James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), and P30 AG072959 (PI James Leverenz, MD).

Author contributions: Guarantors of integrity of entire study, **D.A., V.V., A.R., N.P.O.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **D.A., V.V., B.W., T.A., Z.V., S.D., F.B., D.C.A., A.R., N.P.O.**; clinical studies, **F.B., A.R.**; experimental studies, **V.V., B.W., P.B., A.R.**; statistical analysis, **D.A., V.V., B.W., Z.V., S.D., A.R., N.P.O.**; and manuscript editing, all authors

Disclosures of conflicts of interest: **D.A.** No relevant relationships. **V.V.** No relevant relationships. **B.W.** No relevant relationships. **P.B.** No relevant relationships. **T.A.** No relevant relationships. **Z.V.** No relevant relationships. **S.D.** No relevant relationships. **W.M.v.d.F.** No relevant relationships. **F.B.** Consultant to Roche, Celltrion, Rewind Therapeutics, Merck, Jensen, Combinostics, and IXICO (consulting fees to author); research agreements with Merck, Biogen, GE HealthCare, Roche; participation on Data Safety Monitoring Board or Advisory Board for DSMB or SC member for Prothena, Biogen, Merck, EISAI, and ATRI/ACTC (paid to author); cofounder and shareholder in Queen Square Analytics. **D.C.A.** Cofounder and shareholder in Queen Square Analytics. **A.A.** No relevant relationships. **A.R.** No relevant relationships. **B.M.T.** No relevant relationships. **N.P.O.** Consultant for Queen Square Analytics.

References

1. Johnson KA, Fox NC, Sperling RA, Klunk WE. Brain imaging in Alzheimer disease. Cold Spring Harb Perspect Med 2012;2(4):a006213.
2. Ossenkoppele R, Cohn-Sheehy BI, La Joie R, et al. Atrophy patterns in early clinical stages across distinct phenotypes of Alzheimer's disease. Hum Brain Mapp 2015;36(11):4421–4437.
3. Hedderich DM, Dieckmeyer M, Andrisan T, et al. Normative brain volume reports may improve differential diagnosis of dementing neurodegenerative diseases in clinical practice. Eur Radiol 2020;30(5):2821–2829.
4. Goodkin O, Pemberton H, Vos SB, et al. The quantitative neuroradiology initiative framework: application to dementia. Br J Radiol 2019;92(1101):20190365.
5. Wang J, Knol MJ, Tiulpin A, et al. Gray Matter Age Prediction as a Biomarker for Risk of Dementia. Proc Natl Acad Sci USA 2019;116(42):21213–21218.
6. Young AL, Marinescu RV, Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. Nat Commun 2018;9(1):4273.
7. Liu S, Hou B, Zhang Y, et al. Inter-scanner reproducibility of brain volumetry: influence of automated brain segmentation software. BMC Neurosci 2020;21(1):35.
8. Gebre RK, Senjem ML, Raghavan S, et al. Cross-scanner harmonization methods for structural MRI may need further work: A comparison study. Neuroimage 2023;269:119912.
9. Dewey BE, Zhao C, Reinhold JC, et al. DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. Magn Reson Imaging 2019;64:160–170.
10. Fortin JP, Cullen N, Sheline YI, et al. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 2018;167:104–120.
11. Pagani E, Storelli L, Pantano P, et al. Multicenter data harmonization for regional brain atrophy and application in multiple sclerosis. J Neurol 2023;270(1):446–459.
12. Zhou HH, Singh V, Johnson SC, Wahba G; Alzheimer's Disease Neuroimaging Initiative. Statistical tests and identifiability conditions for pooling and analyzing multisite datasets. Proc Natl Acad Sci USA 2018;115(7):1481–1486.
13. Potvin O, Chouinard I, Dieumegarde L, et al. The Canadian Dementia Imaging Protocol: Harmonization validity for morphometry measurements. Neuroimage Clin 2019;24:101943.
14. Kia SM, Huijsdens H, Rutherford S, et al. Closing the life-cycle of normative modeling using federated hierarchical Bayesian regression. PLoS One 2022;17(12):e0278776.
15. Garcia-Dias R, Scarpazza C, Baecker L, et al. Neuroharmony: A new tool for harmonizing volumetric MRI data from unseen scanners. Neuroimage 2020;220:117127.
16. van der Flier WM, Pijnenburg YA, Prins N, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis 2014;41(1):313–327.
17. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27(4):685–691.
18. Ellis KA, Bush AI, Darby D, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr 2009;21(4):672–687.
19. Frisoni GB, Prestia A, Zanetti O, et al. Markers of Alzheimer's disease in a population attending a memory clinic. Alzheimers Dement 2009;5(4):307–317.
20. Brueggen K, Grothe MJ, Dyrba M, et al. The European DTI Study on Dementia - A multicenter DTI and MRI study on Alzheimer's disease and Mild Cognitive Impairment. Neuroimage 2017;144(Pt B):305–308.
21. Bankó ÉM, Weiss B, Hevesi I, et al. Study protocol of the Hungarian Longitudinal Study of Healthy Brain Aging (HuBA). Ideggyogy Sz 2024;77(1-2):51–59.
22. Cavedo E, Redolfi A, Angeloni F, et al. The Italian Alzheimer's Disease Neuroimaging Initiative (I-ADNI): validation of structural MR imaging. J Alzheimers Dis 2014;40(4):941–952.
23. Beekly DL, Ramos EM, Lee WW, et al. The National Alzheimer's Coordinating Center (NACC) database: the Uniform Data Set. Alzheimer Dis Assoc Disord 2007;21(3):249–258.
24. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 2007;19(9):1498–1507.
25. Galluzzi S, Marizzoni M, Babiloni C, et al. Clinical and biomarker profiling of prodromal Alzheimer's disease in workpackage 5 of the Innovative Medicines Initiative PharmaCog project: a 'European ADNI study'. J Intern Med 2016;279(6):576–591.
26. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779.
27. Gorgolewski KJ, Auer T, Calhoun VD, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 2016;3(1):160044.
28. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33(3):341–355.
29. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One 2017;12(9):e0184661.

30. Monereo-Sánchez J, de Jong JJA, Drenthen GS, et al. Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations - insights from the Maastricht study. Neuroimage 2021;237:118174.
31. Matelsky J, Kiar G, Johnson E, Rivera C, Toma M, Gray-Roncal W. Container-Based Clinical Solutions for Portable and Reproducible Image Analysis. J Digit Imaging 2018;31(3):315–320.
32. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–357.
33. Chen H, Young A, Oxtoby NP, et al. Transferability of Alzheimer's disease progression subtypes to an independent population cohort. Neuroimage 2023;271:120005.
34. Bron EE, Klein S, Papma JM, et al. Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease. Neuroimage Clin 2021;31:102712.
35. Adams HHH, Evans TE, Terzikhan N. The Uncovering Neurodegenerative Insights Through Ethnic Diversity consortium. The Uncovering Neurodegenerative Insights Through Ethnic Diversity consortium. Lancet Neurol 2019;18(10):915.