ARTICLE OPEN

Check for updates

Brain age prediction in generalized anxiety disorder using a convolutional neural network

Corey Richier[1,2✉], André Zugman[2], Anita Harrewijn[2,3], Elise M. Cardinale[2,4], Parmis Khosravi[2], Moji Aghajani[5], Willem B. Bruin[5], Kevin Hilbert[6], Narcis Cardoner[7,8,9], Daniel Porta-Casteràs[7], Marta Cano[7,8], Savannah Gosnell[10], Ramiro Salas[10], Andrea P. Jackowski[11], Pedro M. Pan[11], Giovanni A. Salum[12], Karina S. Blair[13], James R. Blair[14,15], Mohammed R. Milad[16], Katie L. Burkhouse[17], K. Luan Phan[18], Heidi K. Schroeder[19], Jeffrey R. Strawn[19], Katja Beesdo-Baum[20], Neda Jahanshad[21], Sophia I. Thomopoulos[21], Jared A. Nielsen[22,23,24], Jordan W. Smoller[23], Jair C. Soares[25], Benson Mwangi[25], Mon-Ju Wu[25], Giovana B. Zunta-Soares[25], Michal Assaf[26,27], Gretchen J. Diefenbach[27,28], Paolo Brambilla[29,30], Eleonora Maggioni[29], David Hofmann[31], Thomas Straube[31], Carmen Andreescu[32], Rebecca B. Price[32,33], Gisele G. Manfro[34], Federica Agosta[35,36], Elisa Canu[35], Camilla Cividini[35,36], Massimo Filippi[35,36,37,38,39], Milutin Kostić[40,41], Ana Munjiza Jovanovic[40], Brenda Benson[2], Gabrielle F. Freitag[2], Ellen Leibenluft[2], Grace V. Ringlein[2,42], Kathryn Werwath[43], Hannah Zwiebel[2], Hans J. Grabe[44,45], Sandra Van der Auwera[44,45], Katharina Wittfeld[44], Henry Völzke[46], Robin Bülow[47], Nicholas L. Balderston[48], Monique Ernst[49], Lilianne R. Mujica-Parodi[50], Helena van Nieuwenhuizen[51], Hugo D. Critchley[52], Elena Makovac[53], Matteo Mancini[54], Frances Meeten[55], Cristina Ottaviani[56,57], Gregory A. Fonzo[58], Martin P. Paulus[59], Murray B. Stein[60], Raquel E. Gur[61], Ruben C. Gur[61], Antonia N. Kaczkurkin[62], Bart Larsen[61,63], Theodore D. Satterthwaite[61], Jennifer Harper[64], Michael T. Perino[64], Chad M. Sylvester[64], Qiongru Yu[64], Patrick McClure[65], Francisco Pereira[65], Ulrike Lueken[66,67], Dick J. Veltman[68], Paul M. Thompson[21], Nynke A. Groenewold[69], Janna Marie Bas-Hoogendam[70,71,72], Dan J. Stein[69,73], Nic J. A. Van der Wee[70,72], Anderson M. Winkler[2,74], Daniel S. Pine[2] and Chelsea K. Sawyers[2,75]

© The Author(s) 2026

Higher predicted brain age difference has been associated with several psychiatric disorders. Generalized anxiety disorder (GAD) is associated with markers of accelerated aging. In this study, we determined brain predicted age difference (PAD) in individuals with GAD and healthy controls (HC) as well as group differences in PAD variability using voxel-wise structural MRI. The training dataset included 3511 controls, and the testing dataset included 1595 individuals with GAD and 4552 HC from the ENIGMA-Anxiety GAD Working Group. A convolutional neural network model using four input modalities per subject and a model ensemble approach was used to predict brain age. The PAD was then calculated by subtracting chronological age from the predicted age. Model performance was consistent with other image-based brain age prediction models with similar accuracy across the training set (mean absolute error (MAE) = 2.95 years) and HC in the testing set (MAE = 2.94). We found no evidence of accelerated brain aging in individuals with GAD compared to individuals without GAD, though we did find evidence for greater variation in PAD for individuals with GAD (Levene's test: $W = 442.98$, $p < 0.001$) and evidence for greater variability in PAD of those with GAD over 25 years of age. In several exploratory analyses, we found that symptom severity related significantly to PAD, even after controlling for medication and comorbid diagnoses, echoing previous brain age research. These findings underscore the need for consideration of heterogeneity and dimensionality of psychopathology when examining brain age predicted differences.

Translational Psychiatry (2026)16:368 ; https://doi.org/10.1038/s41398-026-04078-3

INTRODUCTION

Generalized anxiety disorder (GAD) is one of the most prevalent psychiatric disorders, having a lifetime prevalence of 6.2% [1–3]. GAD has a peak onset in late adolescence and early adulthood and high comorbidity rate [2–4]. In this study, we leverage available large imaging datasets to investigate whether GAD is associated with an individual's predicted "brain age". Brain age modeling uses neuroimaging data to estimate an individual's chronological age, with deviations between predicted and actual age—referred to as the predicted age difference (PAD)—increasingly regarded as potential biomarkers of atypical neurodevelopment or neurodegeneration [5, 6]. These brain aging estimates may proceed faster or slower than would be expected. A positive predicted age difference (PAD) may indicate accelerated maturation, or an "older" brain than expected given a chronological age; conversely, a negative PAD may indicate a "younger" brain. Such investigations would be particularly relevant for GAD as individuals with this condition appear to have structural, functional, and cognitive differences [7–10], many of which occurring at critical phases of brain development, such as late childhood and adolescence

A full list of author affiliations appears at the end of the paper.

Received: 10 June 2025 Revised: 28 March 2026 Accepted: 30 April 2026
Published online: 24 May 2026

Higher brain PAD has been associated with a wide range of physical and psychological conditions including cognitive impairment [11, 12], traumatic brain injury [13], schizophrenia [14, 15], childhood adversity [16], and major depressive disorder [17]. More broadly, psychiatric disorders are associated with advanced brain age [17–20]. However, findings related to brain age and anxiety disorders are inconsistent. For example, one study found no relationship between dimensional measures of anxiety in youth, but no significant PAD [21]; however another study found a positive PAD of 2.91 years in adults (N = 220) with anxiety disorders, including GAD, panic disorder (PD), and social anxiety disorder (SAD) [19]. Interestingly, this finding was only significant after accounting for medication use. More recently, in a study of individuals with SAD and GAD, adolescents with more severe GAD showed a higher PAD compared to controls, and adolescents with more severe SAD symptoms showed a lower PAD [22]. To date, no work has examined brain age in individuals with GAD as a distinct group.

Given unclear evidence regarding the association of brain age and anxiety disorders, we aim to evaluate PAD and assess its relation to the presence of GAD. To that end, we use a convolutional neural network (CNN) model to predict the age of each individual from their structural magnetic resonance images (MRI) and compute their PADs. CNNs are deep learning models for computer vision inspired by the structure and function of the visual cortex [23], and are commonly used to examine brain age [24–27] as they are well suited for multidimensional data. These models capture spatial relations and features in their input data and are sensitive to subject-level spatial information. This can be a strength relative to univariate models (voxelwise, vertex-wise, or feature-based), as CNNs can extract additional structural information from the image, encoded in spatial patterns. The nonlinear nature of deep learning models (including CNN) allows for detection of complex relations between the input data and the subject's age. Moreover, in healthy individuals, CNN models produce more accurate age estimations than other modeling frameworks, such as linear models or other non-neural network machine learning methods [28].

Our study has two main hypotheses: first, that PAD would be greater in participants with GAD compared to healthy controls (HC); and that larger variance in PAD in GAD. Second, participants with GAD might have a greater prediction error by the CNN, thus leading to increased variance in the PAD for this group. Other brain age research has found greater variability in age predictions for clinical groups relative to individuals without psychopathology [17, 19, 29]. To assess the second hypothesis, we conduct a test to verify the equality of variance between subjects with and without GAD. This could reflect greater heterogeneity in individuals with GAD potentially derived from factors such as treatment and disease severity while controls would form a more homogeneous group. Likewise, this variance might change as a factor of age [30, 31]. Therefore, we also hypothesize that variance in age prediction will depend on age. As secondary goals we investigate if PAD will be related to medication, comorbidity, and age-by-diagnosis interaction. To address these questions, we use structural brain MRI data from the Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) Anxiety Working Group [32, 33]. This was the largest study to date investigating structural alterations between GAD and controls, and found no association between GAD and brain morphometry in a well-powered mega-analysis of 28 sites [33]. The present study builds on that work to examine if "brain age" might reveal structural differences in GAD that were not detected in the mass univariate analysis of the previous work. As PAD is estimated from a multivariate model (CNN), it may have more sensitivity than mass-univariate testing as done in the previous ENIGMA-Anxiety GAD work. Moreover, previous work on other cohorts have shown different results when using "brain age" estimated from features obtained from MRI [21]. Previous work on brain PAD both within other ENIGMA working groups [17] and outside of ENIGMA [21, 22] have used a variety of machine learning methods to build and train brain-age prediction models. Based on the model accuracy findings from Jonnson et al. [34], showing the superiority of a model combining different measures obtained by MRI T1-weighted scans, we build a new voxelwise CNN model rather than via feature extraction (e.g., from regions of interest or cortical parcellations) to limit the loss of information.

METHODS

Sample selection

The present analysis is a pre-registered (https://osf.io/4wtdy) mega-analysis [35] of structural brain MRI data from an international multi-site sample. Our training sample of 3511 individuals includes several publicly available neuroimaging samples of both pediatric and adult healthy controls, including the Human Connectome Project [36], Open Access Series of Imaging Studies 3 (OASIS-3) [37], Philadelphia Neurodevelopmental Cohort (PNC) [38], and the Pediatric Imaging, Neurocognition, and Genetics (PING) study [39]. Training samples were selected as some of the largest publicly available samples with a representative age distribution. The present analysis only used images from a single time point in studies with longitudinal data. OASIS-3 [37] images were randomly selected across participants across all three timepoints to provide a wider distribution of ages within the sample.

The clinical test sample consists of data from the ENIGMA Anxiety Working Group [35] with 6147 subjects from 32 studies. This includes data from the Adolescent Brain Cognitive Development (ABCD) [40], Child-Mind Institute Healthy Brain Network (CMI) [41] and Brazilian High Risk Cohort (BHRC) [42] as well as data provided from ENIGMA Anxiety participating sites. As above, we only used a single time point for each study. In the case of ABCD and BHRC we only include the baseline data. Individuals with GAD had either a lifetime or present diagnosis of GAD. These diagnoses were established at each site via clinician-based assessment [32]. Information regarding comorbidities and medication usage for subjects in the present analysis are included in Supplemental Tables 1 and 2, respectively.

The selection process for determining the final number of images used in the model is summarized in Supplemental Table 3, which illustrates the procedure for inclusion of participants. We excluded subjects due to failed processing, missing covariates (age, sex, diagnosis), comorbid psychosis, bipolar disorder or ASD, and subjects with no GAD diagnosis but any other psychiatric diagnosis.

Neuroimaging data preprocessing

All data were run through the FSL voxel-based morphometry (VBM) pipeline [43, 44], providing voxelwise output [45]. Such output is amenable for use with CNNs, while still capturing between-subject variability on brain morphology. Structural T1-weighted MRI brain images were non-linearly aligned to a standard brain (MNI152) [46], bias-corrected, subjected to removal of non-brain tissue, segmented into gray and white matter maps, modulated, and smoothed. It has been demonstrated that an increased number of imaging modalities results in better prediction [34]; thus, for this analysis we used four images derived from the VBM pipeline: grey matter, white matter, T1-weighted image, and the image of the Jacobian of the non-linear transformation used to map the individual T1-weighted image to the reference standard space [47]; the latter provides a measure of stretches and shrinkages needed to non-linearly align each subject's image to the standard space.

No data harmonization step was implemented (e.g., ComBat; [48]) due to recent demonstrations that not applying site/scanner harmonization produces lower mean absolute error (MAE) for brain age prediction than applying harmonization methods [49]. After the VBM process was completed, images were converted to NumPy arrays [50] to increase speed of subsequent image loading, and were cropped from their original dimension of $91 \times 109 \times 91$ to $75 \times 88 \times 78$ to reduce the amount of empty space surrounding the brain tissue.

Quality control

Quality control procedures consisted of visual inspection to exclude any images that were misshapen or contained any other processing errors. Additionally, subjects who had errors in their covariate entries were excluded. After quality control, the remaining usable sample consisted of

Table 1. Number of Subjects and Descriptive Statistics - Training Sample.

Site	Total n	Average Age	Standard Deviation	Age Min	Age Max
HCP	412	28.71	3.67	22	37
HCP Development	612	14.52	4.04	5.58	21.92
OASIS	334	68.93	9.29	43	93
PING	588	12.42	5.12	3.08	21.67
PNC	1565	15.02	3.66	8.08	23.08
Total/Average	3511	21.23	16.85	3.08	93

HCP human connectome project, *HCP* development human connectome project- development, *OASIS* open access series of imaging studies, *PING* pediatric imaging, neurocognition, and genetics study. *PNC* philadelphia neurodevelopmental cohort.

Table 2. Number of Participants and Descriptive Statistics - Testing Sample.

Site	Total n	Age Min	Age Max	HC Sample	GAD Sample	GAD Average Age (SD)	HC Average Age (SD)	*p*
ABCD	2857	8.91	11	2778	79	10.05 (0.62)	9.92 (0.62)	0.07
Barcelona	69	18	40	46	23	23.61 (4.3)	21.71 (4.48)	0.10
Baylor	226	12	79	128	98	28.84 (9.18)	24.39 (12.84)	**<0.05**
BHRC	767	7.02	21.37	458	309	13.38 (3.15)	11.70 (2.64)	**<0.05**
Boystown	95	13	18	45	50	15.82 (1.35)	15.44 (1.62)	0.23
Chicago-Milad	43	18	59	16	27	30.25 (9.71)	32.37 (12.96)	0.58
Chicago-Phan	141	18	60	41	100	26.36 (8.51)	25.29 (9.75)	0.54
Cincinnati	20	12	18	11	9	13.66 (2.05)	14.18 (1.89)	0.59
CMI-HBN	185	5.01	21.36	97	88	12.24 (3.53)	9.57 (3.21)	**<0.05**
Dresden	93	18	51	46	47	30.23 (9.94)	28.47 (7.90)	0.35
Duke	40	6	10	19	21	7 (1.06)	7.47 (0.99)	0.16
SHIP	32	41	70	22	10	55.9 (7.95)	57.72 (8.01)	0.57
Harvard	237	18	40	51	186	24.64 (4.31)	24.68 (3.77)	0.95
Houston	263	8	68.23	251	12	16.69 (9.02)	29.21 (14.39)	**<0.05**
Milan	91	20.89	72.57	61	30	44.31 (14.91)	34.11 (12.42)	**<0.05**
Muenster	53	19	56	29	24	29.45 (10.01)	27.82 (9.04)	0.54
Pittsburgh-Andreescu	64	19	82	26	38	57.78 (16.51)	56.84 (20.22)	0.84
Pittsburgh-Price	69	18	54	0	69	N/A	N/A	N/A
PROTAIA	43	13	22	25	18	16.67 (2.26)	17.80 (2.54)	0.14
Rome	38	18	55	20	18	29.55 (6.92)	29.15 (9.18)	0.88
San Raffaele	89	23	63	69	20	44.65 (10.61)	45.26 (10.97)	0.83
SDAN	295	8.1	51.14	156	139	12.77 (2.78)	13.70 (4.37)	**<0.05**
SNFA	60	19	50	38	22	31.27 (9.97)	29 (7.17)	0.36
Stony Brook	60	18	49	19	41	22.87 (5.91)	21.57 (5.74)	0.43
UCSD	91	17	53	47	44	28.36 (10.75)	23.40 (8.27)	**<0.05**
Wash U	63	8	12	32	31	9.96 (1.44)	9.90 (1.15)	0.85
IOL	63	19	71	21	42	39.92 (13.38)	41.19 (16.17)	0.76
Total/Average	6147	15.66	46.58	4552	1595	22.16 (13)	14.53 (10.59)	**<0.05**

GAD generalized anxiety disorder, *HC* healthy controls, *p* p value, *ABCD* adolescent brain cognitive development study, *BHRCS* brazilian high risk cohort study, *CMI-HBN* child mind institute healthy brain network, *IOL* institute of living, *HCP* human connectome project, *HCPdev* human connectome project development, *PING* pediatric imaging, neurocognition, and genetics study, *OASIS* open access series of imaging studies, *PNC* philadelphia neurodevelopmental cohort, *PROTAIA* anxiety disorders program for child and adolescent psychiatry, *SDAN* section on development and affective neuroscience, *SNFA* section on neurobiology of fear and anxiety, *UCSD* university of California – San Diego, *WashU* Washington University. *SHIP* study of health in pomerania.

8063 HC and 1595 participants with a diagnosis of GAD, generating a total sample size of 9658. Tables 1 and 2 detail the number of participants included from each sample for the training and testing samples.

Train-test split procedure

For the test and train split procedure, the total number of subjects (9658) was split into two groups: (a) subjects from samples including GAD cases and controls, and (b) subjects from samples composed of only controls. The latter was used as the training sample, which consisted of 3511 individuals (Table 1). Ten percent of the training sample was used to track model performance while training (referred in the results as validation sample). The test sample (n = 6147; Table 2) consisted of data from studies that contained subjects with GAD and controls.

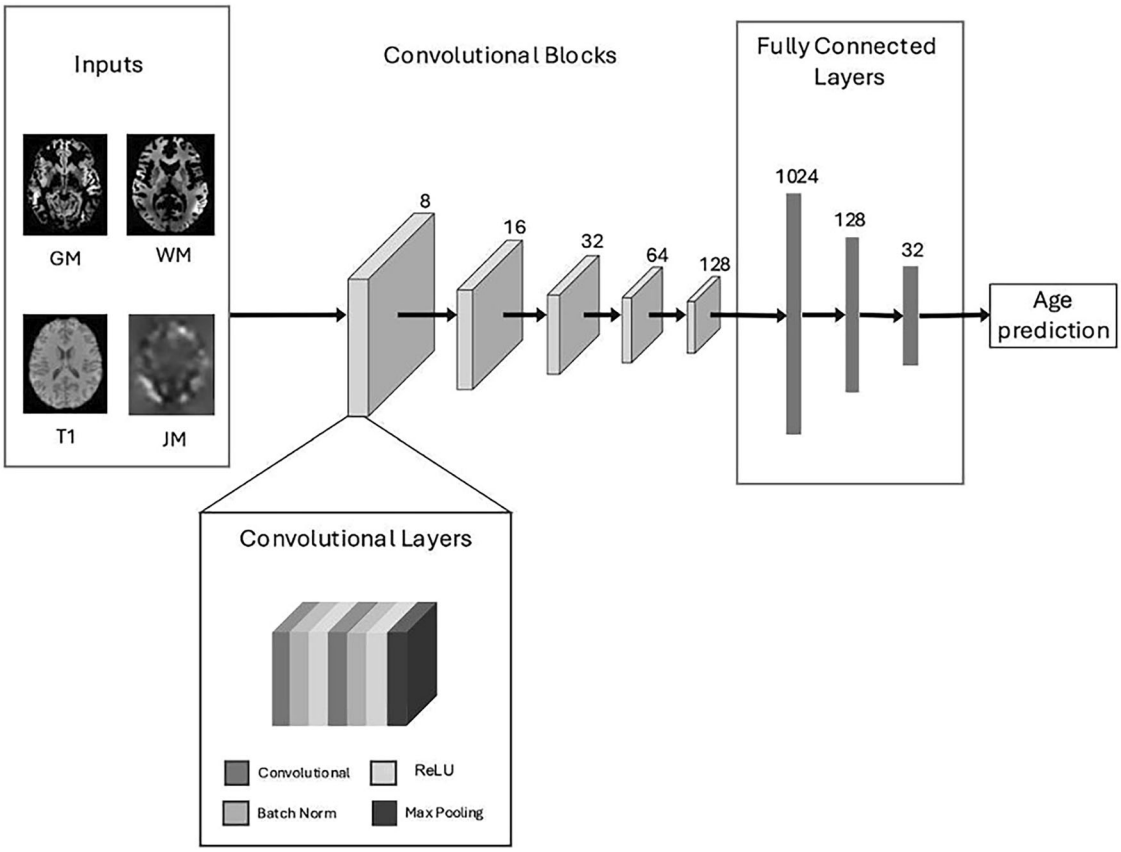


Fig. 1 Overview of the CNN Model Architecture. *Note*. The CNN architecture was designed with four input channels for gray and white matter segmentations, as well as the T1-weighted and Jacobian images. The model consists of five repeated convolutional blocks, each containing 3D convolutional layers, batch normalization, ReLU activations and max pooling. The number of filters used in each block are denoted above. After the last convolutional block, the model includes a flattened layer.

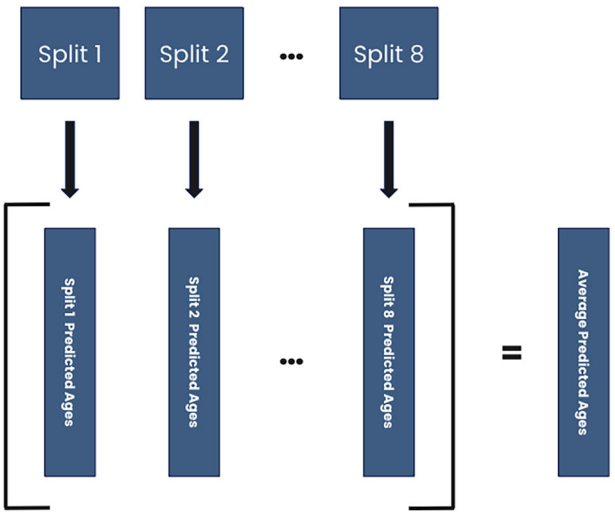


Fig. 2 Model Ensemble Procedure. *Note*. Procedure for generating average predictions for the model ensemble. Eight splits of the training data were created, which used a different 90/10% split of training to validation data. Each of these splits was given to a unique instantiation of the model with identical architecture but instantiated with random weights.

Brain age model

The present analysis used the PyTorch library for the creation of the CNN model [51]. The input to the network was a 5-dimensional array consisting of batch size (the number of subjects given to the model at a time for training) by channels by X image dimension by Y image dimension by Z image dimension. The four channels were the gray and white matter segmentations, as well as the T1-weighted image and Jacobian images, all in MNI152 space.

The network architecture included five repeated convolutional blocks, each containing 3D convolutional layers, batch normalization, ReLU activations and max pooling (Fig. 1). This block was repeated five times, at each step doubling the number of convolutional filters but halving the size of the image via the max pooling layer. Thus, the number of filters was 8 for the first layer, and the final number of filters was 128. We used a $3 \times 3 \times 3$ kernel with padding of size 1. Max pooling layers had a stride distance of 2 to downsize each dimension image by half. Layers following batch normalization utilized a ReLU activation function. After the last block, the image was flattened to a 1024 size fully connected dense layer and then connected to two more fully connected layers of size 128 and 32 before connecting to the output node where age predictions were made with a ReLU function. Between each fully connected layer, a dropout layer was added with a rate of 50 percent, as is recommended for CNNs [52]. The total number of model parameters was 1,020,721.

The MAE was measured across each batch to track model performance and as the model loss function. The mean squared error (MSE) was also tracked but not used to optimize model performance. The Adaptive Moment Estimation (Adam) optimizer [53] was used to update the model weights. Gradient norm clipping was used to prevent exploding gradients [54], which occurs when the gradients of the model parameters become excessively large during training, leading to instability. The maximum gradient value was set to 1 and learning rate was set to 0.0001.

A homogeneous ensemble strategy was used to combine the predictions of multiple models [55]. This procedure is illustrated in Fig. 2. Other machine learning prediction work has used ensemble strategies with deep learning models [55, 56], including brain age prediction [57], which has been found to generate better predictive results. Since neural networks may converge to several possible local minima, it is ideal to use strategies that provide stable results. Instability in results would

complicate downstream analysis and interpretation of results generated by a CNN. As is recommended when using homogeneous ensembles, it is best practice to introduce randomness into the constituent models, usually in the form of a diversified training set [55]. An ensemble of eight model instantiations was created for the present analysis. Each instantiation was initialized with random weights, but the architecture for each was identical. Ten percent of the training data was randomly selected to serve as the validation data for each model, such that each model in the ensemble had a different set of 90% of the full training sample. After the training process, the eight models' age predictions for each subject in the test set were then averaged. These averaged predictions were treated as the overall predicted ages to be used in subsequent analyses and subtracted from the subject's true chronological age to generate the overall PAD for all participants within the entire test set.

Statistical analysis

All statistical tests were conducted in PALM - Permutation Analysis of Linear Models [58]. The PAD measure was entered as the dependent variable for all models. Six general linear models (GLMs) were tested. The first three of these models were preregistered (Supplemental Table 4): one was a simple model with only the diagnosis as the variable of interest, the second was a model with diagnosis and covariates (age, sex, scanner, comorbidities, and medication), and a third model with the same covariates, but with the interaction between diagnosis and age as the variable of interest.

Although not initially planned, we conducted three additional analyses during peer-review (Supplemental Table 5). The models for these additional analyses only included subjects with a GAD diagnosis. The first of these models examined if comorbidity had a significant effect on PAD, with age, sex, scanner and medication as covariates. The remaining two models examined if clinical severity, as obtained from questionnaire data, had a significant relationship to PAD. The first of these models included age, scanner, and sex as covariates, whereas the second included these variables as well as comorbidity and medication.

Since different studies had assessed severity in subjects using distinct measures, three sets of procedures enabled analyses that incorporated these distinct measures into a statistical model. First, all subjects with GAD who had any available questionnaire data were selected. The questionnaires included in the present analysis were the Generalized Anxiety Disorder 7-item Scale [59], the Hamilton Anxiety Rating Scale [60], the Penn State Worry Questionnaire [61], the State-Trait Anxiety Inventory [62], and the Screen for Child Anxiety Related Disorders parent and child GAD scales and total scores [63]. Second, each scale was rescaled with a linear interpolation to numbers between 0, which represented the lowest possible score on the questionnaire, and 1, which represented the maximum score on the questionnaire. A score of 0.5 was treated as the respective clinical cutoff for each questionnaire. Finally, for participants that had more than one questionnaire, their scores were first interpolated and then averaged together. These interpolated scores were then used as an independent variable for the model.

All models included scanner as a covariate and allowed each site to have their own parameter estimates (slopes and intercepts), thus accommodating between-site differences, further dispensing with the need for explicit harmonization, and avoiding that larger sites would dominate the results. Variance groups [58, 64] were used to allow different variances between GAD and controls. The number of permutations (sign-flippings) was set at 10,000. All results were corrected for multiple testing using the familywise error rate (FWER).

Levene's test for equal variance was used to investigate if participants with GAD have a higher variance in the PAD than HCs. Levene's test was also used to examine the difference across diagnosis and in adults compared to children and adolescents (as preregistered with a threshold of 25 years. This was chosen a priori as an appropriate age for the transition from adolescence to young adulthood [11, 65].

We utilized SHAP (SHapley Additive exPlanation; [66, 67]) to explore which brain regions were associated with age predictions. The GradientExplainer module was utilized to interrogate which voxels were most meaningful for age predictions. To obtain Shapley values, which demonstrate the relative contributions of each voxel to predictions, we first constructed a single model whose output was characterized by the mean prediction across all the different models in the ensemble. Shapley values are calculated from the gradients in the ensembled model, using the full set of training data as a reference distribution. For each subject, voxelwise Shapley values were calculated for each of the four input channels. This analysis was not part of the preregistration as it was added during the review of this manuscript.

RESULTS

The MAE for the training set was 2.95 years. The MAE for the validation set was 2.30 years. Figure 3 displays a scatterplot of the predicted ages with respect to the true ages in the testing set. Supplemental Fig. 1 illustrates the predictions of ages broken up across four quartiles of age. Averaged across models in the ensemble, the testing set had a MAE of 3.80 years, 2.94 years for control participants, and 6.27 years for subjects with a GAD diagnosis (Supplemental Table 6). A key consideration was to establish the extent to which each model generated consistent predictions and evaluation metrics. Testing results across models revealed stable performance. This was determined by using intraclass correlation (ICC) of the form [2,K] to measure the reliability of the age predictions [68].

To test the reliability of each model's generated age prediction, ICC was calculated across these predictions. The ICC for the predicted age was 0.99, indicating excellent reliability across models. Table 3 displays the results of the MAE and PAD across samples. The MSE for the overall sample was 47.77, whereas the MSE for the control group was 31.72 and for the GAD group 93.57. Once the average brain age predictions were subtracted from chronological age, the overall PAD was 0.49 years, with a standard deviation of 6.10. For controls, the predicted age difference was 0.36 years (SD = 4.97). For GAD subjects, the PAD was 0.87 years (SD = 8.53). True age and predicted age significantly correlated ($r = 0.86$, $p < 0.001$), as did true age and PAD ($r = -0.26$, $p < 0.001$). The greater MAE of subjects with GAD relative to controls was noteworthy.

None of the preregistered three GLM models that examined the difference between subjects with and without a GAD diagnosis established a statistically significant relationship between PAD and diagnosis (all $p > 0.05$). Three models tested during peer review examined only subjects with GAD. One model, examining whether PAD differed between subjects with or without comorbidities, was nonsignificant. However, both models examining the clinical severity of GAD were significant. The first model, which included age, sex, and scanner as covariates was found to be significant (Aspin-Welch's $v = 6.49$, $p < 0.001$), as was the case for the model that included these variables as well as comorbidity and medication usage (Aspin-Welch's $v = 3.37$, $p < 0.05$). Supplemental Fig. 2 illustrates the relationship between predicted age and symptom severity.

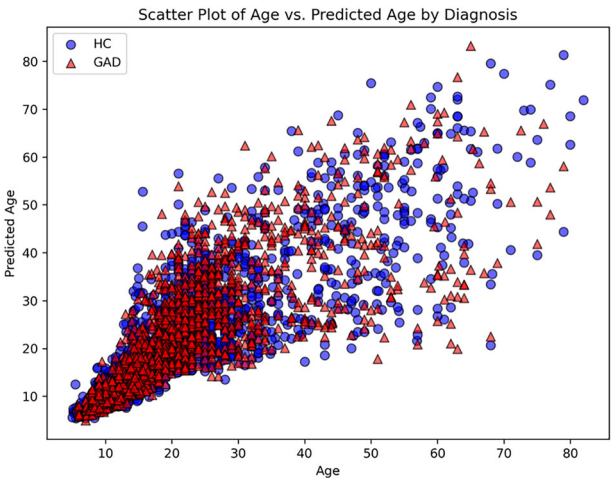


Fig. 3 Scatter Plot of True Age vs. Predicted Age Sorted By Diagnostic Status.

Table 3. Testing Set Age, Predicted Age, and PAD Organized by Sample.

	GAD n	GAD PAD Mean	GAD PAD SD	HC n	HC PAD Mean	HC PAD SD
ABCD	79	0.179	1.18	2778	0.58	1.19
Barcelona	23	−1.45	6.03	46	−1.83	4.79
Baylor	98	5.36	7.71	128	1.06	5.40
BHRC	309	1.60	4.92	458	0.85	4.53
Boystown	50	−2.21	2.18	45	−2.92	1.72
Chicago-Milad	27	4.31	8.52	16	4.89	8.97
Chicago-Phan	100	0.94	5.95	41	1.11	7.38
Cincinnati	9	−1.14	0.96	11	−0.88	1.29
CMI	88	2.22	4.03	97	1.20	2.32
Dresden	47	9.23	7.01	46	10.92	7.41
Duke	21	0.23	1.33	19	−0.48	0.91
SHIP	10	6.25	6.03	22	1.69	7.17
Harvard	186	4.76	6.09	51	4.43	5.36
Houston	12	−0.2.67	5.44	251	−4.61	8.16
Milan	30	−7.04	9.98	61	−8.46	7.63
Muenster	24	12.1	8.72	29	9.04	8.96
Pittsburgh-Andreescu	38	−23.27	11.46	26	−19.83	12.25
Pittsburgh-Price	69	−8.51	5.82	N/A	N/A	N/A
PROTAIA	18	6.51	5.90	25	8.72	8.64
Rome	18	5.24	7.49	20	3.72	7.63
San Raffaele	20	11.65	4.99	69	5.41	8.72
SDAN	139	−0.19	1.42	156	0.14	2.41
SNFA	22	1.58	8.56	38	−0.81	9.93
Stony Brook	41	9.26	7.46	19	12.76	6.74
UCSD	44	−3.09	9.28	47	−0.99	6.33
Wash U	31	2.42	1.13	32	2.47	1.21
IOL	42	−14.80	8.64	21	−13.17	9.31

PAD predicted age difference, *SD* standard deviation, *GAD* generalized anxiety disorder, *HC* healthy controls, *ABCD* adolescent brain cognitive development study, *BHRCS* brazilian high risk cohort study, *CMI-HBN* child mind institute healthy brain network, *IOL* institute of living, *HCP* human connectome project, *HCPdev* human connectome project development, *PING* pediatric imaging, neurocognition, and genetics study, *OASIS* open access series of imaging studies, *PNC* Philadelphia neurodevelopmental cohort, *PROTAIA* anxiety disorders program for child and adolescent psychiatry, *SDAN* section on development and affective neuroscience, *SNFA* section on neurobiology of fear and anxiety, *UCSD* University of California – San Diego, *WashU* Washington university. *SHIP* study of health in Pomerania.

Levene's test of equality of variances suggests a statistically significant difference in the variance of PAD between controls and subjects with GAD ($W = 442.98$, $p < 0.001$), with GAD subjects presenting more variance compared to healthy controls (GAD SD: 8.53, controls SD: 4.97). To examine if age plays a role in the degree of variance between those with and without GAD, the test was implemented after partitioning the data into four groups: healthy volunteers under age 25, individuals with GAD under 25, healthy volunteers over age 25, and individuals with GAD over 25. Subjects with GAD had greater variance, with those over 25 having the greatest variance for all groups. Full results for these tests are shown in Tables 4 and 5. A violin plot illustrating the degree of variance in each group is displayed in Fig. 4.

Two-sample t-tests between the mean ages of individuals with and without GAD were conducted between all sites to examine any significant differences. Several sites were found to have statistically significant differences in age (listed in Table 2). This difference was also significant in the overall sample. To mitigate these age differences, we conducted several supplementary analyses, with only the ENIGMA-Anxiety sites (excluding ABCD, CMI and BHRC). A t-test demonstrated that once these sites were removed, both subjects with and without GAD were age matched. Of note, age predictions conducted by the testing set are

Table 4. Characteristics of PAD by Diagnosis and Age.

Group	Mean PAD	Median PAD	Standard Deviation	Variance
GAD < 25	1.86	0.42	5.25	27.64
HC < 25	0.77	0.51	3.01	9.06
GAD > 25	−1.42	0.10	13.08	171.136
HC > 25	−2.86	−2.70	11.57	134.04

Table 5. Levene's Tests Comparing Variances in PAD across Diagnostic and Age Groups.

Group 1	Group 2	W	P
GAD < 25	HC < 25	348.52	<0.001
GAD < 25	GAD > 25	489.6	<0.001
GAD < 25	HC > 25	394.01	<0.001
HC < 25	GAD > 25	2541.49	<0.001
HC < 25	HC > 25	2221.09	<0.001
GAD > 25	HC > 25	6.74	<0.001

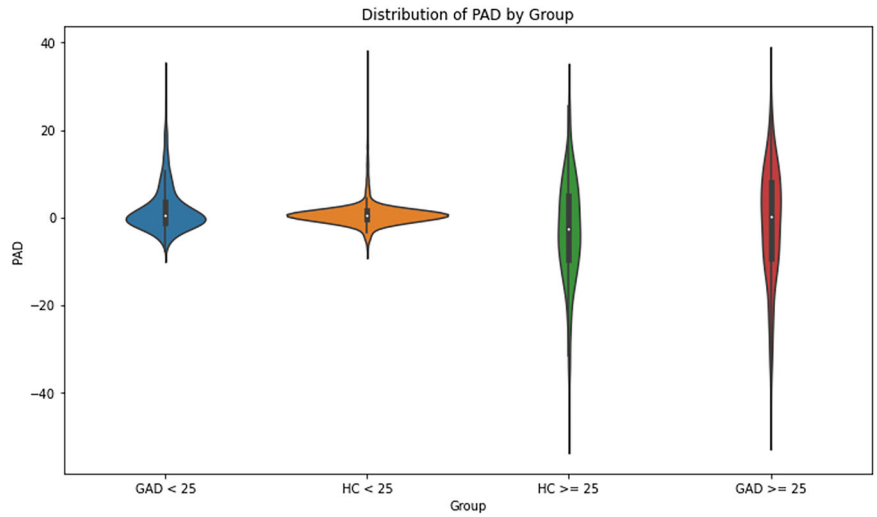


Fig. 4 Violin Plot of Variance in PAD across Age and Diagnostic Groups.

independent—meaning that samples included in the testing set do not influence the age predictions of other samples, any downstream analyses using PAD at the group level can still be influenced by sites with skewed distributions. Once these sites are removed, it removes the potential for any age distribution biases in these supplemental analyses. We reanalyzed the same three GLM models with the large age discrepancies removed. As before, none of these models established a statistically significant relationship between PAD and diagnosis (all $p > 0.05$). These sites were also removed for the set of models examining only individuals with GAD, where significant results were obtained. Additionally, we also re-examined the variance in PAD across age and diagnostic groups. The effect of greater variance in subjects with GAD, and greater variance in GAD subjects above 25 was replicated (See Supplemental Tables 7 and 8).

To explore the extent to which comorbidity may drive variability, we partitioned the individuals with GAD into those with any comorbidity and those with only a GAD diagnosis. Next, we conducted another equality of variances test. The results of this test n suggested a significant difference in variance between groups ($W = 6.33$, $p < 0.05$). Those patients with GAD only had a greater standard deviation ($SD = 9.82$) compared to those with a comorbid disorder ($SD = 7.77$).

Shapley values were used to create visualizations of which voxels were most important for age predictions across the different modalities. Subjects with and without GAD had their respective input images averaged. The top 10% of values in the average images were retained to aid in visualization. Figures 5 and 6 demonstrate the visualization for each image for both individuals with GAD and without a GAD diagnosis.

The values obtained from SHAP were small (range of 0 - 0.00005 on average across modalities in the top 10% of most important features) which reflects that no single voxel or region had a majority of influence in the prediction of age. Identified voxels across the four image modalities were quite diffuse. Furthermore, the voxels implicated in SHAP did not drastically differ between the HC and GAD groups. This may reflect the lack of difference in PAD at the total group level. While diffuse in location across modalities, the maps however do implicate the top 10% voxels largely overlap in regions of the insula, brainstem, cerebellum, parietal lobe and occipital lobe.

DISCUSSION

The current study used a CNN to estimate brain age and compute the PAD across individuals with GAD and controls. We used data from more than 3500 healthy individuals across the lifespan to train our model and then tested the model on more than 6000 independent individuals with and without GAD. The difference in PAD between individuals with GAD and controls was then investigated. With respect to our first hypothesis, we did not find a significant difference in PAD between individuals with GAD and controls. However, consistent with our second hypothesis, we found that individuals with GAD have greater PAD variability, particularly in those older than 25 years. After the PAD for each subject was generated, we conducted secondary analyses to examine (1) the relation of diagnostic status across sites, (2) how GAD severity, medication, comorbid disorders related to PAD, and (3) interaction of diagnosis by age. All three secondary preregistered analyses examining the difference in subjects with and without GAD showed no significant relation between the variables of interest and PAD, even after age matching subjects. There was also no effect of comorbidity on PAD in subjects with a GAD diagnosis. However, a significant effect of clinical severity was found in exploratory analyses, even after controlling for comorbidity and medication usage.

Our findings are consistent with prior work showing no significant differences in PAD between GAD and controls and with dimensional measures of anxiety in youth [22]. However, these findings contrast with some work within anxiety disorders. In general, some previous work found positive PAD in substantially smaller samples of participants with GAD, PD and SAD diagnoses, when accounting for medication effect [19], or a greater PAD in adolescents with GAD [22]. The heterogeneity of GAD in its clinical presentation, may explain portions of these inconsistent findings.

While a group difference between subjects with and without a GAD diagnosis were not found, results from exploratory analyses suggest that clinical severity relates to PAD. Namely, higher severity related to higher PAD in subjects in the GAD group, suggesting advanced biological aging in more severely ill individuals. Previous research has found a relationship between severity of symptoms and increased PAD in mood and anxiety disorders [19].

Although the group difference findings were non-significant, the significantly higher variance in the GAD group is noteworthy, as heterogeneity within psychopathology is a common finding [69, 70]. Other brain age analyses of psychiatric disorders have also found greater variability in the clinical population compared to healthy controls. Han et al. [17] found higher variability among individuals with depression than in healthy individuals. For healthy controls, the MAE was 6.50 and 6.84 years for males and females, respectively. Similarly, within the MDD group, the MAE

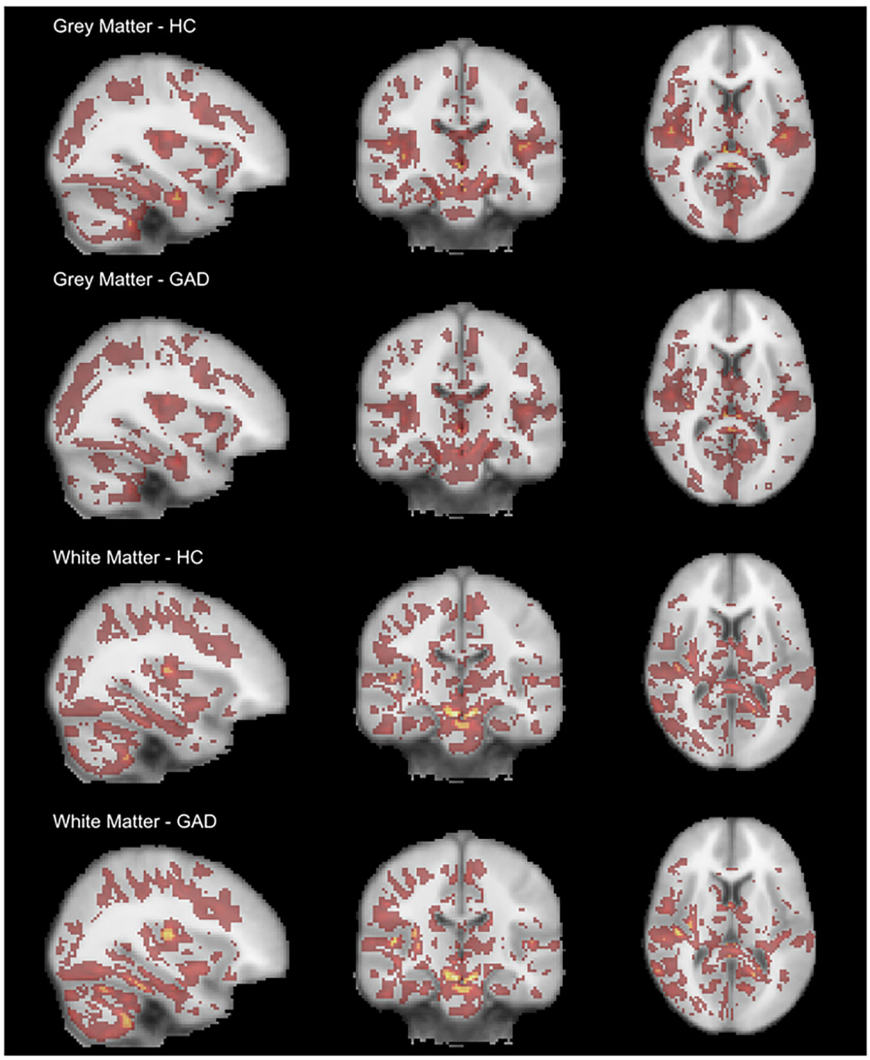


Fig. 5 Visualization of grey and white matter voxels associated with age prediction identified via SHAP. *Note*. SHAP values were derived from a gradient-based explainer applied to an ensemble of deep learning models predicting brain age from structural MRI. The above image presents values identified from the grey and white matter images the model used as input data. Absolute SHAP values were averaged across subjects within each group and thresholded to the top 10% for visualization. Maps for individuals with GAD and those without a diagnosis (healthy control - HC) are displayed on the MNI152 2 mm template with matched slice locations and color scaling. Lighter colors (i.e., more yellow) reflect higher absolute SHAP values.

was 6.72 for males and 7.18 for females. In a replication study, the same authors found that males in the control group had an MAE of 6.85 and females had 7.84. While in the depression group, males showed a larger MAE of 7.57 and females had 9.24, again suggesting greater variability within the clinical population. The same trend appeared in another study examining depression and anxiety disorders [19], where healthy controls manifested an MAE of 5.97, but those in the disorder group had an MAE of 6.73. Finally, Tønnesen et al. [29] found that healthy controls had a MAE of 6.92, patients with bipolar disorder had an MAE = 6.85, and patients with schizophrenia had an MAE = 7.11. This could suggest potentially greater variability in subjects with schizophrenia, but similar variability in bipolar disorder to healthy controls. While notable, these differences in past research were not reported as main findings, nor statistically tested. However, the present study used a statistical procedure to establish a statistically significant difference in variance across those with GAD and controls; this suggests a further avenue for research in brain age prediction for psychiatric disorders. The higher variance in the GAD group indicated that the model performs more poorly in this population. Further investigating the sources of increased variance in the patient population might help understand new mechanisms for the development of psychiatric disorders.

Given that there was a statistically significant age difference in the subjects with GAD, we conducted further supplemental analysis with the non-ENIGMA samples removed to ensure subjects were age matched. This new GLM did not suggest an effect of advanced aging or any other variables having a relationship to the PAD. However, the effect of larger variance in the PAD for individuals with GAD replicated in the age matched sample. This means that the inclusion of sites such as ABCD, where a large number of subjects with narrowly distributed ages did not drive the differences in variances between these groups. The observation of greater PAD variability in individuals with GAD, despite no overall shift in mean brain age relative to individuals without GAD, has meaningful implications, especially in the context of an effect of severity having an effect on PAD. Clinically, it underscores the heterogeneity of GAD and raises the possibility that subgroups of patients may experience neurostructural changes that are obscured when averaging across the diagnostic group (e.g., endophenotypes). This possibility, if true, would partially explain the mixed findings in the literature and raises the

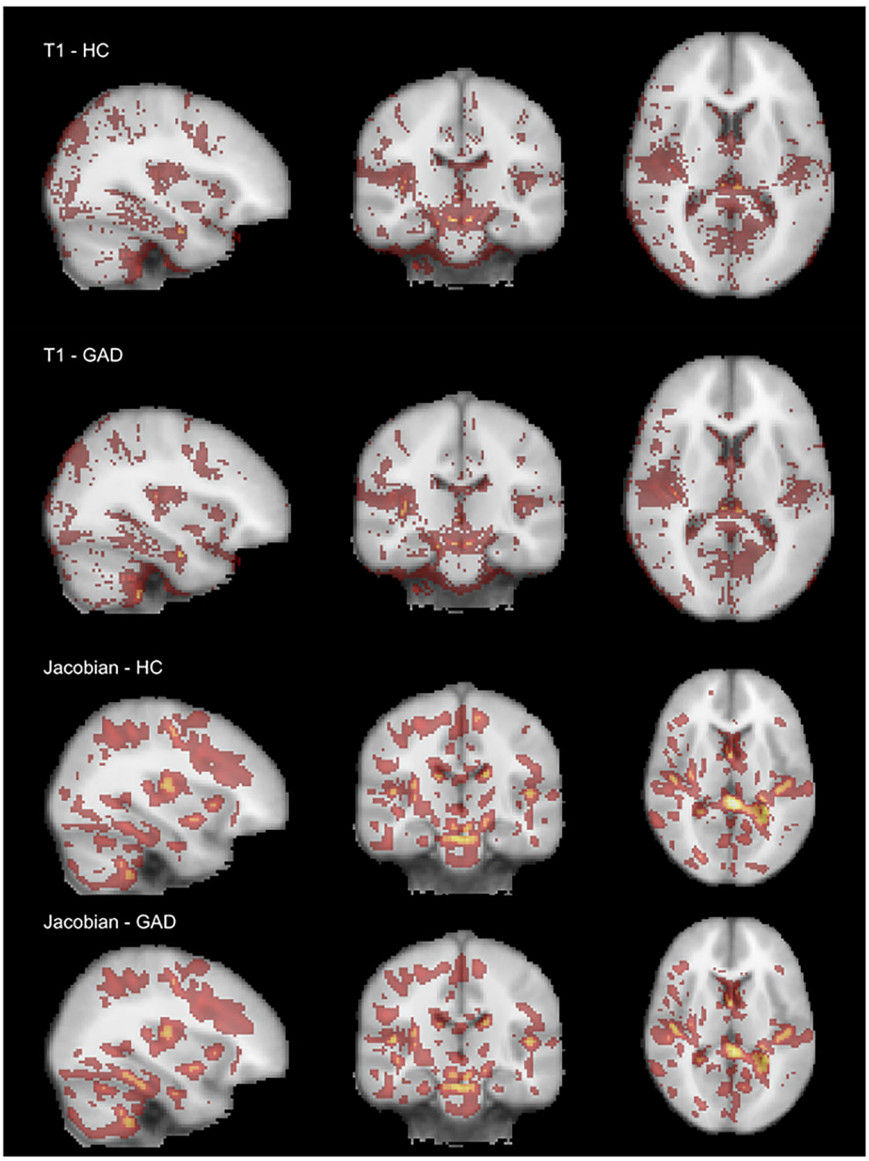


Fig. 6 Visualization of T1 and Jacobian matrix voxels associated with age prediction identified via SHAP. *Note*. SHAP values were derived from a gradient-based explainer applied to an ensemble of deep learning models predicting brain age from structural MRI. The above image presents values identified from the T1 image and the jacobian matrix the model used as input data. Absolute SHAP values were averaged across subjects within each group and thresholded to the top 10% for visualization. Maps for individuals with GAD and those without a diagnosis (healthy control - HC) are displayed on the MNI152 2 mm template with matched slice locations and color scaling. Lighter colors (i.e., more yellow) reflect higher absolute SHAP values.

possibility that PAD variance could represent a transdiagnostic marker of risk or resilience. Moreover, the finding of greater PAD variability in individuals over 25 suggests that brain aging trajectories in GAD may become more divergent in adulthood, highlighting the need for longitudinal research to map the neurostructural progression of the disorder over time. Additionally, it was found that those with only a GAD diagnosis had a greater variability in PAD than those with comorbidity diagnoses. This may be due to greater clinical severity in those with comorbid disorders [71–73]—of whom those with comorbid diagnoses may have an older average predicted age. These results should be taken into consideration with the results of the GLM examining severity in GAD subjects, in which the effect of severity was still detected (but attenuated) when controlling for comorbidity. Future studies should explore whether increased PAD variability is associated with symptom chronicity, cognitive decline, treatment response, or comorbidities.

The results of this study should be viewed in the context of several limitations. First, although this study has a large sample size, there was an uneven age distribution across the samples, with a predominance of younger participants largely due to the inclusion of the ABCD data. To address this, we conducted supplemental analyses with the large sites such as ABCD removed to only include age matched sites. Additionally, PAD was related to true age, which is a common issue in brain age studies [74, 75]; to address this we included age as a covariate, with no significant results. A second limitation is that, while we constructed our model based on the best existing literature and practices on CNN, this area of research continues to develop quickly. The proposed model does not exhaust possible architectures for brain age prediction, or even all possibilities among CNN models. It should also be noted that the use of a whole-brain VBM approach in combination with the CNN architecture does not allow for group comparisons of structural changes in specific regions of interest

(ROI) previously associated with GAD, as is common practice in the neuroimaging literature. However, the CNN approach demonstrates more accurate overall brain age estimations compared to linear ROI approaches in healthy subjects [27]. Thus, the CNN architecture is more sensitive to subject level spatial information making it a better model choice for the proposed analyses with the downside of lacking the granularity of an ROI-based linear approach.

Another limitation of this study is that the model performs declines in older individuals. This is reflected by higher MAE in both the validation and test samples with older age. However, the fact that the findings persist even when removing developmental cohorts, where many of the younger subjects are included, and providing a closer age match between the test and training set (supplemental Fig. 3) suggests that the model can be robust to these effects. We would also highlight that data from older participants also comes from ENIGMA cohorts which have been collected with more variable methodology (i.e., scanner, sequence parameters, inclusion criteria). We chose not to perform any harmonization (i.e., ComBat) as deep learning should be resilient to different data sources and there is evidence that site harmonization may even impair performance [76]. However, future studies are still needed to provide clearer guidance on CNN methods applied to neuroimaging mega-analysis. Other important limitations were that more detailed information about the clinical course of GAD, such as age of onset, and duration were not available. Further research should examine the extent to which these play a role in brain age prediction models.

Regarding the visualization of brain regions associated with age predictions, Shapley values obtained in the present study were not found to be different between those with and without a GAD diagnosis. The voxels do point to regions of the brain that might be implicated with early aging and can be the focus of further research. The voxels identified via SHAP are diffused across grey and white matter. However, these values from SHAP are not to be interpreted as causation or correlation with symptoms or severity in presentation, as they are only maps at the level of the diagnostic group.

In summary, we found no significant difference in PAD between individuals with GAD compared to healthy controls. This is consistent with prior work in adolescents with anxiety symptoms that found no association between PAD and anxiety. We found PAD was not associated with medication use or comorbid disorders related to PAD, and found no interaction between age and diagnostic status with PAD. However, in exploratory analysis undertaken during peer review, we found that there was a relationship between symptom severity and PAD only in subjects with GAD (Supplemental Fig. 2). It is possible that this effect may represent diagnostic heterogeneity present in GAD. We also found a significantly larger variability in PAD for GAD compared to the healthy controls. Additional analysis suggests that individuals with only a GAD diagnosis had a greater degree of variability in PAD than those with comorbid diagnoses in addition to GAD. Future research examining the role of brain aging in psychiatric disorders should take into consideration the possible sources of heterogeneity, either subject level characteristics (comorbidities, symptom severity) or effects of brain age prediction (i.e., variability in PAD) as they may smooth over group level differences, as the findings of the present study suggest. Additionally, comparative models of brain age of anxiety in animals could also be explored [77, 78] to establish relationships between brain aging and anxiety. Lastly, the training and testing samples assembled for this study are substantially larger than previous work within anxiety domains and brain PAD. Reporting these results from a large mega-analysis is important as it may illustrate that a more detailed approach is necessary to fully capture the heterogeneity underlying GAD. It might be that specific subpopulations, treatment exposures, or specific symptoms are associated with differences in brain age; however, it is difficult to disentangle these without harmonized data collection.

REFERENCES

1. Szuhany KL, Simon NM. Anxiety disorders: a review. JAMA. 2022;328:2431–45.
2. Wittchen H-U. Generalized anxiety disorder: prevalence, burden, and cost to society. Depress Anxiety. 2002;16:162–71.
3. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry. 2005;62:617.
4. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry. 2005;62:593.
5. Cole JH, Poudel RPK, Tsagkrasoulis D, Caan MWA, Steves C, Spector TD, et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. NeuroImage. 2017;163:115–24.
6. Smith SM, Vidaurre D, Alfaro-Almagro F, Nichols TE, Miller KL. Estimation of brain age delta from brain imaging. NeuroImage. 2019;200:528–39.
7. Andreescu C, Sheu LK, Tudorascu D, Walker S, Aizenstein H. The ages of anxiety: differences across the lifespan in the default mode network functional connectivity in generalized anxiety disorder. Int J Geriatr Psychiatry. 2013;29:704–12.
8. Brambilla P, Como G, Isola M, Taboga F, Zuliani R, Goljevscek S, et al. White-matter abnormalities in the right posterior hemisphere in generalized anxiety disorder: a diffusion imaging study. Psychol Med. 2011;42:427–34.
9. Hilbert K, Lueken U, Beesdo-Baum K. Neural structures, functioning and connectivity in Generalized Anxiety Disorder and interaction with neuroendocrine systems: a systematic review. J Affect Disord. 2014;158:114–26.
10. Sylvester CM, Corbetta M, Raichle ME, Rodebaugh TL, Schlaggar BL, Sheline YI, et al. Functional network dysfunction in anxiety and anxiety disorders. Trends Neurosci. 2012;35:527–35.
11. Franke K, Luders E, May A, Wilke M, Gaser C. Brain maturation: predicting individual BrainAGE in children and adolescents using structural MRI. NeuroImage. 2012;63:1305–12.
12. Gaser C, Franke K, Klöppel S, Koutsouleris N, Sauer H. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. PLoS ONE. 2013;8:e67346.
13. Cole JH, Leech R, Sharp DJ. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. Annals Neurol. 2015;77:571–81.
14. Koutsouleris N, Davatzikos C, Borgwardt S, Gaser C, Bottlender R, Frodl T, et al. Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders. Schizophr Bull. 2013;40:1140–53.
15. Schnack HG, van Haren NEM, Nieuwenhuis M, Hulshoff Pol HE, Cahn W, Kahn RS. Accelerated brain aging in schizophrenia: a longitudinal pattern recognition study. Am J Psychiatry. 2016;173:607–16.
16. Keding TJ, Heyn SA, Russell JD, Zhu X, Cisler J, McLaughlin KA, et al. Differential patterns of delayed emotion circuit maturation in abused girls with and without internalizing psychopathology. Am J Psychiatry. 2021;178:1026–36.
17. Han LKM, Dinga R, Hahn T, Ching CRK, Eyler LT, Aftanas L, et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. Mol Psychiatry. 2020;26:5124–39.
18. Blake KV, Ntwatwa Z, Kaufmann T, Stein DJ, Ipser JC, Groenewold NA. Advanced brain ageing in adult psychopathology: a systematic review and meta-analysis of structural MRI studies. J Psychiatric Res. 2023;157:180–91.
19. Han LKM, Schnack HG, Brouwer RM, Veltman DJ, van der Wee NJA, van Tol M-J, et al. Contributing factors to advanced brain aging in depression and anxiety disorders. Transl Psychiatry. 2021;11:402.
20. Kaufmann T, van der Meer D, Doan NT, Schwarz E, Lund MJ, Agartz I, et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. Nat Neurosci. 2019;22:1617–23.
21. Cropley VL, Tian Y, Fernando K, Mansour LS, Pantelis C, Cocchi L, et al. Brain-predicted age associates with psychopathology dimensions in youths. Biol Psychiatry Cogn Neurosci Neuroimaging. 2021;6:410–9.
22. Zhou Z, Li Y, Zhang Y, Liu J, Ai H, Liu M, et al. Differential effects of generalized anxiety and separation anxiety on brain structural development during adolescence. J Affect Disord. 2023;339:478–85.
23. Lindsay GW. Convolutional neural networks as a model of the visual system: past, present, and future. J Cogn Neurosci. 2021;33:2017–31.
24. Bashyam VM, Erus G, Doshi J, Habes M, Nasrallah IM, Truelove-Hill M, et al. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. Brain. 2020;143:2312–24.
25. Christman S, Bermudez C, Hao L, Landman BA, Boyd B, Albert K, et al. Accelerated brain aging predicts impaired cognitive performance and greater disability in geriatric but not midlife adult depression. Transl Psychiatry. 2020;10:317.
26. Mendes SL, Pinaya WHL, Pan P, Sato JR. Estimating gender and age from brain structural MRI of children and adolescents: a 3D convolutional neural network

multitask learning model. Computational Intelligence and Neuroscience. 2021;2021:5550914.
27. Ueda M, Ito K, Wu K, Sato K, Taki Y, Fukuda H, et al. An age estimation method using 3D-CNN from brain MRI images. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). p. 380-3 Venice, Italy, IEEE; 2019. https://doi.org/10.1109/ISBI.2019.8759392.
28. Leonardsen EH, Peng H, Kaufmann T, Agartz I, Andreassen OA, Celius EG, et al. Deep neural networks learn general and clinically relevant representations of the ageing brain. NeuroImage. 2022;256:119210.
29. Tønnesen S, Kaufmann T, de Lange A-MG, Richard G, Doan NT, Alnæs D, et al. Brain age prediction reveals aberrant brain white matter in schizophrenia and bipolar disorder: a multisample diffusion tensor imaging study. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020;5:1095–103.
30. Bethlehem RAI, Seidlitz J, White SR, Vogel JW, Anderson KM, Adamson C, et al. Brain charts for the human lifespan. Nature. 2022;604:525–33.
31. Ge R, Yu Y, Qi YX, Fan YN, Chen S, Gao C, et al. Normative modelling of brain morphometry across the lifespan with CentileBrain: algorithm benchmarking and model optimisation. Lancet Digit Health. 2024;6:e211–e21.
32. Bas-Hoogendam JM, Winkler AM, Pine D, Sawyers C, Stein DJ, Richier C, et al. ENIGMA-anxiety working group - Generalized Anxiety Disorder Brain Age Project. OSF Registries; 2023.
33. Harrewijn A, Cardinale EM, Groenewold NA, Bas-Hoogendam JM, Aghajani M, Hilbert K, et al. Cortical and subcortical brain structure in generalized anxiety disorder: findings from 28 research sites in the ENIGMA-Anxiety Working Group. Transl Psychiatry. 2021;11:502.
34. Jonsson BA, Bjornsdottir G, Thorgeirsson TE, Ellingsen LM, Walters GB, Gudbjartsson DF, et al. Brain age prediction using deep learning uncovers associated sequence variants. Nate Commun. 2019;10:5409.
35. Zugman A, Harrewijn A, Cardinale EM, Zwiebel H, Freitag GF, Werwath KE, et al. Mega-analysis methods in ENIGMA: The experience of the generalized anxiety disorder working group. Hum Brain Mapp. 2020;43:255–77.
36. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-minn human connectome project: an overview. NeuroImage. 2013;80:62–79.
37. LaMontagne PJ, Benzinger TLS, Morris JC, Keefe S, Hornbeck R, Xiong C, et al. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. Cold Spring Harbor Laboratory; 2019. https://doi.org/10.1101/2019.12.13.19014902.
38. Satterthwaite TD, Connolly JJ, Ruparel K, Calkins ME, Jackson C, Elliott MA, et al. The philadelphia neurodevelopmental cohort: a publicly available resource for the study of normal and abnormal brain development in youth. Neuroimage. 2016;124:1115–9.
39. Jernigan TL, Brown TT, Hagler DJ, Akshoomoff N, Bartsch H, Newman E, et al. The Pediatric Imaging, Neurocognition, and Genetics (PING) data repository. NeuroImage. 2016;124:1149–54.
40. Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. Dev Cogn Neurosci. 2018;32:43–54.
41. Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Scientific Data. 2017;4:170181.
42. Salum GA, Gadelha A, Pan PM, Moriyama TS, Graeff-Martins AS, Tamanaha AC, et al. High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results. Int J Methods Psychiatr Res. 2015;24:58–73.
43. Ashburner J, Friston KJ. Voxel-based morphometry: the methods. NeuroImage. 2000;11:805–21.
44. Douaud G, Smith S, Jenkinson M, Behrens T, Johansen-Berg H, Vickers J, et al. Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. Brain. 2007;130:2375–86.
45. Lu B, Li H-X, Chang Z-K, Li L, Chen N-X, Zhu Z-C, et al. A practical Alzheimer's disease classifier via brain imaging-based deep learning on 85,721 samples. Journal of Big Data. 2022;9:101.
46. Good CD, Johnsrude IS, Ashburner J, Henson RNA, Friston KJ, Frackowiak RSJ. A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage. 2001;14:21–36.
47. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. NeuroImage. 2011;54:313–27.
48. Fortin J-P, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. NeuroImage. 2018;167:104–20.
49. Yu Y, Cui HQ, Haas SS, New F, Sanford N, Yu K, et al. Brain-age prediction: systematic evaluation of site effects, and sample age range and size. Hum Brain Mapp. 2024;45:e26768.
50. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585:357–62.
51. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems, 8024–35.
52. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1929–58.
53. Kingma DP, Ba J Adam: a method for stochastic optimization. arXiv:1412.6980 [Preprint]. 2014. Available from: https://doi.org/10.48550/arXiv.1412.6980.
54. Pascanu, R., Mikolov, T., & Bengio, Y. On the difficulty of training recurrent neural networks. In International conference on machine learning (pp. 1310-1318). Pmlr. 2013.
55. Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: a review. Engineering Applications of Artificial Intelligence. 2022;115:105151.
56. Khosla M, Jamison K, Kuceyeski A, Sabuncu MR. Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction. NeuroImage. 2019;199:651–62.
57. Kuo C-Y, Tai T-M, Lee P-L, Tseng C-W, Chen C-Y, Chen L-K, et al. Improving individual brain age prediction using an ensemble deep learning framework. Front Psychiatry. 2021;12:626677.
58. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. NeuroImage. 2014;92:381–97.
59. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166:1092–7.
60. Maier W, Buller R, Philipp M, Heuser I. The hamilton anxiety scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. J Affect Disord. 1988;14:61–8.
61. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther. 1990;28:487–95.
62. Spielberger, C. D., Gonzalez-Reigosa, F., Martinez-Urrutia, A., Natalicio, L. F., & Natalicio, D. S. The state-trait anxiety inventory. Revista Interamericana de Psicologia/Interamerican journal of psychology, 5(3 & 4). 1971.
63. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry. 1997;36:545–53.
64. Winkler AM, Webster MA, Vidaurre D, Nichols TE, Smith SM. Multi-level block permutation. Neuroimage. 2015;123:253–68.
65. Blakemore SJ. Development of the social brain in adolescence. J R Soc Med. 2012;105:111–6.
66. Ravagnani Salto AB, Azank F, Voltolini MC, Fogaca Doretto V, de Giusti C, Jackowski AP, et al. Machine learning for anxiety diagnosis using structural MRI does not generalize to unseen data: results from a large developmental cohort. Int Rev Psychiatry. 2025;37:719–30.
67. Lundberg SM, Lee S-I A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA: Curran Associates Inc.; 2017. p. 4768–77.
68. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine. 2016;15:155–63.
69. Feczko E, Fair DA. Methods and challenges for assessing heterogeneity. Biol Psychiatry. 2020;88:9–17.
70. Olshan SM, Richier CJ, Baacke KA, Miller GA, Heller W. Making the most of errors: utilizing erroneous classifications generated by machine-learning models of neuroimaging data to capture disorder heterogeneity. Journal of Psychopathology and Clinical Science. 2024;133:678–89.
71. Klein Hofmeijer-Sevink M, Batelaan NM, van Megen HJGM, Penninx BW, Cath DC, van den Hout MA, et al. Clinical relevance of comorbidity in anxiety disorders: a report from the Netherlands Study of Depression and Anxiety (NESDA). J Affec Disord. 2012;137:106–12.
72. Kaufman J, Charney D. Comorbidity of mood and anxiety disorders. Depress Anxiety. 2000;12:69–76.
73. Angst J, Sellaro R, Ries Merikangas K. Multimorbidity of psychiatric disorders as an indicator of clinical severity. Eur Arch Psychiatry Clin Neurosci. 2002;252:147–54.
74. Liang H, Zhang F, Niu X. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. Human Brain Mapping. 2019;40:3143–52.
75. More S, Antonopoulos G, Hoffstaedter F, Caspers J, Eickhoff SB, Patil KR. Brain-age prediction: a systematic comparison of machine learning workflows. NeuroImage. 2023;270:119947.
76. Lee J, Han J, Kim M, Kim Y, Kim TS, Lee WH, editors. The impact of MRi data harmonization on brain age prediction. 2025 International Conference on

Artificial Intelligence in Information and Communication (ICAIIC); 2025 18-21 Feb. 2025.
77. Zhou J, Chen Y, Jin X, Mao W, Xiao Z, Zhang S, et al. Fusing multi-scale functional connectivity patterns via multi-branch vision transformer (MB-ViT) for macaque brain age prediction. Neural Networks. 2024;179:106592.
78. Ji GJ, Xu W, Zhang J, Li Q, Yang Y, He K, et al. Distinct antidepressant therapies act on a common brain network. Nat Commun. 2026;17:1176.

ACKNOWLEDGEMENTS

We would like to acknowledge the late Professor Dan J Stein for his longstanding leadership within ENIGMA-Anxiety and for his invaluable contributions to psychiatry, neuroscience, and mental health research. His intellectual generosity, mentorship, and unwavering support shaped this work and the broader field in lasting ways. His scientific legacy and his kindness to colleagues continue to inspire us.

AUTHOR CONTRIBUTIONS

CR, AZ: conceptualization; data curation; formal analysis; investigation; methodology; software; visualization; writing—original draft; writing—review and editing. AH, EMC, PK: data curation; project administration; writing—review and editing. EL: data curation; supervision; writing—review and editing. PMT: conceptualization; methodology; writing—review and editing; supervision. NAG, JMB-H, DJS, NJAVdW: conceptualization; writing—review and editing. AMW: conceptualization; methodology; software; formal analysis; writing—original draft; writing—review and editing; visualization; supervision. DSP: conceptualization; formal analysis; resources; writing—original draft; writing—review and editing; supervision. CKS: conceptualization; methodology; supervision; writing—review and editing. AH, EMC, PK, MAg, WBB, KH, NC, DP-C, MC, SG, RS, APJ, PMP, GAS, KSB, JRB, MRM, KLB, KLP, HKS, JRS, KB-B, NJ, SIT, JAN, JWS, JCS, BM, M-JW, GBZ-S, MAs, GJD, PB, EMa, DH, TS, CA, RBP, GGM, FA, EC, CC, MF, MK, AMJ, BB, GFF, GVR, KWa, HZ, HJG, SVdA, KWe, HV, RBu, NLB, ME, LRM-P, HvN, HDC, EMk, MM, FM, CO, GAF, MPP, MBS, REG, RCG, ANK, BL, TDS, JH, MTP, CMS, QY, PMc, FP, UL, DJV: data curation; writing—review and editing.

FUNDING STATEMENT

This work was supported by National Institute of Mental Health Intramural Research Program project ZIA-MH002781.

COMPETING INTERESTS

PMP has received speakers' honoraria from Daiichi Sankyo, Libbs, Sandoz, Novartis, and Abbot. NJ is MPI of a research related grant from Biogen Inc, for work unrelated to this manuscript. JCS is a consultant for PFIZER, SUNOVIAN, SANOFI, JOHNSON & JOHNSON, Livanova, and Boehringer Ingelheim. Material support (for IOL site, led by MA and GJD) was provided by Neuronetics. FA is Section Editor of NeuroImage: Clinical; has received speaker honoraria from Philips, Novartis, and Biogen Idec; and receives or has received research supports from the Italian Ministry of Health, AriSLA (Fondazione Italiana di Ricerca per la SLA), and the European Research Council. EC has received research supports from the Italian Ministry of Health. MF is Editor-in-Chief of the Journal of Neurology; received compensation for consulting services and/or speaking activities from Biogen Idec, Merck-Serono, Novartis, and Teva Pharmaceutical Industries; and receives research support from Biogen Idec, Merck-Serono, Novartis, Teva Pharmaceutical Industries, Roche, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla, and ARiSLA (Fondazione Italiana di Ricerca per la SLA). MF is Editor-in-Chief of the Journal of Neurology; received compensation for consulting services and/or speaking activities from Biogen Idec, Merck-Serono, Novartis, and Teva Pharmaceutical Industries; and receives research support from Biogen Idec, Merck-Serono, Novartis, Teva Pharmaceutical Industries, Roche, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla, and ARiSLA (Fondazione Italiana di Ricerca per la SLA). GAF owns equity in Alto Neuroscience and has served as a consultant for SynpaseBio AI and Alto Neuroscience. MPP is an advisor to Spring Care, Inc., a behavioral health startup, he has received royalties for an article about methamphetamine in UpToDate. MBS has in the past 3 years been a consultant for Actelion, Acadia Pharmaceuticals, Aptinyx, Bionomics, BioXcel Therapeutics, Clexio, EmpowerPharm, Epivario, GW Pharmaceuticals, Janssen, Jazz Pharmaceuticals, Roche/Genentech, and Oxeia Biopharmaceuticals; and has stock options in Oxeia Biopharmaceuticals and Epivario. DJS has received consultancy honoraria from Discovery Vitality, Johnson & Johnson, Kanna, L'Oreal, Lundbeck, Orion, Sanofi, Servier, Takeda and Vistagen. KH is a scientific advisor to the Aury Care GmbH, which develops an AI-based chatbot providing mental health support. He holds virtual stock options in Aury Care GmbH. Previously, this relationship was with the Mental Tech GmbH, from which the Aury Care GmbH took over product development. This research was supported [in part] by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-026-04078-3.

Correspondence and requests for materials should be addressed to Corey Richier.

Reprints and permission information is available at http://www.nature.com/reprints

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

Data and code sharing: The data analyzed in this study is a combination of open datasets and the collaborative work of ENIGMA-Anxiety. The ENIGMA-Anxiety Working Group accepts research proposals from bona fide researchers. The data are shared by each participating PI after approval of the project. For more information please see: https://enigma.ini.usc.edu/ongoing/enigma-anxiety/. The code used in the analysis described here will be available at https://github.com/coreyjr2/ENIGMA_BrainAge_GAD_structural.

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

© The Author(s) 2026

[1]Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [2]Emotion and Development Branch, National Institute of Mental Health, Bethesda, MD, USA. [3]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands. [4]Department of Psychology, The Catholic University of America, Washington, DC, USA. [5]Leiden University, Institute of Education & Child Studies, Section Forensic Family & Youth Care, Leiden, The Netherlands. [6]Department of Psychology, HMU Health and Medical University Erfurt, Erfurt, Germany. [7]Sant Pau Mental Health Research Group, Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain. [8]Centro de Investigación Biomédica En Red en Salud Mental (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. [9]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain. [10]Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA. [11]LiNC, Department of Psychiatry, Federal University of São Paulo, São Paulo, Brazil. [12]Section on Negative Affect and Social Processes, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [13]Center for Neurobehavioral Research, Boys Town National Research Hospital, Boys Town, NE, USA. [14]Child and Adolescent Mental Health Center, Copenhagen University Hospital – Mental Health Services CPH, Copenhagen, Denmark. [15]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [16]Department of Psychiatry, NYU School of Medicine, New York University, New York, NY, USA. [17]Department of Psychology, Penn State University, University Park, PA, USA. [18]Department of Psychiatry and Behavioral Health, The Ohio State University, Columbus, OH, USA. [19]Department of Psychiatry & Behavioral Neuroscience, University of Cincinnati, Cincinnati, OH, USA. [20]Behavioral Epidemiology, Institute of Clinical Psychology and Psychotherapy, TUD Dresden University of Technology, Dresden, Germany. [21]Imaging Genetics Center, Mark and Mary Stevens

Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA, USA. [22]Center for Brain Science & Department of Psychology, Harvard University, Cambridge, MA, USA. [23]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [24]Psychology Department & Neuroscience Center, Brigham Young University, Provo, UT, USA. [25]Center Of Excellence On Mood Disorders, Louis A. Faillace, MD, Department of Psychiatry and Behavioral Sciences, The University of Texas Health Science Center at Houston, Houston, TX, USA. [26]Olin Neuropsychiatry Research Center, Institute of Living, Hartford Hospital, Hartford, CT, USA. [27]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. [28]Anxiety Disorders Center, Institute of Living, Hartford Hospital, Hartford, CT, USA. [29]Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [30]Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy. [31]Institute of Medical Psychology and Systems Neuroscience, University of Muenster, Muenster, Germany. [32]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA. [33]Department Psychology, University of Pittsburgh, Pittsburgh, PA, USA. [34]Anxiety Disorder Program, Hospital de Clínicas de Porto Alegre, Department of Psychiatry, Federal University of Rio Grande do Sul, Porto Alegre, Brazil. [35]Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy. [36]Vita-Salute San Raffaele University, Milan, Italy. [37]Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. [38]Neurophysiology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. [39]Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. [40]Institute of Mental Health, University of Belgrade, Belgrade, Serbia. [41]Department of Psychiatry, School of Medicine, University of Belgrade, Belgrade, Serbia. [42]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [43]Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA. [44]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [45]German Center for Neurodegenerative Diseases (DZNE), Site Rostock/Greifswald, Greifswald, Germany. [46]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [47]Institute for Diagnostic Radiology and Neuroradiology, University Medicine Greifswald, Greifswald, Germany. [48]Center for Neuromodulation in Depression and Stress, University of Pennsylvania, Philadelphia, PA, USA. [49]Section on Neurobiology of Fear and Anxiety, National Institute of Mental Health, Bethesda, MD, USA. [50]Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA. [51]Department of Physics, Stony Brook University, Stony Brook, NY, USA. [52]Department of Neuroscience, Brighton and Sussex Medical School, University of Sussex, Brighton, UK. [53]Department of Psychology, Brunel University London, Uxbridge, UK. [54]"Enrico Fermi" Research Center, Rome, Italy. [55]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [56]Department of Psychology, Sapienza University of Rome, Rome, Italy. [57]IRCCS Santa Lucia Foundation, Rome, Italy. [58]Department of Psychiatry and Behavioral Sciences, The University of Texas at Austin Dell Medical School, Austin, TX, USA. [59]Laureate Institute for Brain Research, Tulsa, OK, USA. [60]Department of Psychiatry, School of Medicine and Herbert Wertheim School of Public Health, University of California, San Diego, La Jolla, CA, USA. [61]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. [62]Department of Psychology, Vanderbilt University, Nashville, TN, USA. [63]Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA. [64]Department of Psychiatry, Washington University, St. Louis, MO, USA. [65]Machine Learning Core, National Institute of Mental Health, Bethesda, MD, USA. [66]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany. [67]German Center for Mental Health (DZPG), partner site Berlin-Potsdam, Berlin, Germany. [68]Department of Psychiatry, Amsterdam UMC, location VUMC, Amsterdam, The Netherlands. [69]Department of Psychiatry & Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [70]Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands. [71]Department of Developmental and Educational Psychology, Institute of Psychology, Leiden University, Leiden, The Netherlands. [72]Leiden Institute for Brain and Cognition, Leiden, The Netherlands. [73]South African Medical Research Council Unit on Risk & Resilience in Mental Disorders, Department of Psychiatry & Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [74]University of Texas Rio Grande Valley, Brownsville, TX, USA. [75]Division of Child and Adolescent Psychiatry, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA.
✉email: coreyjr2@illinois.edu