Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Translating Imaging-based Brain Age Models for Predicting Accelerated Brain Ageing in African Populations: A Pilot Study on a Nigerian Cohort.

Udunna Anazodo

udunna.anazodo@mcgill.ca

Montreal Neurological Institute, McGill University

Tolulope Olusuyi

Medical Artificial Intelligence Laboratory (MAI Lab), Lagos https://orcid.org/0009-0007-7182-0785

Jasmine Cakmak

Montreal Neurological Institute, McGill University, Montreal, Canada

Oumayma Soula

University of Sfax, Sfax, Tunisia

Confidence Raymond

Montreal Neurological Institute, McGill University, Montreal, Canada

Reza Rajabli

Montreal Neurological Institute, McGill University, Montreal, Canada

Charity Umoren

Medical Artificial Intelligence Laboratory (MAI Lab), Lagos

Ifeoluwa Oladeji

Medical Artificial Intelligence Laboratory (MAI Lab), Lagos

Anu Gbadamosi

Medical Artificial Intelligence Laboratory, MAI Lab

Oluyemisi Toyobo

Medical Artificial Intelligence Laboratory (MAI Lab), Lagos

Farouk Dako

University of Pennsylvania, USA

Abbas Muhammad

Bayero University, Kano

Suwaid Abba

Bayero University, Kano

Rachel Akinola

Lagos State University Teaching Hospital (LASUTH), Lagos

Harrison Aduluwa

Medical Artificial Intelligence Laboratory (MAI Lab), Lagos

Chukwuebuka Eze
Regions Healthcare Hospitals and Specialist Clinics, Owerri

Benjamin Anyanwu
Regions Healthcare Hospitals and Specialist Clinics, Owerri

Chinedum Anosike
Accuread Radiology Nigeria, Lagos

Jasmit Shah
Brain and Mind Institute, Aga Khan University, Nairobi

Chinedu Udeh-Momoh
Brain and Mind Institute, Aga Khan University, Nairobi https://orcid.org/0000-0002-7357-4692

James Cole
UCL https://orcid.org/0000-0003-1908-5588

D. Collins
McGill University https://orcid.org/0000-0002-8432-7021

Abiodun Fatade
Crestview Radiology Ltd, Lagos

Maruf Adewole
Medical Artificial Intelligence Laboratory (MAI Lab), Lagos https://orcid.org/0000-0002-6562-2834

Article

Keywords: Brain age, Chronological age, Africa, MRI, Gaussian process regression, machine learning, deep learning, population-specific modelling

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11076290/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Translating Imaging-based Brain Age Models for Predicting Accelerated Brain Ageing in African Populations: A Pilot Study on a Nigerian Cohort.

Abstract

Background: Brain age is an emerging imaging biomarker for early detection of neurodegenerative disorders, based on the estimation of the brain's biological age. Existing predictive models are predominantly trained on Global North (GN) neuroimaging data, limiting their sensitivity when applied to underrepresented populations due to geographic, imaging scanner, and biological heterogeneity. We investigated this bias by adapting the standard Gaussian Process Regression (GPR) method to a population-specific (PS) brain age model.

Methods: Using a multi-institutional T1-weighted brain MRI dataset from 309 Nigerian adults, we implemented a robust low-compute image analysis pipeline which includes repurposing clinical scans and high-resolution cortical parcellation with multi-stage quality control. The PS-GPR model was trained on regional brain volumetric features from the Nigerian MRI scans and its performance compared to four common GN pre-trained models.

Results: The proposed PS-GPR model showed a significantly lower prediction error (mean absolute error = 9.82 years) without age-bias correction and few seconds inference time compared to GN-trained machine learning models (mean absolute error = 10.00 – 16.70 years). SPARROW, a generalizable deep learning framework achieved an MAE of 7.87 but failed on 6% of the test set. This early work demonstrates the importance of representative data to build resource-efficient clinically viable models.

Conclusion: This population-specific model is a step towards sensitive low-compute African brain age models. With the expansion of imaging data beyond Nigeria through the Africa Neuroimaging Archive (AfNiA), we will further develop the PS model to identify individuals with accelerated brain ageing for early intervention in at-risk adults.

Keywords: Brain age, Chronological age, Africa, MRI, Gaussian process regression, machine learning, deep learning, population-specific modelling

Plain Language Summary

Brain age is a promising tool that can use MRI scans to estimate how 'old' a brain appears relative to the individual's actual age. If the brain appears older, it may signal risk for neurological diseases such as dementia, making accurate prediction important. However, most brain age models have been developed with data from North America and Europe. This study trained a regression model using Nigerian MRI scans and tested whether established Western models would better predict brain ageing in Nigerian populations, as a step towards applying brain age across African populations. The model was more accurate, demonstrating that models for assessing brain ageing must be developed using data from the populations they aim to serve.

Introduction

The ageing process is accompanied by structural and functional changes in the human brain[1]. Brain age has emerged as a powerful imaging biomarker that quantifies this ageing process by estimating the biological age of an individual's brain from neuroimaging data[2]. A prediction model is trained on brain imaging data from a large sample of healthy individuals and used to estimate the novel individual's age[3]. The discrepancy between an individual's predicted brain age and their chronological age known as Brain Age Gap (BAG), is proposed to reflect the brain's health[4,5]. Studies have shown that normal ageing involves gradual changes in brain structure and function whereas pathological ageing is distinguished by the magnitude, distribution, and the rate of these changes, ultimately manifesting as clinical impairment[6]. Increased brain ageing (a positive BAG) has been associated with a range of negative outcomes, including cognitive decline, the onset of neurodegenerative diseases such as Alzheimer's, and increased mortality[4,7]. As such, brain age holds significant promise as a tool for early detection and risk stratification in clinical neurology, particularly in resource-constraint settings.

However, the development and validation of brain age models are subject to a critical, yet often overlooked limitation - the geographic bias in the training of neuroimaging data[8]. The majority of models published to date have been trained and validated on datasets almost exclusively from the Global North (GN) regions (North America, Europe, Australia etc). This creates a fundamental problem of representation beyond ethnicity[9]. Brain structure is not only influenced by genetics but also by a lifetime of environmental, social, and economic factors[10–12]. Neuroimaging data is also sensitive to technical variables such as the magnetic field strength of the scanner, scanner manufacturer, and acquisition protocols[13]. When a model learns from these demographic-specific datasets, it may capture population-specific and scanner-specific confounding variables as much as it captures the 'true' biological signal of ageing. The direct application of such models to populations such as those in sub-Saharan Africa, not represented in the training data, risks generating inaccurate and potentially misleading results. A model might interpret normal, population-specific anatomical variations as signs of accelerated ageing, or miss true pathological changes that fall outside its learned patterns. Thus, the use of biased models could lead to misdiagnosis, inappropriate patient counselling, and a failure to identify individuals at genuine risk within these underserved populations[14]. This represents a significant gap in global health equity where advanced computational tools intended to improve care, may instead widen the existing disparities[15].

Nigeria, as Africa's most populous nation and representative of many low- and middle-income country (LMIC) settings[16,17], is a compelling use-case for addressing this bias. Its clinical neuroimaging landscape is characterised by multi-delivery (public/private and hospital/diagnostic clinic), multi-scanner environments, often relying on faster two dimensional (2D) acquisition protocols to increase patient throughput[18]. This results in heterogeneous MRI scans that are structurally different from the high-resolution standardized three dimensional (3D) datasets typically used to train existing brain age models. To address this bias, this study adapted a kernel-based Bayesian model and tailored it to be computationally efficient and clinically deployable for population-specific (PS) prediction in resource-constrained settings, such as Nigeria. The primary objectives are to (1) curate a multi-site and multi-scanner normative African T1-weighted MRI dataset from routine clinical scans, (2) use this PS data to train a computationally efficient brain age prediction model based on a

Gaussian Process Regression (GPR) framework, and (3) benchmark the PS-GPR model's performance and its computational demands by comparing its performance metrics (mean absolute error R^2, etc) against established publicly available machine learning (ML) and deep learning (DL) brain age models. In the pilot Nigerian cohort, the PS-GPR method achieved superior predictive accuracy at significantly lower computational costs compared to the complex non-representative ML and DL alternatives, demonstrating the significance of population-specific approaches in assessing brain health in Africa and other resource-constrained settings.

Methods

Study Population and Imaging Data

The study used a multi-institutional T1-weighted brain MRI dataset derived from previously acquired clinical scans of over 11,000 participants from the Africa Neuroimaging Archive (AfNiA)[19]. The database has been described elsewhere and the scans included in this study were collected from reports of scans deemed to be neurologically normal (no reported abnormality, described as normal, or unremarkable) by experienced radiologists[20]. Participants below the age of 18 years and those whose age were unknown were excluded. This dataset reflects the real-world clinical and research landscape of many African and low-and-middle income settings, consisting of images acquired across clinical protocols and MRI systems (1.5T magnetic field strength), common in the regions, and exhibiting a wide range of spatial resolutions[18] (see data summary in Table 1). The retrospective data from the multi-instutional study was approved by the Western University Research Ethics Board (HSREB 121287), the institutional review board of the College of Medicine of the University of Lagos (CMUL/HREC/04/22/1090), Lagos State University Teaching Hospital (LREC/06/10/1952), National Hospital Abuja (NHA/EC/049/2023), and Regions Healthcare Hospital (RH-HREC/021/2025).

Image Preprocessing and Quality Assurance

The scans were processed and quality checked using a series of steps summarized in Figure 1 and outlined below. Briefly, the 2D scans were first converted to standard high-resolution 3D isotropic ($1mm^3$) images using the default settings of SynthSR (version 2.0), a deep learning technique designed for this purpose[21]. This step enhanced the 2D spatial resolutions to the required isotropic resolution for volumetric-based brain age modelling. A region-level tissue segmentation and anatomical labelling was performed using the FreeSurfer (version 7.4.1) recon-all pipeline with standard parameters[22,23]. Cortical surface parcellation was done using Destrieux brain atlas[24] and the resulting cortical thickness and subcortical volumes were extracted using FreeSurfer functions (aparcstats2table and asegstats2table).

Rigorous quality checks were performed before and after image segmentation to assess image quality distribution among the varying scanner and acquisition protocols (Figure 1). First, a pre-segmentation quality check was performed on all acquired (2D and 3D) MRI scans and the synthesized (SynthSR outputs) 3D volumes using MRIQC (version 25.1.0)[25]. MRIQC computes various image quality metrics (IQMs) including signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), entropy focus criterion (EFC) and coefficient of joint variation (CJV), from T1-weighted MRI scans. The IQMs were benchmarked against the publicly available Transdiagnostic Connectome Project (TCP) dataset[26], as a reference standard for high-quality neuroimaging data. Although the clinical data may

not match the quality profile of the TCP research-grade scans, the benchmark assesses the quality status of the clinical data. An overall higher SNR and CNR, and lower EFC and CJV indicate less image noise, better tissue delineation, less signal intensity nonuniformity from excessive head motion and other artefacts, and minimal presence of ghosting or blurring, respectively[25]. An additional multi-stage quality control was performed after FreeSurfer segmentation and parcellation. The FreeSurfer outputs were progressively passed through various quality checks (Figure 1): 1) manual inspection of gray matter (GM) and white matter (WM) segmentation outputs of all 3D volumes using FSL (version 6.0.7.18)[27], and 2) outlier detection of all outputs to select top-quality segmentations by excluding participants with mean Euler number greater than 50[23,28]. Euler number has been shown to be a reliable quality control strategy for brain segmentation [29]. The threshold of 50 was determined empirically based on the distribution of Euler numbers in the dataset. This approach is consistent with recommendations of literature[29] emphasizing that Euler number thresholds vary by scanner, acquisition protocol, and population characteristics.

Data Processing

Neurological cortical and subcortical features were extracted based on the Destrieux atlas parcellations of quality-controlled FreeSurfer outputs. Extraction was done using Freesurfer's aparcstats2table function (aparc.a2009s) for both left and right hemispheres and asegstats2table for subcortical volumes. This extracted 187 features that were compiled into a feature matrix (*X*), with chronological age as the target vector (*y*). The dataset was randomly split into training (70%), validation (15%), and held-out test (15%) sets using the train_test_split from scikit-learn (version 1.4.2), with shuffling enabled and a fixed random seed (RANDOM_SEED = 42) for reproducibility. Features were standardized to zero mean and unit variance based on the training set statistics, and the same transformation applied to the validation and test sets.

Population-Specific Gaussian Process Regressor Model Architecture and Training

The GPR model is a non-parametric modelling approach employed in several brain age prediction techniques[30–33]. We trained a GPR *de novo* on the Nigerian data, implemented using scikit-learn (GaussianProcessRegressor). The kernel function was defined as a Radial Basis Function (RBF) multiplied by a constant, given in equation 1 as:

Equation 1

$$C(1.0,\ (1e^{-3},\ 1e^{3}))\ *\ RBF(lengthscale\ =\ 1.0,\ length\ scale\ bounds\ =\ (1e^{-2},\ 1e^{2}))$$

This configuration allows the model to learn the smoothness and scale of the relationship between brain structure and age. The model's internal noise parameter (alpha) was set to $1e^{-10}$, and target values were normalized (normalize_y=True). Within the training set, we performed 5-fold cross-validation to estimate model stability and the final model was retrained on the entire training set (n=216). To minimize known age prediction bias from systematic overestimation of age in younger individuals and underestimation in older individuals[34], a bias correction adjustment step was implemented. This standard post-processing step leads to a non-zero correlation between the BAG and chronological age and ensures the derived BAG is independent of age, making it a more valid assessment for clinical correlational analyses. The slope (β) and intercept (α) of this relationship

derived from regressing BAG against chronological age in the validation set, were used to adjust the test set predictions[35,36], defined in equation 2 as:

Equation 2 $$Corrected\ prediction = Raw\ prediction - (\alpha + \beta * Chronological\ age)$$

Comparative Brain Age Model Benchmarking

The PS-GPS model performance was evaluated on the held-out test set and an external independent test set, using standard regression metrics, without bias correction. Bias correction could not be reliably performed on the comparative GN models, due to lack of individual chronological age data from their training datasets. The primary benchmarking metric was the mean absolute error (MAE) between predicted and chronological age, while secondary metrics were Root Mean Squared Error (RMSE), the coefficient of determination (R^2), and Pearson's correlation coefficient (r). The distribution of BAG was summarized by its mean and standard deviation. The model's computational demand was recorded for comparative efficiency analysis. The performance of the following established brain age models, each pre-trained on high quality (research grade), multi-site, and multi-scanner GN scans of healthy participants was evaluated on the Nigerian test sets, without retraining, fine-tuning, or further data processing:

1. PyBrainAge (beta version): an open-source machine learning model based on the Extra Trees model and pre-trained on 187 FreeSurfer features (cortical thickness, ventricles, and regional volumes) from T1-weighted MRI scans of 29,175 participants, primarily from the UK Biobank (age: range = 2 to 100 years, mean = 46.9 ± 24.4 years and 51.1% female) scanned across 76 sites[37,38]. The model was implemented on a workstation as outlined in the PyBrainAge model repository (https://github.com/james-cole/PyBrainAge).
2. ENIGMA-Brain Age: a pre-trained model based on the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium's framework[39,40]. We implemented the model using the PhotonAI Python wrapper (https://photon-ai.com/), which provides a standardized web-based interface for executing the ridge regression machine learning pipeline, previously trained by the consortium on 2188 T1-weighted MRI scans (age 18 to 75 years and 56.5% female) acquired across 19 cohorts. We uploaded a csv file (one for each sex) on PhotonAI containing the 77 ENIGMA-required FreeSurfer subcortical regions, thickness and surface areas values to get the biological age predictions.
3. SynthBA: a publicly-available 3D Densely Connected Convolutional Network (3D DenseNet), specifically designed for brain age regression from unprocessed T1-weighted images[41]. SynthBA pre-trained synthetic MRI volumes generated from 7060 T1-weighted images of 4105 healthy participants (6-95 years old, 53% female; group mean ages is unspecified), to improve the models generalizability in learning features that are invariant to contrast and resolution across varying scanners and acquisition protocols, without retraining. Predictions were made by directly applying the model, using its pre-trained weights and a batch size of 2 (https://github.com/LemuelPuglisi/SynthBA), to predict biological age of both acquired 2D and 3D scans, as well as the synthesized (SynthSR) 3D scans, rather than extracted volumetric features.
4. SPARROW: a Simple Pseudo-fully-convolutional Architecture with Residuals, Regularized, Optimized and Weight-consolidated 3D neural network (SPARROW), based on the SFCN-reg architecture[42] with residual connections for more stable training. The model trained on

preprocessed T1-weighted images primarily from the UK Biobank[43] and fine-tuned on the Mayo Clinic Study of Aging (MCSA), Open Access Series of Imaging Studies 1 & 3 (OASIS-1 & OASIS-3), Information eXtraction from Images (IXI), and 1000 Functional Connectomes Project (FCON1000) datasets to improve generalizability across different cohorts. We applied the pre-trained weights[44] of this model directly to both the acquired and synthesized 3D images.

To assess the extent of generalizability, we applied all models including PS-GPR directly to an independent external Nigerian cohort (n=25; age=18 - 75 years; mean = 36.36 ± 11.45) acquired from a separate source (0.3T and 1.5T scanners, 0.39-0.98mm x 0.39-0.98mm x 5-8mm image resolution)[45]. The external data was quality-checked, preprocessed and volume extracted as described above, and not retrained or used in any stage of model training, validation, or hyperparameter tuning. All data analysis including image processing, quality assurance, model training and testing (except for SPARROW) were performed on a workstation running on Ubuntu 20.04.6 LTS with an 11th Gen Intel® Core™ i7-11700 CPU (2.50 GHz), 16 GB RAM, and an NVIDIA GeForce GT 730 GPU (driver version 390.157).

Results

Of the 11,000 MRI cases, 343 T1-weighted images were identified as neurologically normal adult scans (Table 1) and prior to any processing, underwent automated quality assessment using MRIQC. The MRIQC IQMs were checked against the TCP reference data. The IQM comparisons confirmed that the Nigerian clinical dataset exhibited greater variance and more frequent outliers across multiple metrics compared to the TCP reference, reflecting the inherent heterogeneity of multi-site clinical acquisitions. The outliers were largely observed in the synthesized acquisitions, as shown in Figure 2. The outliers (n=57) were visually inspected and those deemed to present with no apparent artefact or blurring were retained (n=315) for Freesurfer processing. Following Freesurfer's recon-all preprocessing, six participants were excluded due to incomplete reconstructions.

The final cohort comprised 309 T1-weighted neurologically normal Nigerian adults (164 females, 145 males), with a mean age of 40.66 ± 18.19 years (range: 18 – 88 years) which were allocated to the training, validation, and test sets (Supplementary Table 1). The training dataset had 216 participants (70%) while the validation and test sets had 46 and 47 participants respectively (15% each). Overall, 66% of the Nigerian data (204 of 309) had mean Euler numbers between 0 and -50, indicating generally good reconstruction quality across the dataset. Crucially, 8% of the data had Euler numbers exceeding -200 representing the most challenging scans. The distribution of Euler numbers of all 309 participants is described in the Supplementary (Supplementary Table 2, Supplementary Figure 1-2). Visual inspection of these scans revealed no apparent gray or white matter artefacts and were retained, although they are not research-grade, they represent the clinical reality in Nigeria where infrastructures can produce different quality standards. Expert visual inspection of the most extreme outliers using FSLeyes confirmed the automated findings and ensured no scan with severe segmentation failure was included in downstream analyses. In total, 309 images were used for model training, validation, and testing.

<u>Population-Specific Gaussian Process Regression (PS-GPR): Training and Bias Correction</u>

The performance of all evaluated models on the Nigerian data set is summarised in Table 2, with sex-straified model performance outlined in Supplementary Table 3. The PS-GPR model trained on the Nigerian dataset achieved a cross-validated MAE of 11.02 years (R^2=0.46, r=0.68) during training, an MAE of 9.87 years (R^2=0.39, r=0.63) on the validation set (N=46), and 9.82 years (R^2=0.55, r=0.76) on the test set (N=47) before bias correction. As expected, a strong age-bias was present, with the Brain Age Gap significantly correlated with chronological age (slope β = -0.56, R = -0.73, $p < 0.001$). Following a linear age-bias correction using the validation-derived β and intercept α, the model's performance improved substantially on the test set, yielding an MAE of 6.39 years, an R^2 of 0.83, and a Pearson correlation of 0.93 between predicted and chronological age. The correction successfully centered the BAG, with a final distribution of -1.22 ± 7.61 years and no significant correlation to age ($p > 0.05$).

Comparative Performance of Established Models on Internal Validation (Held-out Test Set)

The performance of selected existing, pre-trained models on Nigerian dataset revealed significant performance gaps (Table 2 and Supplementary Table 4). PyBrainAge achieved an MAE of 10.0 years (R^2 = 0.53) representing a 112% increase from its original MAE (4.7 years). ENIGMA-BrainAge performed substantially worse, with an MAE of 16.70 years and a negative R^2 (-0.30). SynthBA achieved an MAE of 12.29 years on original 2D and 3D images (R^2 = 0.18) and an MAE of 17.34 years on a combination of SynthSR-synthesized and original 3D images, compared to its reported MAE of 4.9. SPARROW failed completely on 3 of the 47 test participants due to preprocessing or registration errors. All 3 test cases were synthesized 3D images. On the remaining 44 where inference was successful, it achieved an MAE of 7.87, about 207% increase from its reported MAE of 2.56 years on GN cohorts. For head-to-head comparison, we report the performance of all models on the 44 test set only (Table 2 and Figure 3). However, generalization errors varied by sex. PyBrainAge's MAE was worse for males (11.14 years) than for females (10.06 years). Both ENIGMA-BrainAge and SynthBA models also performed worse for males than for females. SPARROW maintained a similar performance for both (male, MAE = 7.72; female, MAE = 7.98) (Figure 4).

In terms of compute and inference time, all feature-based models (PyBrainAge, ENIGMA, PS-GPR) require Freesurfer's recon-all pipeline for cortical parcellation and feature extraction. For the full data of 309 participants, the cumulative preprocessing was 40 days, 18 hours, 48 minutes, and 20 seconds of continuous processing. This is the single practical bottleneck in the workflow that highlights a critical barrier to clinical translation in resource-limited environments. We summarized the inference time for each model along with the hardware specifications used in Table 3.

Comparative Performance of Established Models on External Validation (Independent Set)

In the external unseen Nigerian cohort, PS-GPR achieved an MAE of 13.29 years uncorrected and an MAE of 8.78 years after age-bias correction, an increase from the internal test set (MAE = 6.39 years) (Supplementary Table 5). It maintained a moderate correlation between predicted and chronological age (r = 0.71, $p = 8.19e^{-05}$) with a negative R^2 value (-0.05) which indicates that the model does not explain any variance beyond predicting the mean age. This suggests that the external cohort may differ systematically from the training distribution in ways that the current feature set cannot capture. All GN pre-trained models performed poorly on the external test set with uniformly negative R^2 values indicating that none explained meaningful variance in chronological age (Supplementary

Table 6). PyBrainAge achieved an MAE of 13.87 years (R^2 = -1.64, r = 0.02, p > 0.05). Its error increased by 38% compared to its internal test set performance (10.0 years). ENIGMA-Brain Age performed poorly with an MAE of 33.25 years (R^2 = -13.25, r = 0.53). The correlation suggests a linear relationship but the massive error and negative R^2 indicate systematic bias. SynthBA and SPARROW showed similar performance to PyBrainAge (SynthBA, MAE = 12.76, R^2 = -1.03, r = 0.51; SPARROW, MAE = 10.52, R^2 = -0.23, r = 0.60) failing to capture age-related variance.

Discussion

This study curated normative T1-weighted MRI scans to predict the brain age of healthy adults in the Nigerian population using low-resource machine learning algorithms and provided evidence on the transportability of established brain age models trained on non-representative, Global North datasets. Although a retrospective observational study[46,47] design was used, the quality control analysis revealed that the heterogeneous clinical Nigerian dataset matches real-world imaging settings common in Africa[45], with lower-quality scans than research-grade GN datasets like TCP[26].

While research-grade anatomical MRI scans, including from TCP, UK Biobank, and other large imaging repositories have provided robust estimation of brain age, the performance and generalizability/transportability of existing brain age models developed using these GN datasets (PyBrainAge[37], ENIGMA-Brain Age[40], SynthBA[41], and SPARROW[43]), showed a marked degradation in their performance when applied to the Nigerian cohort compared to their reported metrics. This is consistent with prior work highlighting demographic bias in models trained predominantly on Global North datasets[48–50]. This bias arises from models learning population-specific anatomical standards, scanner-related artifacts, and sociodemographic variables present in training data[51,52]. This could partly account for the broader application of SPARROW (MAE = 7.87 years, R^2 = 0.71) compared to other methods. It is the most sophisticated deep learning model in this benchmark, trained on the largest multi-study cohort (35,815 participants) and incorporating advanced regularization that could augment inter-population bias[43]. Although this may suggest that deep learning architectures may offer an advantage over simpler machine learning approaches in capturing age-related brain patterns when applied to new populations, the lower performance of SynthBA compared to two machine learning models (PyBrainAge and PS-GPR), require further investigation. It is quite possible that SynthBA mixed input shows its input-dependency. Despite being designed for synthesized images, it did not generalise to the study's SynthSR outputs and combined acquired 3D images, suggesting that simply harmonizing image resolution is insufficient without also addressing the biological and technical heterogeneity of the dataset. Equally, the failure of SPARROW on 6% of the test set due to preprocessing errors reflects the model's assumption about input image quality, spatial normalization, etc. These assumptions may hold for standardized, research-quality UK Biobank data but can break down when confronted with the greater anatomical and acquisition heterogeneity of a multi-site Nigerian dataset.

Despite a modest sample size (n=309) and an imbalanced age distribution -common challenges of underrepresented populations, the proposed PS-GPR model showed a superior performance even prior to age-bias correction. The uncorrected model exhibited moderate prediction error (MAE = 9.82) and clear age-dependent bias (slope β = -0.56) which is characterised by the overestimation in younger participants and underestimation in older individuals, a popular pattern in brain age modelling that reflects regression-to-the-mean effects[34,35,46]. Bias corrected PS-GPR model achieved

an MAE of 6.39 years, a performance comparable to models trained on large, homogenous Global North cohorts e.g. UK Biobank studies (MAE ~3-6 years)[47]. This implies that population-specific modelling can compensate for limited data by improving feature relevance. An instructive finding, demonstrated by the uniformly negative R^2 values on the external test set observed for all models including PS-GPR. These values inform us about the proportion of the variance in the dependent variable (Y; chronological age) that can be explained by the independent variables (X; extracted features). Thus a negative R^2 indicates that the models explain no variance and may be systematically biased. A potential explanation for this could be that the external cohort differs substantially from the training distribution in unmeasured ways, including different scanner types (e.g. acquisition parameter) or population characteristics (e.g. age range and sample size)[53] not captured by volumetric features.

One of the study's core objectives was to evaluate the deployability of brain age models in resource-constrained settings. Like the majority of established volumetric ML models (ENIGMA-Brain Age, PyBrainAge etc), the proposed PS-GPR is fundamentally dependent on the FreeSurfer recon-all pipeline for cortical parcellation and feature extraction. For the study cohort of 309 participants, this preprocessing step required approximately 980 hours of processing time, representing a dominant computational cost. Based on current Google Cloud Platform pricing[54] (using a standard 4-vCPU virtual machine), this corresponds to an estimated compute cost of approximately USD $140/month on the lowest cost per core E2 series (running on a free Debian/Ubuntu operating system), excluding storage and network charges. Including temporary storage (~150 GB) and data egress yields a total estimated cloud cost of approximately USD $170–180/month. This denotes that anatomical preprocessing represents the dominant computational and economic bottleneck for deploying volumetric brain age models in resource-constrained settings and its timely clinical adoption. In contrast, end-to-end deep learning models (e.g SPARROW) aim to bypass explicit feature extraction by learning directly from raw images. However, the computational demand is shifted to the need for extensive GPU resources particularly for training[55], a luxury that is often scarce in low-resource settings. The study's PS-GPR model leverages the well-validated morphometrics from FreeSurfer and then applies an efficient lightweight model that confers negligible added compute cost to the overall analysis. However, the study's results reinforce that future innovation must prioritize the development of accurate less resource-intensive segmentation tools to truly enable equitable access to promising early biomarkers as brain age.

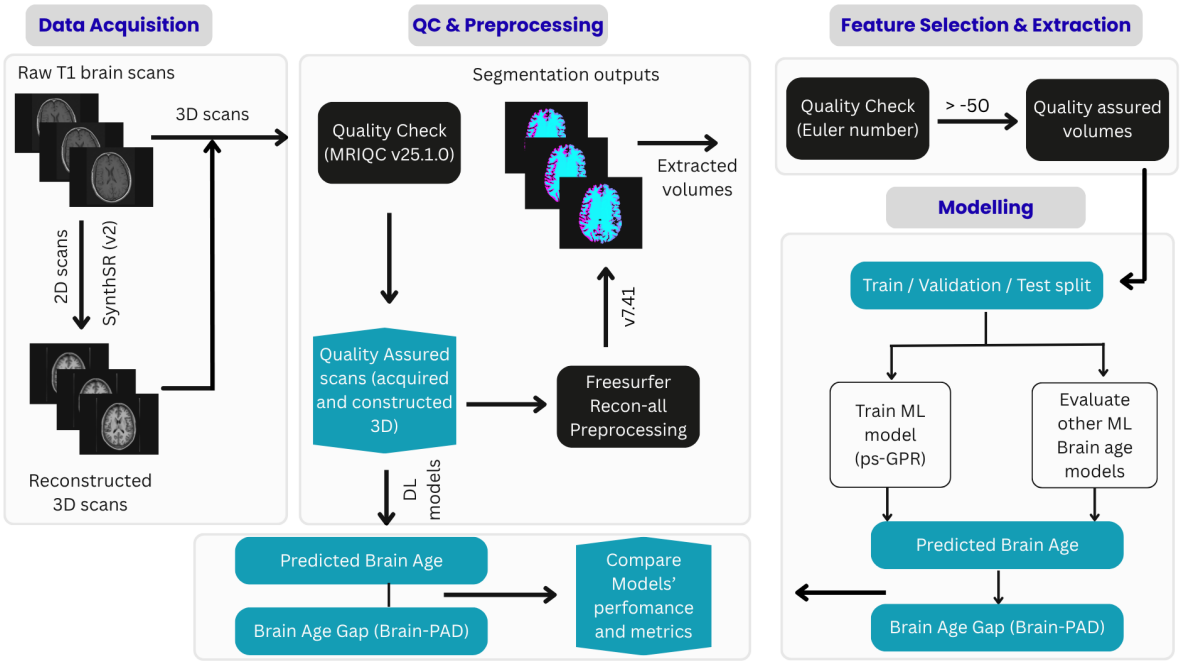
Limitations and Future Directions

This study has some important limitations. First, the sample size limits the complexity of the model that could have been used for training and equally may contribute to the relatively higher prediction error seen in all models compared to reported GN results. Second, brain age models may capture sex-specific anatomical patterns[56,57], but training datasets with imbalanced sex distributions such as the one used in this study can introduce bias in modelling, limiting the investigation of these brain patterns. The study's sample size was simply insufficient for a robust sex-stratified analysis and the sex distribution of some of the comparative GN datasets are not publicly available. While, sex-specific brain ageing remains unclear in African populations, this study unfortunately, was not designed to systematically investigate the biological or sociodemographic determinants of these sex differences. As such, larger multi-national sex- and age-stratified African cohorts are needed to build more robust

models, leveraging ML resource-efficient approaches such as those presented here. New studies should incorporate experimental designs that include measures of hormonal status (e.g. estrogen, testosterone, progesterone levels), and other sex-related endocrine states along with readily available clinical brain MRI. This effort is underway with curation of larger brain imaging data across Africa, to establish the proposed approach and validate a more representative brain age model. Third, the dependency on FreeSurfer preprocessing remains a major constraint. Future work needs to explore alternative, lightweight segmentation models or incorporate transfer learning from models trained on larger, diverse datasets to reduce this bottleneck. Fourth, while the study evaluated several benchmark models, a comprehensive comparison with all state-of-the-art brain age models is beyond scope. Future studies can explore a wider array of machine learning and deep learning architectures on the same African population to establish reliable population-specific brain age index. Finally, external validation in other African populations beyond Nigeria, is essential to assess the pan-African applicability of the proposed population-specific model and to distinguish population-specific from site-specific effects.

In conclusion, this work demonstrates that equitable and accurate brain age assessment in underrepresented populations is achievable but requires a deliberate use of population-specific real-world data for model training to ensure a clinically deployable biomarker. By providing a validated model on an initial Nigerian population and transparent benchmark of existing tools, this study offers a practical pathway for improving neuroimaging-based risk stratification in African clinical neuroscience. It highlights that the future of global brain health does not depend on universally applied ‘one-size-fits-all’ models, but on an ecosystem of validated, population-aware tools.

Figure 1: **Workflow for developing and evaluation of the population-specific brain age prediction model.**



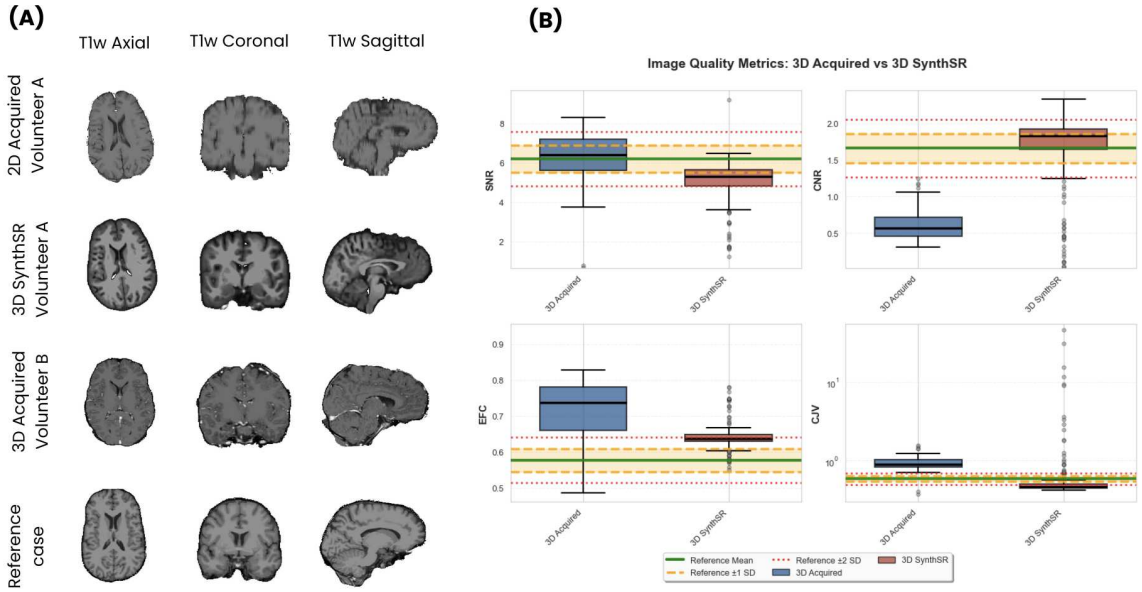
The performance of known machine learning (ML) and deep learning (DL) models were tested on the same population data. Three dimensional (3D) isotropic images were reconstructed from two dimensional (2D) T1-weighted (T1) brain scans before further preprocessing, quality checks, and segmentation to regional gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) volumes were carried out. Quality-assured regional volumes were used to train the adapted population-specific Gaussian Process Regression (PS-GPR) model and to evaluate four other established models (PyBrainAge (ML), ENIGMA (ML), SynthBA (DL), and SPARROW (DL).

Table 1: Study Population and MRI Data Summary.

Scanner Type (1.5T Model)	N (Participants)	Age (years)	Mean (SD) (years)	Sex M/F	Acquisition type (3D/2D)	Image Resolution (x/y/z mm)
GE Medical (Signa Explorer)	1	22.0	-	0 / 1	1 / 0	0.94 x 0.94 x 2.0
Philips (Achieva)	109	18.0 – 84.0	49.83 (18.56)	56 / 53	100 / 9	0.23 – 0.94 × 0.23 – 0.94 × 1.0 – 5.0
Siemens Healthineers (Magnetom Essenza)	219	18.0 – 88.0	36.48 (16.23)	102 / 117	20 / 199	0.31 – 1.15 × 0.31 - 1.15 × 1.2 – 5.0
Toshiba (Vantage Elan)	1	24.0	-	1 / 0	1 / 0	0.5 x 0.5 x 1
United Imaging (uMR 588)	13	22.0 - 35.0	28.69 (4.55)	6 / 7	13 / 0	1 x 1 x 1
Total	**343**	**18 - 88**	**40.48 (17.99)**	**165 / 178**	**135 / 208**	-

The curated normative dataset comprised 343 male (M) and female (F) participants retrospectively acquired across five 1.5 Tesla (T) scanners from General Electric (GE) HealthCare, Philips, Siemens Healthineers, Toshiba/Canon Medical Systems, and United Imaging. Age distribution ranged from 18 to 88 years, with the Philips and Siemens scanners contributing the majority of data. Both three-dimensional (3D) and two-dimensional (2D) acquisitions were included, with voxel resolutions varying from 0.23 × 0.94 × 0.23 mm to 1.15 × 1.2 × 5.0 mm, reflecting scanner-specific acquisition characteristics.

Figure 2: Samples of scans used in this study and image quality of original and synthesized 3D images.



(A) Axial, Coronal, and Sagittal T1w slices of 2 participants showing two-dimensional (2D) acquired image (Volunteer A), three-dimensional (3D) SynthSR-synthesized image of same subject, and 3D acquired image (Volunteer B) with Freesurfer Bert's image as a reference. (B) Comparison of image quality metrics of synthesized and acquired 3D images. Distribution of signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), entropy focus criterion (EFC), and coefficient of joint variation (CJV) are shown with reference values.

Table 2: Comparison of models performance on Internal Nigerian dataset.

Model	Reported results					Nigerian-data results					
	N (M/F)	Age (SD) (years)	MAE (years)	R^2	r	N (M/F)	Age (SD) (years)	MAE (years)	R^2	r	BrainPAD (years)
PyBrainAge	29,175 (14,267/14,908)	46.9(24.4)	4.7	0.42	0.66	44 (19/25)	41.29 (18.77)	10.52	0.51	0.73	-2.71 ± 12.86
ENIGMA BrainAge (M/F)	2188 (952/1236)	43.09 (15.32) / 39.37 (15.69)	6.5/6.8	0.72/ 0.69	0.85/ 0.83	44 (19/25)	41.29 (18.77)	17.51	-0.35	0.47	-13.41 ± 17.24
SynthBA	4105 (1929/2176)	42-95	4.9	-	-	44 (19/25)	41.29 (18.77)	12.33	0.19	0.54	-5.03 ± 16.07
SPARROW	40,497 (18,766/21,548)*	66(8.23)	2.56	0.98	0.95	44 (19/25)	41.29 (18.77)	7.87	0.71	0.85	-0.58 ± 10.15
PS-GPR	216 (101/115)	41.15(18.41)	-	-	-	44 (19/25)	41.29 (18.77)	10.24	0.53	0.74	-1.39± 12.74

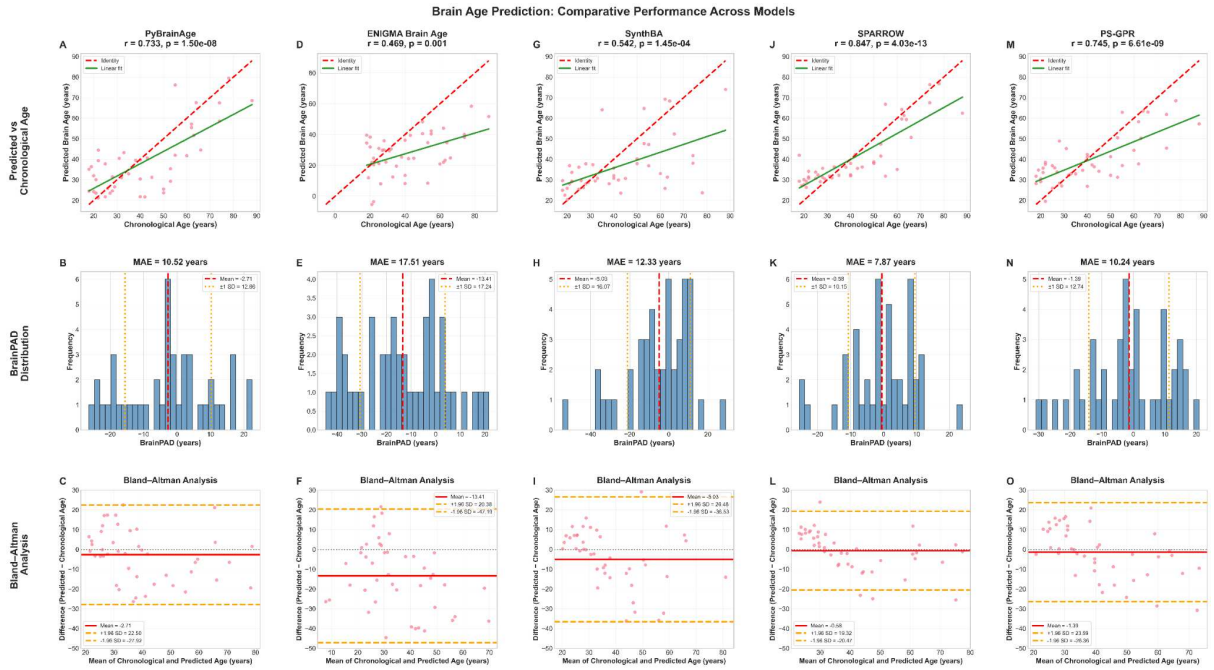
Reported results of models from the original publications are shown alongside performance on the Nigerian population dataset (without bias correction). Metrics include mean absolute error (MAE, years), coefficient of determination (R^2), Pearson correlation coefficient (r), and Brain-Predicted Age Difference (BrainPAD, mean ± SD, years). PS-GPR achieved the lowest MAE and highest R^2 and r on the Nigerian dataset, indicating superior predictive performance in this cohort. N, number of participants; MAE, mean absolute error; R^2, R-Squared (coefficient of determination); r, Pearson's correlation coefficient. *SPARROW reported 183 unidentified participants (unknown sex).

Table 3: Resource Efficiency of Brain Age Models.

Model	Prediction Time	System specs	Setup
PyBrainAge	28.24s (all participants)	Intel i5-6300U, 8GB RAM	Requires python environment, specific versions of dependencies, longer setup
ENIGMA (Male / Female)	3.88s / 3.17s (all participants)	Intel i5-6300U, 8GB RAM	Requires a csv file, specified columns, upload and get predictions (after FreeSurfer preprocessing)
SynthBA	1hr 47mins (all participants)	Intel i7-11700, 16GB RAM	No preprocessing required, minimal setup required, high computation and longer time required
SPARROW	~3mins (per participant)	Intel i7-8700T CPU, 32GB RAM	High computational demands and longer time required
PS-GPR	11.73s (including training time)	Intel i5-6300U, 8GB RAM	Load data and get predictions (after FreeSurfer preprocessing)

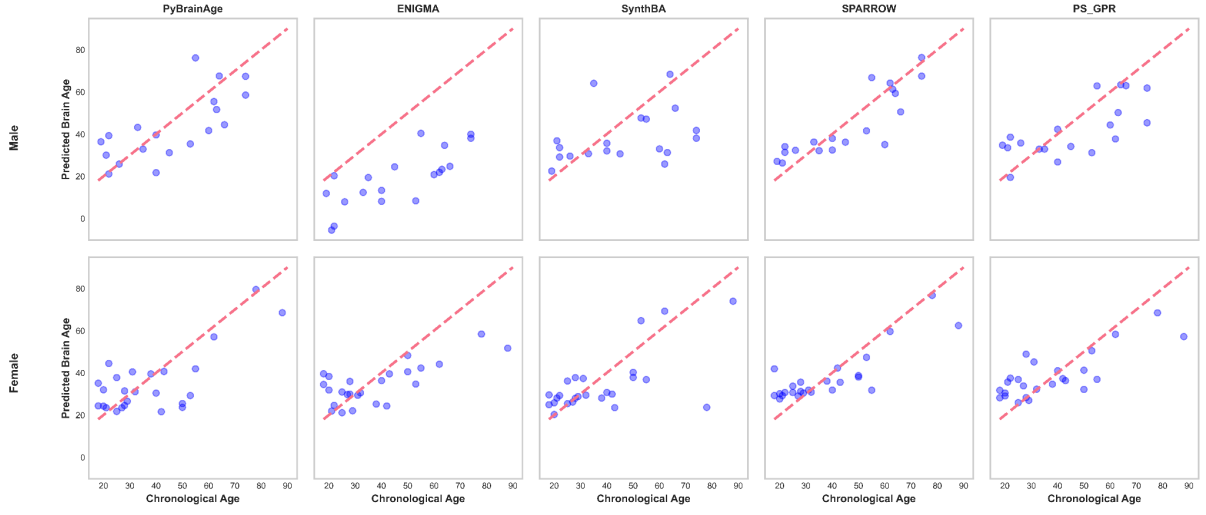
Inference time for all participants (per participant in SPARROW) is shown for each model on the specified hardware. Freesurfer preprocessing required for all ML models totaled 980 hours and is not included in times reported. All ML models were tested on the same Hp laptop “Intel i5-6300U CPU (2.4 GHz), 8 GB RAM, Windows 11 x64” and SynthBA DL model on a Dell Desktop “Intel i7-11700 @ 2.50 GHz, 16GB RAM, Ubuntu 20.04.6 LTS x86_64”; Freesurfer preprocessing for ENIGMA, PyBrainAge, and PS-GPR models took 40 days, 18 hours, 48 minutes, 20 seconds for 309 participants.

Figure 3: Comparison of models performance on test set (n=44)



Top row shows predicted vs chronological age with the identity line (red dashed). The middle row shows Bland-Altman plots illustrating model bias and limits of agreement. Bottom row shows residual distributions (predicted - chronological age) indicating error spread and bias. Among the machine learning models, PS-GPR demonstrates the strongest agreement with chronological age and the lowest error and SPARROW outperforming SynthBA.

Figure 4: Sex-stratified performance of all models.



Predicted brain age is plotted against chronological age for male (top row) and female (bottom row) participants, with the identity line (dashed) indicating perfect agreement. Across both sexes, PS-GPR and SPARROW show stronger alignment with the identity line, reflecting better predictive accuracy. Overall, model performance is consistent across sexes, though slightly improved agreement is observed in females for most models.

References

1. Lee, J. *et al.* Deep learning-based brain age prediction in normal aging and dementia. *Nat Aging* **2**, 412–424 (2022).

2. Franke, K. & Gaser, C. Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? *Front. Neurol.* **10**, (2019).

3. Liem, F. *et al.* Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage* **148**, 179–188 (2017).

4. Cole, J. H. *et al.* Brain age predicts mortality. *Mol Psychiatry* **23**, 1385–1392 (2018).

5. Cole, J. H. & Franke, K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in Neurosciences* **40**, 681–690 (2017).

6. Lee, J. & Kim, H.-J. Normal Aging Induces Changes in the Brain and Neurodegeneration Progress: Review of the Structural, Biochemical, Metabolic, Cellular, and Molecular Changes. *Front Aging Neurosci* **14**, 931536 (2022).

7. Franke, K., Ziegler, G., Klöppel, S. & Gaser, C. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage* **50**, 883–892 (2010).

8. Gravina, M., Pontillo, G., Shawa, Z., Cole, J. H. & Sansone, C. Assessing demographic bias in brain age prediction models using multiple deep learning paradigms. *Pattern Recognition Letters* **199**, 246–253 (2026).

9. Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C. & Mechelli, A. Machine learning for brain age prediction: Introduction to methods and clinical applications. *eBioMedicine* **72**, 103600 (2021).

10. Alijanpourotaghsara, A. *et al.* Morphometric Analysis of Neocortical and Infratentorial Structures: Genetic and Environmental Insights from a Twin Neuroanatomical Study. *Medicina* **61**, 261 (2025).

11. Joshi, A. A. *et al.* The contribution of genes to cortical thickness and volume. *Neuroreport* **22**, 101–105 (2011).

12. Kweon, H. *et al.* Human brain anatomy reflects separable genetic and environmental components of socioeconomic status. *Sci Adv* **8**, eabm2923 (2022).

13. Antal, A. *et al.* Imaging artifacts induced by electrical stimulation during conventional fMRI of the brain. *NeuroImage* **85**, 1040–1047 (2014).

14. Johnston, A. L. *et al.* At the intersection of place and race: Neighborhood and racialized exposures associated with brain age gap in midlife. *Brain and Environment* **8**, 100027 (2026).

15. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. & Chin, M. H. Ensuring Fairness in Machine Learning to Advance Health Equity. *Ann Intern Med* **169**, 866–872 (2018).

16. Atama, C. S., Eze, O. J. & Asogwa, U. S. Population data and resource allocation in Nigeria: Evidences from the challenges of distributing COVID-19 palliatives. *Social Sciences & Humanities Open* **10**, 101186 (2024).

17. Vitalis Jafla Pontianus & Oruonye E.D. The Nigerian population: A treasure for national development or an unsurmountable national challenge. *Int. J. Sci. Res. Arch.* **2**, 136–142 (2021).

18. Umoren, C., Adewole, M., Oladele, A., Oyeniran, O. & Anazodo, U. Mapping Neuroimaging Capacity in Africa. *Alzheimers Dement* **21**, e108454 (2025).

19. Anazodo, U. C., Adewole, M. & Dako, F. AI for Population and Global Health in Radiology. *Radiol Artif Intell* **4**, e220107 (2022).

20. The BraTS-Africa Dataset: Expanding the Brain Tumor Segmentation Data to Capture African Populations | Radiology: Artificial Intelligence. https://pubs.rsna.org/doi/10.1148/ryai.240528.

21. Iglesias, J. E. *et al.* Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. *NeuroImage* **237**, 118206 (2021).

22. FreeSurfer - PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC3685476/.

23. recon-all - Free Surfer Wiki. https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all.

24. DESTRIEUX, C., FISCHL, B., DALE, A. & HALGREN, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).

25. Esteban, O. *et al.* MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One* **12**, e0184661 (2017).

26. Sidhant Chopra *et al.* Transdiagnostic Connectome Project. OpenNeuro https://doi.org/10.18112/OPENNEURO.DS005237.V1.1.3 (2025).

27. FSLeyes. https://zenodo.org/records/14933238.

28. Freesurfer recon-all Euler number. https://uw-chn.github.io/CHN-User-Manual/flywheel/freesurfer-recon-all/#euler-number.

29. Rosen, A. F. G. *et al.* Quantitative assessment of structural image quality. *NeuroImage* **169**, 407–418 (2018).

30. Da Costa, P. F., Dafflon, J. & Pinaya, W. H. L. Brain-Age Prediction Using Shallow Machine Learning: Predictive Analytics Competition 2019. *Front. Psychiatry* **11**, (2020).

31. Gutierrez Becker, B., Klein, T., Wachinger, C., & Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging Biomarkers and Lifestyle flagship study of ageing. Gaussian process uncertainty in age estimation as a measure of brain abnormality. *Neuroimage* **175**, 246–258 (2018).

32. Kalc, P., Dahnke, R., Hoffstaedter, F. & Gaser, C. BrainAGE: Revisited and reframed machine learning workflow. *Hum Brain Mapp* **45**, e26632 (2024).

33. Ziegler, G., Ridgway, G. R., Dahnke, R. & Gaser, C. Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. *Neuroimage* **97**, 333–348 (2014).

34. Zhang, B., Zhang, S., Feng, J. & Zhang, S. Age-level bias correction in brain age prediction. *NeuroImage: Clinical* **37**, 103319 (2023).

35. Beheshti, I., Nugent, S., Potvin, O. & Duchesne, S. Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme. *NeuroImage: Clinical* **24**, 102063 (2019).

36. Lange, A.-M. G. de & Cole, J. H. Commentary: Correction procedures in brain-age prediction. *NeuroImage : Clinical* **26**, 102229 (2020).

37. Cole, J. james-cole/PyBrainAge. (2025).

38. Rutherford, S. *et al.* Charting brain growth and aging at high spatial precision. *eLife* **11**, e72904 (2022).

39. ENIGMA. https://enigma.ini.usc.edu/.

40. Han, L. K. M. *et al.* Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol Psychiatry* **26**, 5124–5139 (2021).

41. Puglisi, L. *et al.* SynthBA: Reliable Brain Age Estimation Across Multiple MRI Sequences and Resolutions. in *2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)* 559–564 (2024). doi:10.1109/MetroXRAINE62247.2024.10796114.

42. Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A. & Smith, S. M. Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis* **68**, 101871 (2021).

43. Rajabli, R., Soltaninejad, M., Fonov, V. S., Bzdok, D. & Collins, D. L. Brain Age Prediction: Deep Models Need a Hand to Generalize. *Hum Brain Mapp* **46**, e70254 (2025).

44. Rajabli, R. GodPascal/brain-age-prediction. (2026).

45. A labeled Clinical-MRI dataset of Nigerian brains | Scientific Data. https://www.nature.com/articles/s41597-025-04743-0.

46. Liang, H., Zhang, F. & Niu, X. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. *Hum Brain Mapp* **40**, 3143–3152 (2019).

47. Baecker, L. *et al.* Brain age prediction: A comparison between machine learning models using region- and voxel-based morphometric data. *Human Brain Mapping* **42**, 2332–2346 (2021).

48. Cole, J. H. *et al.* Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage* **163**, 115–124 (2017).

49. Isamah, N. *et al.* Variability in Frontotemporal Brain Structure: The Importance of Recruitment of African Americans in Neuroscience Research. *PLoS One* **5**, e13642 (2010).

50. Levakov, G., Rosenthal, G., Shelef, I., Raviv, T. R. & Avidan, G. From a deep learning model back to the brain—Identifying regional predictors and their relation to aging. *Human Brain Mapping* **41**, 3235–3252 (2020).

51. Dhamala, E. *et al.* Proportional intracranial volume correction differentially biases behavioral predictions across neuroanatomical features, sexes, and development. *Neuroimage* **260**, 119485 (2022).

52. van der Meer, D. *et al.* Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Mol Psychiatry* **25**, 3053–3065 (2020).

53. de Lange, A.-M. G. *et al.* Mind the gap: Performance metric evaluation in brain-age prediction. *Human Brain Mapping* **43**, 3113–3129 (2022).

54. Pricing Overview. *Google Cloud* https://cloud.google.com/pricing.

55. Joo, Y. *et al.* Brain age prediction using combined deep convolutional neural network and multi-layer perceptron algorithms. *Sci Rep* **13**, 22388 (2023).

56. Rehman, A. U. *et al.* SA-AVAE: A sex-Aware adversarial variational autoencoder network for biological brain age estimation in IoMT-Enabled multimodal neuroimages. *Internet of Things* **37**, 101893 (2026).

57. Sanford, N. *et al.* Sex differences in predictors and regional patterns of brain age gap estimates. *Hum Brain Mapp* **43**, 4689–4698 (2022).

Data Availability

The unprocessed and pre-processed (3D Synthesized) de-identified PS-GPR training and internal test dataset is being prepared for public release. In the interim, the dataset can be made available to researchers upon request.

Code Availability

The code used in this study can be made available to researchers upon request.

Supplementary Materials

https://docs.google.com/document/d/1yqc_5GmtE_xyfNpd-gztV3FfNsqjXa_mtY9-KaxBnlc/edit?usp=sharing

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.pdf