Investigating the validity and interpretability of longitudinal brain age underlying cognition in Asian children and older adults

Running title: Longitudinal brain age underlying cognition

Susan F. Cheng[1,2], Wan Lin Yue[1,2], Kwun Kei Ng[2], Xing Qian[2], Siwei Liu[2], Trevor W.K. Tan[1,2], Kim-Ngan Nguyen[2], Ruth L.F. Leong[2], Saima Hilal[2,3,4], Ching-Yu Cheng[5,6], Evelyn C. Law[7,8], Peter D. Gluckman[7,9], Christopher Li-Hsian Chen[2,4], Michael J. Meaney[2,7,10], Michael W.L. Chee[2], B.T. Thomas Yeo[1,2,11,12*], and Juan Helen Zhou[1,2,11,*†]

[1]Integrative Sciences and Engineering Programme, NUS Graduate School, National University of Singapore, Singapore, Singapore
[2]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[3]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore
[4]Memory Aging and Cognition Centre, National University Health System, Singapore
[5]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore
[6]Duke-NUS Medical School, Singapore
[7]Singapore Institute for Clinical Sciences (SICS), A*STAR Research Entities (ARES), Singapore
[8]National University Health System, Singapore
[9]Liggins Institute, University of Auckland, Auckland, New Zealand
[10]Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada, and the Strategic Research Program, A*STAR Research Entities (ARES), Singapore
[11]Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore
[12]N.1 Institute for Health & Institute for Digital Medicine (WisDM), National University of Singapore, Singapore
[*]Joint senior authors
[†]Corresponding author: helen.zhou@nus.edu.sg

Key words: brain age, longitudinal, development, aging, machine learning, T1-weighted MRI

ABSTRACT

Brain age has emerged as a powerful tool to understand neuroanatomical aging and its link to health outcomes like cognition. However, most brain age models are trained and tested on cross-sectional data from primarily Caucasian, adult participants. It is thus unclear how well these models generalize to non-Caucasian participants, especially children. Furthermore, there is a lack of studies investigating the longitudinal change in brain age gap and its relationship to cognition. Here, we tested a previously published deep learning model on Singaporean elderly participants ($55 - 88$ years old) and children ($4 - 11$ years old). We found that the model directly generalized to the elderly participants, but model finetuning was necessary for children. After finetuning, we found that the longitudinal change in brain age gap was associated with future executive function performance in both elderly participants and children. We further found that lateral ventricles and frontal areas contributed to brain age prediction in elderly participants, while white matter and posterior brain regions were more important in predicting brain age of children. Taken together, our results suggest that there is potential for generalizing brain age models to diverse populations. Moreover, the longitudinal change in brain age gap reflects developing and aging processes in the brain, relating to future cognitive function.

1 INTRODUCTION

The human brain undergoes coordinated, multidimensional anatomical changes throughout the lifespan, which can be measured noninvasively by magnetic resonance imaging (MRI) [1]. These anatomical changes occur in parallel with age-related changes to other measurable phenotypes, such as cognition [2], [3]. Abnormal aging in late life and abnormal development in early life have both been implicated with increased risk of neuropsychiatric disorders [4], [5]. Thus, efforts have been made to quantify the heterogenous effects of aging/development on the brain through the concept of "brain age." Brain age uses machine learning to predict age from neuroimaging data. A higher brain age suggests more advanced aging/development relative to one's chronological age. This helps summarize complex patterns into a single number that preserves individual variations [6].

Historically, most brain age studies first use specialized software to preprocess MRI data and extract features such as gray matter volume, cortical thickness, or surface area. A machine learning model is then trained to predict age from the extracted features. The model is typically trained using cognitively normal participants, with chronological age acting as the ground truth. The model is then applied to new participants to predict their brain age [7]–[10].

More recently, deep learning models have gained popularity over traditional machine learning models for brain age prediction [11]–[16]. Unlike previous machine learning methods, deep learning models can learn relevant features directly from the unprocessed (or minimally processed) MRI scan. This reduces the need for specialized preprocessing to extract features, which is time-consuming, requires expert knowledge, and involves laborious quality control. This also allows deep learning models to train on increasingly numerous and heterogeneous data. Some of these pretrained models have been made publicly available and shown impressive generalization performance on completely unseen test data [14], [15]. However, both training and testing data still primarily consist of Caucasian participants, which could bias the models [17], [18]. Deep learning models can use finetuning (i.e. transfer learning) to help overcome this bias and achieve good performance even in small datasets [16]. But to our knowledge, previous work has not examined the performance, with and without finetuning, of a Caucasian-centric model on non-Caucasian children and elderly participants, such as Asian children and elderly participants. This generalization is important to establish as the prevalence of developmental disorders [19] and dementia [20] are on the rise in Asia.

In addition to being able to predict age, predictions made by the model should show utility in relating to other phenotypes of interest [21]. The deviation from expected aging is often quantified as the brain age gap (BAG; also referred to as brainPAD, brainAGE, etc.), which is calculated by subtracting chronological age from brain age. The BAG has shown broad associations with brain disorders [9], risk of mortality [8], and cognitive function [22], to name a few (see [23], [24] for recent reviews). However, there are relatively few longitudinal studies in healthy participants. Previous work found associations with early life or cross-sectional measures [22], [25]–[27], but there was only a weak link to future age-related cognitive decline, which did not survive multiple comparison correction [22]. Notably, these studies only used brain age measured at

	Elderly		Children
Characteristic	**EDIS (N = 694)**	**SLABS (N = 215)**	**GUSTO (N = 678)**
Age (years)	69.91 ± 6.46 $(60-88)$	68.17 ± 6.77 $(55-85)$	5.85 ± 1.68 $(4.2-11.3)$
Sex (M/F)	340/354	101/114	346/332
Ethnicity (C/M/I/O)	276/184/234/0	215/0/0/0	370/187/120/1*
Education (years)	6.18 ± 4.63 $(0-22)$	12.02 ± 3.45 $(0-21)$	N/A
MMSE score	24.13 ± 3.59 $(10-30)$	28.29 ± 1.27 $(26-30)$	N/A
Imaging follow up (years)	N/A	4.00 ± 3.33 $(0-9.59)$	3.49 ± 2.41 $(0-6.69)$
Cognition sample size	N = 694	N = 81 to 212	N = 217 to 239

Table 1: Participant characteristics at baseline. EDIS was cross-sectional, while SLABS and GUSTO were longitudinal. Reported as mean ± standard deviation (range). *: GUSTO ethnicities were based on the mother. **Key:** M/F - Male/Female; C/M/I/O - Chinese/Malay/Indian/Other; MMSE – Mini-Mental State Examination; EDIS – Epidemiology of Dementia in Singapore; SLABS – Singapore Longitudinal Aging Brain Study; GUSTO - Growing Up in Singapore Towards healthy Outcomes

one time point. There is evidence that cross-sectional and longitudinal brain measures may reflect different factors and predispositions of individuals [28], suggesting combining them could provide more predictive power. However, to our knowledge, the additional utility of longitudinal changes in brain age has not been tested in healthy participants.

Thus, in this work, we leverage a state-of-the-art deep learning brain age model trained on over 30,000 individuals across the lifespan [14] to test generalizability to Asian elderly participants and children. We also finetune the model to explore how much predictions improve. We then examine the longitudinal utility of brain age in associating with future cognition. Finally, we investigate model interpretability using guided backpropagation. Our findings provide insight into the generalizability of brain age models and the importance of longitudinal measurements.

2 RESULTS

Figure 1 shows the study design. For the brain age model, we used the publicly available Simple Fully Convolutional Network (SFCN) pretrained on 34,285 T1 MRI scans from 21 non-overlapping datasets across the lifespan [14]. This pretrained model (a.k.a. pyment) was found to have the highest accuracy and test-retest reliability in a recent comparison of publicly available brain age models [29]. While featuring an unusually large and heterogenous training set for brain age prediction, there was still a relative lack of training data from very young and/or non-Caucasian participants.

Thus, to test generalizability to Asian elderly participants and children, we used three datasets from Singapore: 1) the cross-sectional Epidemiology of Dementia in Singapore (EDIS) study [30]–[32], consisting of 694 non-demented elderly (226 with no cognitive impairment (NCI) and 468 with cognitive impairment no dementia (CIND)); 2) the longitudinal Singapore Longitudinal Aging Brain Study (SLABS) [33], consisting of 215 healthy elderly participants; and 3) the longitudinal Growing Up in Singapore Towards healthy Outcomes (GUSTO) study [34], consisting of 678 healthy children. These datasets are detailed in the Methods (Section 5). Table 1 and Supplementary Table S1 summarize the participant demographic and cognitive characteristics.

2.1 Brain age predictions

We first input minimally preprocessed T1 scans directly to the pretrained model (all baseline and follow-up data). We also finetuned the model for our local datasets using 10-fold cross-validation (Figure 1A,B; Section 5.2). Figure 2 shows the brain age predictions for the pretrained and finetuned models on all datasets.

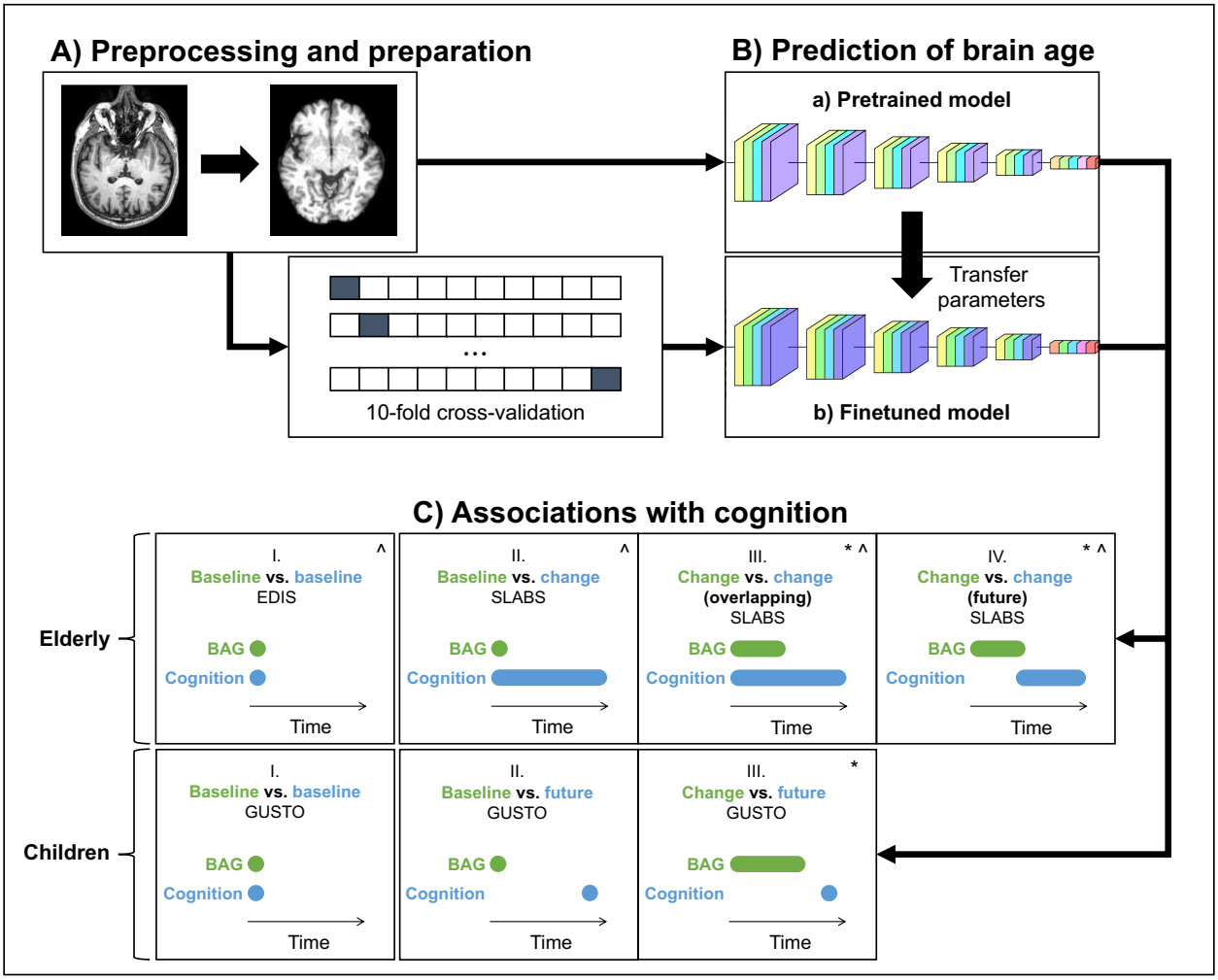


Figure 1: Study design schematic. **(A, B)** T1 MRI scans were minimally preprocessed according to the SFCN pipeline [14]. These were a) directly input into the pretrained brain age model, or b) split into 10 cross-validation folds to finetune the model. The finetuned model transferred the weights from the pretrained model for initialization. All layers were then retrained. Age predictions were obtained on the test folds. BAG was calculated by subtracting chronological age from predicted age. Model interpretability was interrogated using guided backpropagation. **(C)** Cross-sectional and longitudinal association of BAG and cognitive performance were tested using multiple linear regression models in both elderly and children. Time intervals for BAG and cognition, based on data availability, are shown schematically. Annual rate of change was calculated from a linear regression with time for each participant. All models included chronological age and sex as covariates. ^: models for elderly also included years of education as a covariate; *: models with (annual rate of) change in BAG also included baseline BAG as a covariate. **Key:** EDIS – Epidemiology of Dementia in Singapore; SLABS – Singapore-Longitudinal Aging Brain Study; GUSTO - Growing Up in Singapore Towards healthy Outcomes; BAG – brain age gap

In EDIS and SLABS (elderly), the pretrained model performed well, as evidenced by the high correlation ($r = 0.7389$ for EDIS and $r = 0.8136$ for SLABS) and low MAE (MAE = 3.9895 for EDIS and MAE = 3.4668 for SLABS; Figure 2A, first two rows). After finetuning, correlations and MAEs slightly improved ($r = 0.7445$ for EDIS and $r = 0.8138$ for SLABS; MAE = 3.3232 for EDIS and MAE = 3.2653 for SLABS; Figure 2B, first two rows), but the predictions were generally similar to those made by the pretrained model (correlation between finetuned and pretrained predictions = 0.9143 for EDIS and 0.9231 for SLABS).

In contrast, the pretrained model did not perform as well in GUSTO (children). The MAE was lower than in elderly (MAE = 2.57), but the age range of GUSTO was also much smaller. Importantly, predictions did not distinguish among younger ages, leading to a low correlation ($r = 0.5426$; Figure 2A, last row). After finetuning, the correlation and MAE drastically improved ($r = 0.9411$; MAE = 0.6286; Figure 2B, last row). The variance in predicted ages also increased as chronological age increased (Supplementary Figure S1). Unlike EDIS and SLABS, the finetuned predictions were not similar to the pretrained predictions (correlation = 0.5732).

2.2 Associations with cognition

2.2.1 Elderly

To validate the brain age model with a large and cognitively heterogeneous sample, we first tested cross-sectional associations in EDIS (N = 694) using multiple linear regression models (Figure 1C). We included chronological age, sex, and years of education as covariates. Higher baseline BAG was broadly associated with lower baseline cognitive performance (i.e. negative associations; Supplementary Table S4). The associations were significant after multiple comparison correction for global cognition ($p_{corr} = 0.0006$), executive function ($p_{corr} = 0.0076$, Figure 3A), language ($p_{corr} = 0.0047$), visuomotor speed ($p_{corr} = 0.0136$), visuoconstruction ($p_{corr} = 0.0136$), verbal memory ($p_{corr} = 0.0034$), and visual memory ($p_{corr} = 0.0002$). The association was not significant for attention ($p_{corr} = 0.2461$). These results were consistent after finetuning (Supplementary Figure S2A and Supplementary Table S5). Similar broad negative associations were also observed in SLABS at baseline (N = 212), but these were not significant (Supplementary Tables S6 & S7).

To investigate longitudinal utility of brain age in healthy elderly, we tested associations in a longitudinal subset of SLABS (N = 81) using similar multiple linear regression models. We first related baseline BAG and early change in BAG to long-term cognitive change (Figure 1C; Methods Section 5.3). Baseline BAG generally failed to show associations with longitudinal cognitive changes (Supplementary Table S8). While higher baseline BAG was associated with faster long-term decline in executive function, it was not significant after multiple comparison correction ($p = 0.0406$, $p_{corr} = 0.2433$, Figure 3B). On the other hand, the early rate of BAG change was negatively associated with long-term rate of executive function change ($p = 0.0017$, $p_{corr} = 0.0100$, Figure 3C). This negative association held after removing the temporal overlap between BAG and cognition, looking only at the future rate of executive function change ($p = 0.0033$, Figure 3D). Notably, these associations were independent of baseline BAG, chronological age, sex, and years of education. The associations were also specific to executive function (Supplementary Table S10). Results were again consistent after finetuning (Supplementary Figure S2B-D and Supplementary Tables S9 & S11).

2.2.2 Children

To extend our analyses to healthy children, we tested cross-sectional and longitudinal associations in GUSTO using multiple linear regression models similar to above (Figure 1C; Methods Section 5.3). Since longitudinal cognitive data was not available for GUSTO, we used the cognitive scores themselves instead of the change. Furthermore, since finetuning the model drastically improved prediction accuracy in GUSTO, we used the finetuned predictions for our main analyses. We did not find a significant association between baseline BAG and baseline IQ score (p = 0.3809, Supplementary Figure S3B). Similarly, we did not find significant associations between baseline BAG and future cognitive scores (at 8.5 years old; $p \geq 0.4086$, Figure 4A, Supplementary Table S13). However, the early rate of BAG change (from 4.5 to 7.5 years old) was positively associated with future inhibition (at 8.5 years old; $p = 0.0103$, $p_{corr} = 0.0411$, Figure 4B). The early rate of BAG change was also positively associated with future switching, but it was not significant after correcting for multiple comparisons ($p = 0.0221$, $p_{corr} = 0.0663$, Supplementary Table S15). These associations were independent of baseline BAG, chronological age, and sex. Notably, in contrast to older adults, the direction

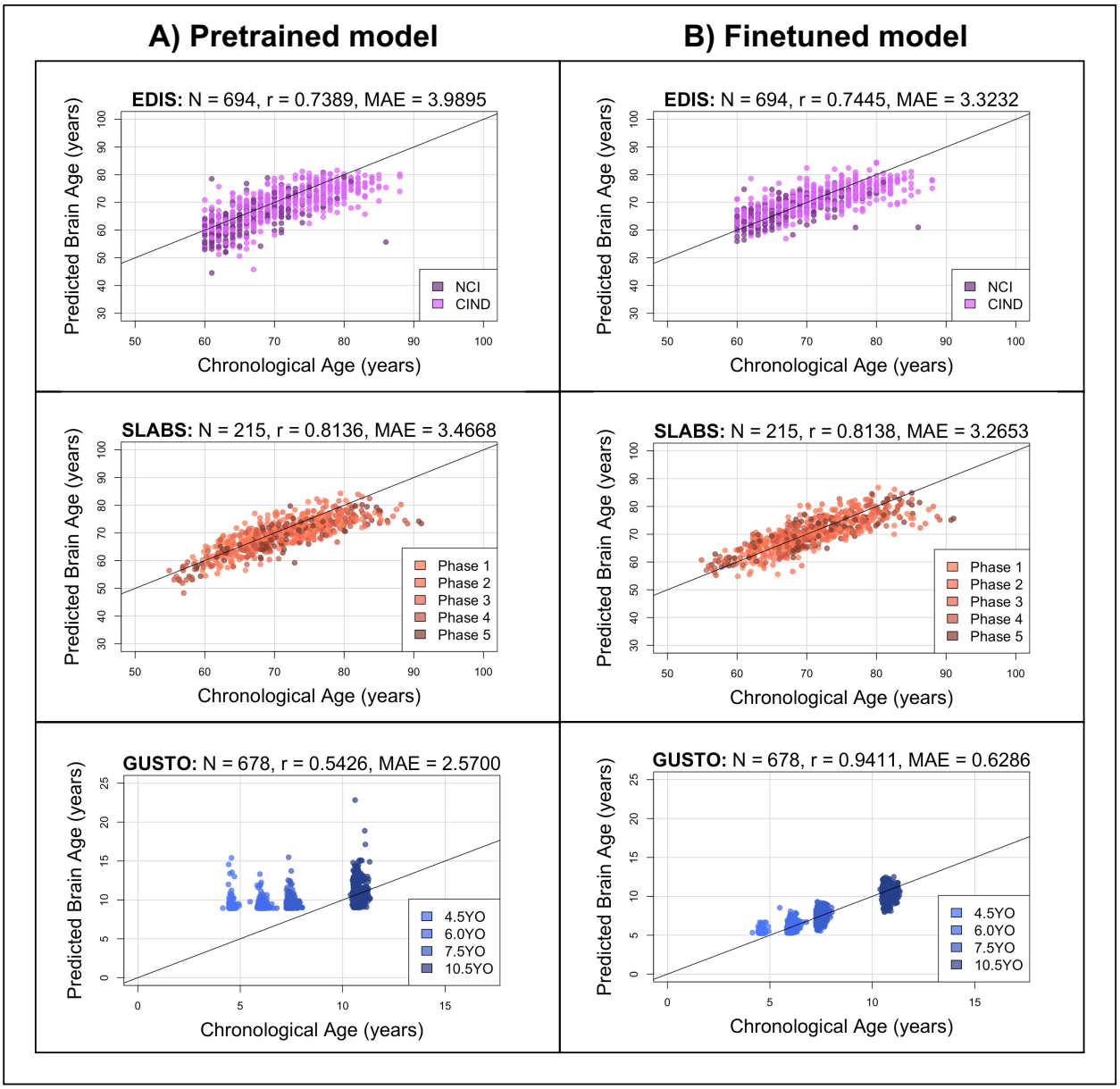


Figure 2: The pretrained brain age model performs well in elderly, while the finetuned model performs well in both elderly and children. Black identity lines representing perfect prediction are included for reference. **(A)** Predicted brain ages from the **pretrained** model are plotted against chronological age. They are highly correlated for EDIS and SLABS (elderly), but not GUSTO (children). **(B)** Predicted brain ages from the **finetuned** model are plotted against chronological age. They are highly correlated in all three datasets. **Key:** EDIS – Epidemiology of Dementia in Singapore; SLABS – Singapore-Longitudinal Aging Brain Study; GUSTO - Growing Up in Singapore Towards healthy Outcomes; N – Number of participants; r – Pearson's correlation coefficient; MAE – Mean Absolute Error; NCI – No Cognitive Impairment; CIND – Cognitive Impairment No Dementia

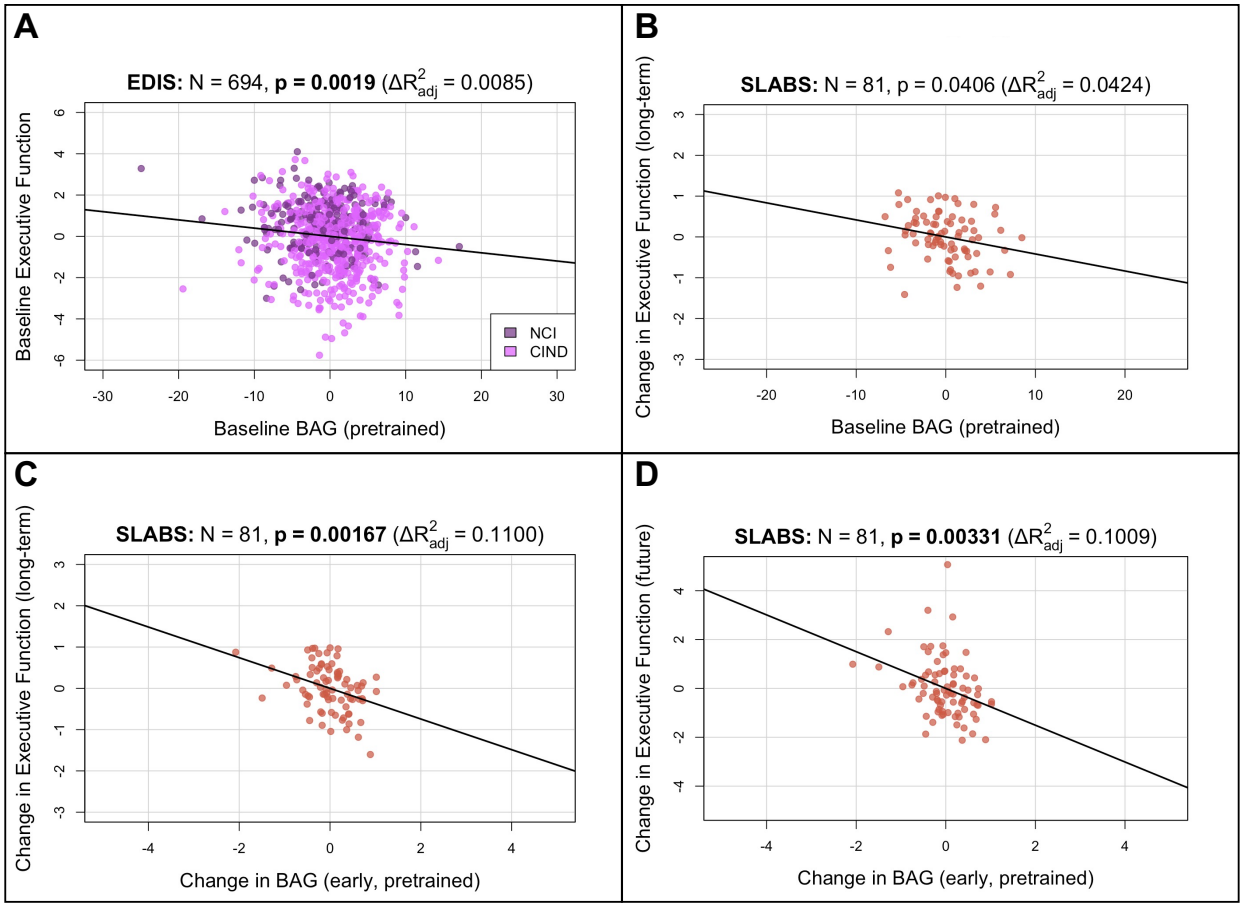


Figure 3: Brain age gap from the **pretrained** model is negatively associated with executive function in **elderly**. Bolded p-values indicate significance after Holm-Bonferroni correction ($p_{corr} < 0.05$). ΔR^2_{adj} indicates the change in adjusted R^2 after adding the variable of interest. All models include chronological age, sex, and years of education as covariates. Models with change in BAG also include baseline BAG as a covariate. Results are similar after finetuning (Supplementary Figure S2) **(A)** Partial regression plot between baseline BAG and executive function in EDIS, colored by cognitive status. A significant negative association is observed. **(B)** Partial regression plot between baseline BAG and long-term rate of change in executive function (mean follow-up time $= 7.8 \pm 1.0$ years) in SLABS. A negative association is observed, but it is not significant after correcting for multiple comparisons. **(C)** Partial regression plot of early longitudinal rate of change in BAG (mean follow-up time $= 3.6 \pm 0.8$ years) when added to the model in (B). A significant negative association and increase in R^2 is observed. **(D)** Partial regression plot as in (C), but with future rate of change in executive function (mean follow-up time $= 4.2 \pm 1.1$ years), removing the overlap with early change in BAG. A significant negative association is again observed. **Key:** N – number of participants; p – p-value for variable of interest (x-axis); ΔR^2_{adj} - change in adjusted R^2 when adding variable of interest; BAG – Brain Age Gap; NCI – No Cognitive Impairment; CIND – Cognitive Impairment No Dementia; EDIS – Epidemiology of Dementia in Singapore; SLABS – Singapore-Longitudinal Aging Brain Study

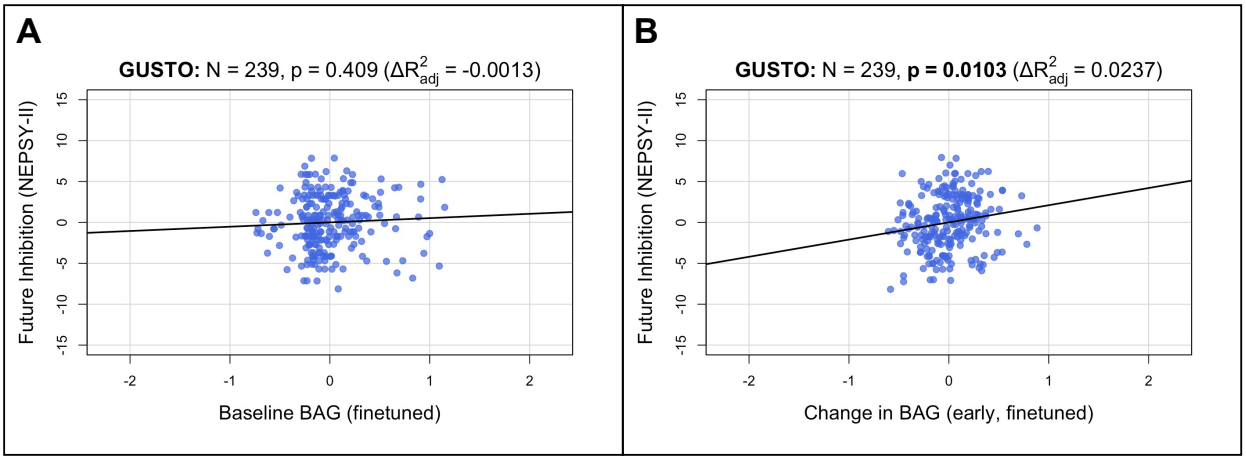


Figure 4: Longitudinal brain age gap from the **finetuned** model is positively associated with inhibition in **children**. Bolded p-values indicate significance after Holm-Bonferroni correction ($p_{corr} < 0.05$). ΔR^2_{adj} indicates the change in adjusted R^2 after adding the variable of interest. All models include chronological age and sex as covariates. Models with change in BAG also include baseline BAG as a covariate. **(A)** Partial regression plot between baseline BAG (calculated from 4.5 or 6.0 years old) and future NEPSY-II inhibition scaled subscore (measured at 8.5 years old). No significant association is observed. **(B)** Partial regression plot of early longitudinal rate of change in BAG calculated from 4.5 to 7.5 years old (mean follow-up time $= 2.4 \pm 0.7$ years) when added to the model in (A). A significant positive association and increase in R^2 is observed. **Key:** N – number of participants; p – p-value for variable of interest (x-axis); ΔR^2_{adj} - change in adjusted R^2 when adding variable of interest; BAG – Brain Age Gap; GUSTO - Growing Up in Singapore Towards healthy Outcomes

of association was now positive, meaning increased early rate of BAG change was associated with better future executive function performance. There were no significant associations using the pretrained model (Supplementary Figures S3A & S4 and Supplementary Tables S12 & S14).

2.3 Model interpretability

Finally, to investigate model interpretability, we used guided backpropagation [35] to derive group-level saliency maps for brain age prediction (Methods Section 5.4). Figure 5 shows the top 10% of contributing voxels to age prediction in four representative slices (left) for all datasets. Full 3D maps will also be made available online. Relative contributions of white and gray matter features across the whole brain are shown on the right. Areas near the lateral ventricles are labeled in red, while areas more prominent in elderly than children are labeled in magenta, and areas more prominent in children than elderly are labeled in blue.

Both EDIS and SLABS show similar profiles (Figure 5A&B), suggesting important features are stable across the elderly datasets. Regions near the lateral ventricles all make strong contributions, making up 7 of the 15 highest ranking features. Substantial contributions can also be seen in frontal/association areas corresponding to the default mode, control, and salience/ventral attention networks. Areas near the hippocampus/fornix, thalamus, and somatomotor network also contribute. These findings are consistent when using the pretrained model for EDIS and SLABS (Supplementary Figure S5A&B).

In contrast, with the children of GUSTO, notable differences can be found (Figure 5C). While the fornix is still the strongest contributor, it and other anterior areas near the ventricles (genu and body of corpus callosum, caudate) do not contribute as much. The overall prominence of white matter is also increased, especially in the brainstem (corticospinal tract and pontine crossing tract) and posterior regions (sagittal stratum, superior longitudinal fasciculus, and posterior limb of internal capsule). Furthermore, gray matter networks generally decrease in prominence, except the visual and limbic networks, which increase in prominence. The hippocampus, amygdala, and thalamus continue to make substantial contributions. Unlike in elderly, these features are not consistent when using the pretrained model (Supplementary Figure S5C).

3 DISCUSSION

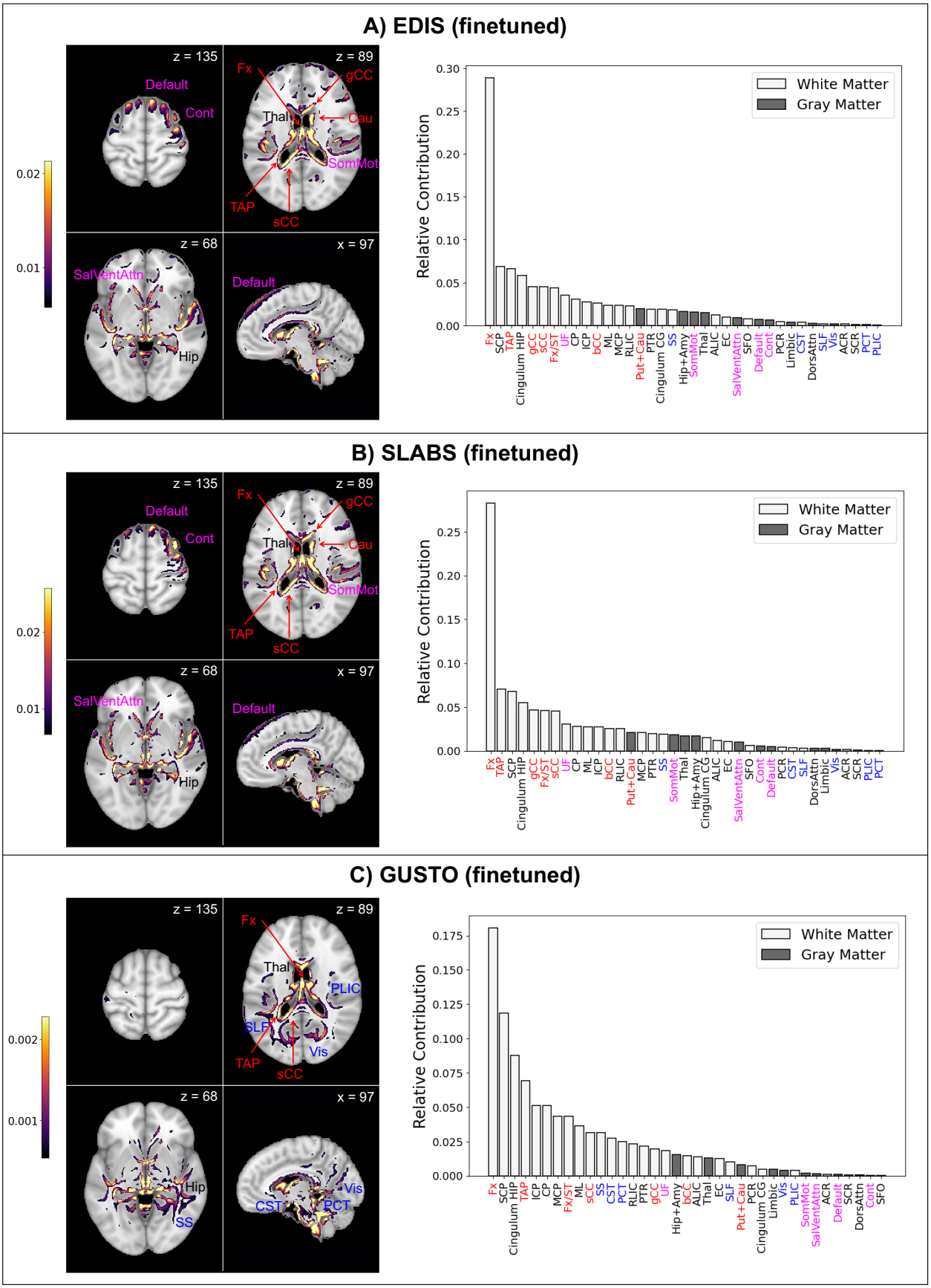
Our findings are the first, to our knowledge, to show the age-dependent generalizability of a pretrained brain age model to non-Caucasian participants - specifically Singaporean children and elderly. We also present novel results on the informativeness of longitudinal changes in brain age, independent of baseline brain age, to future executive function in healthy participants. Finally, we show that accurate models focus on distinct features in elderly and children, suggesting that the brain age model can extract relevant age-related information.

3.1 Generalizability of pretrained brain age models to local datasets may be age-dependent

Overall, our results suggest the pretrained SFCN model could be directly applied to Singaporean elderly participants, but it needed to be finetuned for Singaporean children. Previous work with the SLABS dataset showed that aging-related changes in Chinese Singaporean elderly were comparable to previous studies conducted with primarily Caucasian participants [36]. However, it was not initially clear whether this would carry over to a multidimensional index like brain age or to elderly datasets that better reflect the ethnic diversity of Singapore. Encouragingly, we found high accuracy in predicting age (i.e. low MAE and high correlation) in both EDIS (Chinese, Malay, and Indian participants) and SLABS (Chinese participants only) using the pretrained model. Furthermore, our results after finetuning were generally consistent with the original findings in elderly. This suggests that similarities were not specific to the SLABS sample, but could generalize to Singaporean elderly as a whole. In addition to similar aging patterns, the success of the pretrained model in this age range can be attributed to the abundance of training data around $60-80$ years old, mostly from UKBiobank [14].

However, we found the pretrained model did not perform as well in children (i.e. low correlation). While the MAE was actually lower in children than elderly, this was likely due to the smaller age range [37]. After finetuning, the MAE reduced dramatically, further demonstrating the inadequacy of the pretrained model in this case. Previous work has indicated that brain structural differences between Chinese Singaporeans and non-Asian Americans may be more pronounced in young adults than elderly [38]. This could conceivably extend to childhood and other ethnicities. However, another important factor is the model training age distribution, which only included 147 participants 5 years old or younger and had its earliest (and smallest) peak around 10 years old [14]. Notably, the pretrained model tended to predict all GUSTO ages around 10

Figure 5 *(following page)*: **Finetuned** brain age models focus on distinct features in children and elderly. The top 10% of features are shown for four representative brain slices on the left. Relative contributions for gray and white matter features across the whole brain are shown on the right. Regions near the lateral ventricles are labeled in red. Features more prominent in elderly than children are labeled in magenta, while features more prominent in children than elderly are labeled in blue. Features and relative contributions are generally consistent between **(A)** EDIS and **(B)** SLABS, but key differences can be seen in **(C)** GUSTO. **Key:** EDIS – Epidemiology of Dementia in Singapore; SLABS – Singapore-Longitudinal Aging Brain Study; GUSTO - Growing Up in Singapore Towards healthy Outcomes; MCP – Middle cerebellar peduncle; PCT – Pontine crossing tract; gCC – Genu of corpus callosum; bCC – Body of corpus callosum; sCC – Splenium of corpus callosum; Fx – Fornix (column and body); CST – Corticospinal tract; ML – Medial lemniscus; ICP – Inferior cerebellar peduncle; SCP – Superior cerebellar peduncle; CP – Cerebral Peduncle; ALIC – Anterior limb of internal capsule; PLIC – Posterior limb of internal capsule; RLIC – Retrolenticular part of internal capsule; ACR – Anterior corona radiata; SCR – Superior corona radiata; PCR – Posterior corona radiata; PTR – Posterior thalamic radiation; SS – Sagittal stratum; EC – External capsule; Cingulum CG – Cingulum (cingulate gyrus); Cingulum HIP – Cingulum (hippocampus); Fx/ST – Fornix (cres) / Stria terminalis; SLF – Superior longitudinal fasciculus; SFO – Superior fronto-occipital fasciculus; UF – Uncinate fasciculus; TAP – Tapetum; Vis – Visual network; SomMot – Somatomotor network; DorsAttn – Dorsal attention network; SalVentAttn – Salience/Ventral attention network; Limbic – Limbic network; Cont – Control/frontoparietal network; Default – Default mode network; Hip+Amy – Hippocampus + amygdala; Put+Cau – Putamen + caudate; Thal – Thalamus


A) EDIS (finetuned)
z = 135
z = 89
z = 68
x = 97
Default
Cont
Fx
gCC
Thal
Cau
SomMot
TAP
sCC
SalVentAttn
Hip
0.02
0.01
Relative Contribution
White Matter
Gray Matter
Fx SCP TAP Cingulum HIP gCC sCC Fx/ST UF CP ICP bCC ML MCP RLIC Put+Cau PTR Cingulum CG SS Hip+Amy SomMot Thal ALIC EC SalVentAttn SFO Default Cont PCR Limbic CST DorsAttn SLF Vis ACR SCR PCT PLIC
B) SLABS (finetuned)
Fx TAP SCP Cingulum HIP gCC Fx/ST sCC UF CP ML ICP bCC RLIC Put+Cau MCP PTR SS SomMot Thal Hip+Amy Cingulum CG ALIC EC SalVentAttn SFO Cont Default PCR CST SLF DorsAttn Limbic Vis ACR SCR PLIC PCT
C) GUSTO (finetuned)
PLIC
SLF
Vis
CST
PCT
SS
Fx SCP Cingulum HIP TAP ICP CP MCP Fx/ST ML sCC SS CST PCT RLIC PTR gCC UF Hip+Amy bCC ALIC Thal EC SLF Put+Cau PCR Cingulum CG Limbic Vis PLIC SomMot SalVentAttn ACR Default SCR DorsAttn Cont SFO

years old, suggesting it may have been impacted by this imbalanced distribution.

Fortunately, finetuning the model produced distinct age groups, along with higher correlation and lower MAE. As discussed below, finetuning the model in children also shifted feature saliency and revealed a significant association with future executive function that was not found using the pretrained model. This suggests the model underwent a greater change in children, compared to elderly, to become both more accurate and meaningful. Furthermore, we found that the variance of finetuned predictions was the lowest at 4.5 years old and increased steadily with age, consistent with previous reports [39]. This implies the "brain maintenance" account of aging, where individuals start with the same or similar offsets, and different slopes result in increased variability over the lifespan [28]. This also suggests that the variance in brain age predictions at later ages is likely due to stable, lifelong factors as well as ongoing changes. Thus, looking at longitudinal changes in brain age could help separate these influences.

3.2 Longitudinal changes in brain age are informative of future executive function

Our results with baseline BAG were largely consistent with previous work in elderly. With a large sample of community-dwelling, non-demented participants from EDIS, we found significant associations with baseline cognition across multiple cognitive domains, consistent with a recent review [23]. Furthermore, with a smaller longitudinal sample of healthy participants from SLABS, we matched previous work finding an association with future decline in cognition, despite being not significant after multiple comparison correction [22]. In GUSTO children, we did not find a significant association between baseline BAG and baseline cognition at 4.5 years old or future cognition at 8.5 years old. To our knowledge, while brain age associations with cognition have been reported in samples spanning $3 - 22$ years old [40]–[43], they have not been explored around the early age of 4.5 years old specifically. This is notable since our generalization analyses revealed that, after finetuning, brain age variability is lowest at this age and increases with age. Thus, cross-sectional associations with cognition may only occur at later ages, when there is more variability in brain age. Alternatively, more complex models may be needed to reveal cross-sectional structure-cognition association at such young ages.

Although we generally did not find associations between baseline BAG and future cognition in healthy participants, the story was different when including the early rate of change in BAG. Due to the lack of available longitudinal cognitive data in GUSTO, we could only look at future cognition at a single time point instead of the rate of change. This likely introduced noise into our analyses as it did not account for the initial variation in cognition. The future cognitive scores in GUSTO were also all related to executive function, so we could not investigate other cognitive domains in children. Interestingly, within subdomains of executive function in children, we found the strongest association with inhibition, which has been proposed as a primary driver of domain-general executive function [44]. Thus, we found that early longitudinal changes in BAG associated with future executive function performance in both children and elderly, independent of baseline BAG. This is notable in light of recent evidence that early-life factors may affect cross-sectional brain measurements, but not longitudinal changes [45]. While the association was specific to executive function in elderly, it is presently not clear whether this was biased by the modest sample size of 6- to 10-year longitudinal data or whether brain age is particularly sensitive to this domain [23]. Previous work found associations between early- or mid-life factors and cross-sectional brain age [25]–[27], but associations with the longitudinal change of brain age were not investigated.

Our findings thus suggest that early-life factors could influence the cross-sectional brain age, but they are not the only influence throughout the lifespan. Early life factors may dominate the (relatively low) inter-individual variability in brain age at 4.5 years old. However, as normal development occurs in the next several years, children mature at different rates, leading to increasing variability in brain age predictions. This variance is also related to individual differences in executive function. In late life, these changes have accumulated to produce more variable brain age predictions. However, baseline brain age predictions do not associate with future cognitive performance, possibly since they reflect past factors. Information about ongoing changes in brain aging is needed to reveal associations with future rates of executive function decline. Taken together, these findings suggest that brain age, when measured longitudinally, can capture ongoing processes of healthy aging.

While early longitudinal changes in BAG associated with future executive function performance in both elderly and children, one notable difference lies in the direction of association. In elderly, increases in BAG

were associated with worse executive function decline. In children, increases in BAG were associated with better future inhibitory performance. There have been somewhat conflicting reports on the direction of association between BAG and cognition in youth [40]–[43]. However, our results are unique in examining the longitudinal change in BAG rather than the cross-sectional BAG. Thus, our results could reflect previously reported cognitive decline with increasing age in late life [2] and cognitive gains with increasing age in early life [3]. One of the few longitudinal studies in development related white matter and executive function development, and found that white matter growth in adolescence was associated with better inhibitory control, while growth in adulthood was associated with worse performance [46]. Our brain age paradigm, based on multivariate features of the brain, further support these findings in children and elderly. Specifically, a faster increase in BAG may imply that a child is developing ahead of schedule, resulting in more rapidly maturing cognitive functioning. Conversely, a slower increase in BAG at an older age may reflect mechanisms of brain maintenance at work, prolonging a more "youthful" brain and sustained optimal cognitive performance.

3.3 Salient features of the brain age model differ between elderly and children

Our work builds on recent efforts to interpret deep learning brain age models in aging [11]–[13]. While the datasets, brain age models, and interpretability methods all differed among these studies, the most consistent finding was the importance of the lateral ventricles in elderly. This was evident in our models as well. Like other popular methods for extracting feature importance from deep learning models, our guided backpropagation method tended to highlight boundaries between regions and tissue types (i.e. edges). Thus, strong contributions from white matter areas such as the fornix and corpus callosum were likely at least partly due to the size of the lateral ventricles. These regions all ranked highly in elderly, suggesting the overall importance of the lateral ventricles.

Our other findings in elderly also broadly align with prior research in aging. We find important contributions around subcortical regions and frontal/association areas that are observed to degenerate more prominently in aging [47]. Among areas near the lateral ventricles, the fornix particularly stands out as the strongest contributor. This could be due to its connections with the hippocampus, suggesting fornix contributions may also reflect age-related hippocampus atrophy. Fornix was previously the strongest contributor in brain age models focused on white matter derived from diffusion MRI, and it showed the highest absolute correlations with age [48]. Corpus callosum and cerebellar peduncle were also found to strongly contribute in a separate white matter brain age model [49]. Additionally, the importance of the thalamus, putamen and caudate, ventral attention network, and somatomotor network could indicate the importance of frontostriatal circuits. Frontostriatal changes have been proposed as a hallmark of healthy aging [47], and the role of these and related regions [50] in supporting executive function could underlie the observed association between BAG and executive function.

We also show clear differences in feature importance between elderly and children, in line with prior research in development. The consistency between elderly datasets reinforces that these differences are not simply artefacts of using a different dataset. Most strikingly, we find evidence of a posterior to anterior pattern [5], [51] going from childhood to elderly. For instance, posterior areas near the lateral ventricles (tapetum and splenium of corpus callosum) continue to rank highly, while more anterior areas decrease in prominence. We also find a general increase in the relative importance of white matter, with greater increases in posterior regions. This is in line with a previous brain age model showing stronger contributions of white matter relative to gray matter in youth [39]. The focus on the development of white matter could also underlie the observed association with executive function development [46], [52], [53]. Finally, we find increased contributions from the brain stem, which is consistent with its large volume changes in youth [54].

3.4 Limitations

Our study is not without limitations. While we find encouraging signs that the model generalizes to Singaporean elderly, we can not completely rule out more subtle issues that may have arisen from applying the model to these participants. For instance, finetuning the model slightly increased prediction accuracy and generally strengthened associations with cognition in elderly, suggesting the pretrained model may not have performed optimally. Furthermore, we have not tested the model in other non-Caucasian participants, which would be needed for a more comprehensive test of generalizability. The current study also only includes par-

ticipants from very early and late life. Thus, future work would be needed to extend our results across the lifespan, with more participants and even longer follow up times, in order to achieve a more complete picture.

4 Conclusion

Here, we used a previously published brain age model to reveal age-dependent generalization to Asian participants, as well as age-dependent associations and interpretability of brain age. Specifically, we found the brain age model could be directly applied to Singaporean older adults, but it needed to be finetuned for Singaporean children. Furthermore, longitudinal changes in brain age were related to future executive function in both children and elderly participants. However, the direction of association was positive in children and negative in elderly. Combined with the identified salient features for brain age prediction, we conclude that increased brain age in early life could indicate more mature development, especially in white matter and posterior areas. Conversely, increased brain age in late life could suggest greater degeneration, especially around the lateral ventricles and frontal areas. Our results provide early evidence of the generalization capability of the brain age model and the ability of longitudinal measurements to capture ongoing aging process in the brain.

5 METHODS

5.1 Sample characteristics

We analyzed three datasets from Singapore, which are detailed in the following.

5.1.1 Participants

EDIS EDIS was a cross-sectional study to measure the prevalence of cognitive impairment and dementia in Singapore, which has been described previously [30]–[32]. We analyzed T1 MRI and cognitive data from 694 community-dwelling older adults. The same participants were used for all analyses. Ethics approval for the EDIS study was obtained from the Singapore Eye Research Institute and the National Healthcare Group Domain Specific Review Board. The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained, in the preferred language of participants, by bilingual study coordinators prior to recruitment into the study.

SLABS SLABS was a longitudinal, community-based study to characterize age-related brain changes and cognitive performance in healthy elderly in Singapore, which has been described previously [33]. Participants underwent at most 5 phases of neuroimaging and neuropsychological assessments at approximately 2-year intervals. Neuropsychological assessments were performed within 3 months of neuroimaging. To test prediction accuracy of the brain age model, we first used 598 T1 scans from N = 215 participants with MMSE score ≥ 26 at baseline (mean follow-up time = 4.0 ± 3.3 years). To investigate longitudinal associations in healthy elderly, we identified a subset of N = 81 participants with: (1) longitudinal T1 and cognitive data in the first three phases; (2) additional cognitive data in the last two phases (to study future cognitive decline, see Figure 1C); and (3) MMSE score ≥ 26 at baseline. Thus, mean follow-up time in this subset was 3.6 ± 0.8 years for T1 scans (total number = 228), while mean follow-up time was 7.8 ± 1.0 years for cognitive scores (total number = 355). The study was approved by the Institutional Review Board of the National University of Singapore. All participants provided written informed consent prior to participation.

GUSTO GUSTO is a longitudinal birth cohort study to characterize early development in Singapore, which has been described previously [34]. Participants were scanned at 4.5, 6.0, 7.5, and 10.5 years old. Neuropsychological assessments were taken at 4.5 and 8.5 years old. To test prediction accuracy of the brain age model in children, we first used 1,702 T1 scans from N = 678 normally developing children (mean follow-up time = 3.5 ± 2.4 years). To investigate cross-sectional and longitudinal associations in healthy children, we identified subsets of participants (N = 217 to 239) with the requisite imaging and cognitive

data available, similar to SLABS. For the cross-sectional analysis, this included participants with both T1 and cognitive data at 4.5 years old (N = 217). For the longitudinal analyses, this included participants with longitudinal T1 data from 4.5 to 7.5 years old and cognitive data at 8.5 years old (N = 220 or 239). The study was approved by the National Healthcare Group Domain Specific Review Board (NHG DSRB) and the Sing Health Centralized Institutional Review Board (CIRB). Written informed consent was obtained from mothers. When children reached 6 years of age, children also provided oral consent.

5.1.2 Neuropsychological assessments

EDIS Trained research psychologists administered a neuropsychological battery locally validated for Singaporean elderly, as described previously [30]. Briefly, the battery assessed the following seven cognitive domains using the corresponding tests: (1) Executive function: Frontal Assessment Battery and Maze Task; (2) Attention: Digit Span, Visual Memory Span, and Auditory Detection; (3) Language: Boston Naming Test and Verbal Fluency; (4) Visuomotor speed: Symbol Digit Modality Test and Digit Cancellation; (5) Visuoconstruction: Weschler Memory Scale-Revised Visual Reproduction Copy task, Clock Drawing, and Weschler Adult Intelligence Scale-Revised subtest of Block Design; (6) Verbal memory: Word List Recall and Story Recall; (7) Visual memory: Picture Recall and Weschler Memory Scale-Revised Visual Reproduction.

For each individual test, raw scores were transformed to standardized z-scores using the mean and standard deviation (SD) of that test (across all of EDIS, not just the imaging sample included here). Composite z-scores for each domain were obtained by averaging all individual test z-scores within that domain. These domain-specific z-scores were then standardized using their own mean and SD. A global cognition z-score was calculated by averaging over all domain-specific z-scores and standardized using its own mean and SD. CIND was defined as impairment in at least one cognitive domain using education adjusted cut-off values of 1.5 SDs below the established normal means on individual tests. Failure in at least half of the tests in a domain constituted failure in that domain. Note that CIND was not a formal clinical diagnosis.

SLABS Trained researchers administered neuropsychological assessments within 3 months of the neuroimaging scan, as described previously [36]. Briefly, the following five cognitive domains were assessed using the corresponding tests: (1) Executive function: Categorical Verbal Fluency Test and Design Fluency Test in the Delis-Kaplan Executive Function System[1]; (2) Attention: Digit Span Test and Spatial Span Test in Wechsler Memory Scale-Third Edition; (3) Processing speed: Symbol Digit Modalities Test, Symbol Search Task in the Wechsler Memory Scale-Third Edition, and Trail Making Test A; (4) Verbal Memory: Rey Auditory Verbal Learning Test; and (5) Visuospatial Memory: Visual Paired Associates Test. Composite T-scores (T-score = (z-score$\times$10)+50) were obtained for each domain and for global cognition following a similar procedure as EDIS.

GUSTO To maintain consistency with EDIS and SLABS, we selected standardized cognitive summary scores measured at 4.5 (baseline) and 8.5 (future) years old. These included the Kaufman Brief Intelligence Test Second Edition (KBIT-2) composite IQ standard score, the Wisconsin Card Sorting Test (WCST) total errors standard score, and the Developmental Neuropsychological Assessment Second Edition (NEPSY-II) scaled domain scores. The KBIT-2 was administered at 4.5 years and is a measure of abbreviated intelligence for children and adults aged 4 years to 90 years of age. The WCST is a lab-based measure of set-shifting/cognitive flexibility and was administered at age 8.5 years. The NEPSY-II was administered at 8.5 years and consisted of a word interference task requiring working memory recall (i.e. naming) and a Stroop task requiring predominantly inhibition in one condition and switching in another condition.

5.1.3 Image acquisition and preprocessing

EDIS MRI scans were performed on a 3T Siemens Magnetom Tim Trio System (Siemens, Erlangen, Germany) at the Clinical Imaging Research Centre, National University of Singapore. High-resolution T1-weighted structural MRI was acquired using magnetization-prepared rapid gradient echo (MPRAGE) sequence (192 continuous sagittal slices, TR/TE/TI = 2300/1.9/900 ms, flip angle = 9°, FOV = 256 $\times$ 256 mm^2, matrix = 256 $\times$ 256, isotropic voxel size = 1.0 $\times$ 1.0 $\times$ 1.0 mm^3, bandwidth = 240 Hz/pixel).

[1]Unlike previous studies, Trail Making Test B was not included in this study due to missing data in later phases.

SLABS For the first three phases, MRI scans were performed on a 3T Siemens Magnetom Tim Trio System (Siemens, Erlangen, Germany) at the Centre for Cognitive Neuroscience, Duke-NUS Medical School. High-resolution T1-weighted structural MRI was acquired using a MPRAGE sequence (192 continuous sagittal slices, TR/TE/TI = 2300/2.98/900 ms, flip angle = 9°, FOV = 256 × 240 mm^2, matrix = 256 × 240, isotropic voxel size = 1.0 × 1.0 × 1.0 mm^3, bandwidth = 240 Hz/pixel).

For the last two phases, following a scanner upgrade, MRI scans were performed on a 3T Siemens Magnetom Prisma Fit System (Siemens, Erlangen, Germany). High-resolution T1-weighted structural MRI was again acquired using a MPRAGE sequence (192 continuous sagittal slices, TR/TE/TI = 2300/2.28/900 ms, flip angle = 8°, FOV = 256 × 240 mm^2, matrix = 256 × 240, isotropic voxel size = 1.0 × 1.0 × 1.0 mm^3, bandwidth = 200 Hz/pixel).

GUSTO For scans taken at 4.5 and 6.0 years, MRI scans were performed on a 3T Siemens Magnetom Skyra System (Siemens, Erlangen, Germany) at KK Women's and Children's Hospital. High-resolution T1-weighted structural MRI was acquired using a MPRAGE sequence (192 continuous sagittal slices, TR/TE/TI = 2000/2.08/877 ms, flip angle = 9°, FOV = 192 × 192 mm^2, matrix = 192 × 192, isotropic voxel size = 1.0 × 1.0 × 1.0 mm^3).

For scans taken at 7.5 and 10.5 years, MRI scans were performed on a 3T Siemens Magnetom Prisma Fit System (Siemens, Erlangen, Germany) at the National University of Singapore. The scanning parameters were the same as for 4.5 and 6.0 years.

Preprocessing For all datasets, we used the minimal preprocessing pipeline performed on the SFCN training set, as described previously [14]. Briefly, images were first skull-stripped with FreeSurfer [55], then reoriented to standard FMRIB (Oxford Centre for Functional MRI of the Brain) Software Library (FSL) [56] orientation and linearly registered to Montreal Neurological Institute (MNI) 152 space using the FSL linear registration tool (FLIRT) [57]. Images were then cropped to 167 × 212 × 160 voxels, and voxel intensity values were normalized between 0 and 1. These minimally preprocessed images were input to the SFCN brain age model (Figure 1A). Similar to the original model [14], we adopted a lenient manual quality control before conducting analyses, removing scans where a significant portion of the brain was missing or there was a registration failure. This excluded 2 scans/participants from EDIS and 12 scans from 11 participants from GUSTO.

5.2 Brain age predictions

After preprocessing, we directly applied the pretrained brain age model [14] to each of the datasets (all baseline and follow-up data) to generate brain age predictions. We used the regression variant of SFCN due to its superior generalization performance [14]. Performance was assessed using Pearson's correlation and mean absolute error (MAE) between brain age and chronological age. The model was considered to have performed well if both correlation was high and MAE was low. BAG was calculated by subtracting chronological age from brain age.

We then finetuned the model on each dataset separately to mitigate effects from the domain shift (i.e. the change in distribution from training data to testing data). We used all scans from each dataset without any exclusions. We split scans into 10 cross-validation folds, where each participant was included in the testing set exactly once. Of the remaining data for each fold, 80% was used for training and 20% was used for validation. In the case of longitudinal data, it was ensured that all scans from the same participant were kept in either the training, validation or test set to avoid biased estimates. We used the pretrained model weights as initialization, then retrained all layers (Figure 1B).

We built on the original model code using the Keras [58] interface of Tensorflow 2.11 [59]. We used the Adam optimizer with mean squared error loss. Upon recommendation of the original study authors, we set the dropout rate = 0.3 and weight decay = 1e-3. We selected the initial learning rate from {1e-3, 1e-4, 1e-5} using the validation sets of each fold. Supplementary Table S2 shows the optimal initial learning rate for each study and fold. We used a cosine learning rate decay over 25 epochs and trained the models for 35 epochs total. The final weights were taken from the epoch with the lowest validation MAE. Models were trained on a NVIDIA RTX 3090 GPU with 24GB RAM on top of cuda 11.0 with a batch size of 4.

5.3 Associations with cognition

To examine cross-sectional and longitudinal associations with cognition in both elderly and children, we conducted several analyses, which are shown schematically in Figure 1C. Supplementary Table S3 shows the model equations. Statistical results were corrected for multiple comparisons across cognitive domains using the Holm-Bonferroni method [60]. Change in adjusted R^2 (ΔR^2_{adj}) was calculated from the difference between a model including the variable of interest and covariates and a model only including covariates. Variance inflation factors were confirmed to be less than five to rule out multicollinearity among baseline BAG, change in BAG (when included), and other covariates (especially chronological age). Analyses were performed in R 4.2.1 [61] with RStudio [62]. Packages dplyr [63] and tidyr [64] were used for data organization and packages car [65] and scales [66] were used for visualization.

Elderly For each cognitive domain in EDIS, we related baseline BAG to baseline cognitive score, with chronological age, sex, and years of education as covariates. For the longitudinal analyses in SLABS, we first calculated annual rates of change in BAG and cognition using linear regressions with time for each participant. For each cognitive domain in SLABS, we related baseline BAG to long-term rate of cognitive change (calculated from all five phases). Next, again for each cognitive domain, we related early rate of BAG change (calculated from the first three phases) to long-term rate of cognitive change (calculated from all five phases). If this relation was significant for a domain, we lastly related early rate of BAG change (calculated from the first three phases) to future rate of cognitive change (calculated from the last measurement of BAG onwards). All models included chronological age, sex, and years of education as covariates. Models with rate of BAG change also included baseline BAG as a covariate.

Children For cross-sectional analyses, we related baseline BAG to baseline cognition, both at 4.5 years old. For longitudinal analyses, we again calculated annual rates of change in BAG using linear regressions with time for each participant. We then related baseline BAG to future cognition and early rate of change in BAG to future cognition. Here, early rate of BAG change was calculated from 4.5 to 7.5 years, while future cognition was measured at 8.5 years. Chronological age and sex were included as covariates in all models. Models with rate of BAG change also included baseline BAG as a covariate.

5.4 Model interpretability

For each dataset, we investigated model interpretability using all scans from the same participants as the associations with cognition. Guided backpropagation [35] was used to compute individual saliency maps for both the pretrained and finetuned models. Guided backpropagation was previously shown to give similar results as occlusion for a brain age model, at a higher resolution and lower computational cost [11]. For finetuned models, the fold where the participant was included in the test set was used. Maps were registered to a common space using Advanced Normalization Tools (ANTs) [67]. Specifically, for each participant, input (minimally preprocessed) images were nonlinearly registered to MNI 152 space using the default parameters. This transformation was then applied to each participants' saliency maps.

To identify brain features that contributed the most to brain age predictions, we first averaged saliency maps over all participants in a study. We then retained the top 10% of voxels and calculated gray and white matter network/regional contributions. We used the 400-area Schaefer parcellation [68] assigned to 7 functional networks [69] for cortical gray matter. We averaged over all voxels in all parcels for each network. For subcortical gray matter, we used the automated anatomical labelling atlas 3 (AAL3) [70] to identify regions containing the hippocampus and amygdala, the putamen and caudate, and the thalamus. For white matter, we used the ICBM-DTI-81 atlas [71] with 48 ROIs. For all regions, we averaged over all voxels in both hemispheres. Contributions were normalized to sum to 1, giving the relative contribution. To visualize saliency maps in 2D, we set the maximum value to the 99th percentile and overlaid select slices (x = 97, z = 68, z = 89, z = 135) over a MNI 152 template brain. Full 3D saliency maps will be made available online.

Data and code availability

Custom Python and R code for this study will be made publicly available at `https://github.com/susan-cheng/brain-age-longitudinal`. Data that support the findings of this study are available from co-authors C.L.H.C (EDIS), M.W.L.C. (SLABS), and M.J.M. (GUSTO) upon collaborative request.

DISCLOSURE STATEMENT

The authors have nothing to disclose.

ACKNOWLEDGEMENTS

We would like to thank all participants and the research teams of GUSTO, EDIS, and SLABS for their contributions. We would also like to thank Esten Leonardsen and team for making the SFCN pretrained model and code available and for providing suggestions on hyperparameters. Finally, we would like to thank Zijiao Chen, Yilei Wu, Yichi Zhang, Yao Feng Chong, and Joanna Su Xian Chong for helpful discussions.

The Epidemiology of Dementia in Singapore (EDIS) study is supported by the National Medical Research Council (NMRC), Singapore (NMRC/CG/NUHS/2010 [Grant no: R-184-006-184-511]) and Bright Focus Foundation [R-608-000-248-597]. The research conducted in this study is also supported by the Singapore National Research Foundation under the Translational and Clinical Research (TCR) Flagship (GUSTO), Healthy Longevity Catalyst Awards (Zhou), Open-fund Young Individual Research Grant (Ng), Open Fund Large Collaborative Grant (OFLCG) Programmes (Zhou) by Singapore Ministry of Health's National Medical Research Council (NMRC) (Singapore – NMRC/TCR/004-NUS/2008; NMRC/TCR/012-NUHS/2014; HLCA23Feb0004; OFLCG/MOH-000504; MOH-OFYIRG19may-0012), National Medical Research Council Singapore Grants NMRC/STaR/0004/2008, NMRC/STaR/015/2013, and STAR19may-0001 (Chee), NMRC/CIRG/1446/2016 (Chen), NMRC/CIRG/1390/2014 and NMRC/CBRG/0088/2015 (Zhou), and from the Biomedical Research Council, Singapore (BMRC 04/1/36/372, Zhou). Additional funding is provided by the Duke-NUS Medical School Signature Research Program funded by Ministry of Health, Singapore, and Centre for Sleep and Cognition funded by Yong Loo Lin School of Medicine, National University of Singapore and the Brain – Body Initiative of A*STAR.

References

[1] R. a. I. Bethlehem, J. Seidlitz, S. R. White, *et al.*, "Brain charts for the human lifespan," en, *Nature*, pp. 1–11, Apr. 2022, Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: `10.1038/s41586-022-04554-y`.

[2] T. Hedden and J. D. E. Gabrieli, "Insights into the ageing mind: A view from cognitive neuroscience," en, *Nature Reviews Neuroscience*, vol. 5, no. 2, pp. 87–96, Feb. 2004, Number: 2 Publisher: Nature Publishing Group, ISSN: 1471-0048. DOI: `10.1038/nrn1323`.

[3] D. R. Roalf, R. E. Gur, K. Ruparel, *et al.*, "Within-Individual Variability in Neurocognitive Performance: Age and Sex-Related Differences in Children and Youths From Ages 8 to 21," *Neuropsychology*, vol. 28, no. 4, pp. 506–518, Jul. 2014, ISSN: 0894-4105. DOI: `10.1037/neu0000067`.

[4] W. M. v. d. Flier and P. Scheltens, "Epidemiology and risk factors of dementia," en, *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. suppl 5, pp. v2–v7, Dec. 2005, Publisher: BMJ Publishing Group Ltd, ISSN: 0022-3050, 1468-330X. DOI: `10.1136/jnnp.2005.082867`.

[5] R. Marsh, A. J. Gerber, and B. S. Peterson, "Neuroimaging Studies of Normal Brain Development and Their Relevance for Understanding Childhood Neuropsychiatric Disorders," *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 47, no. 11, pp. 1233–1251, Nov. 2008, ISSN: 0890-8567. DOI: `10.1097/CHI.0b013e318185e703`.

[6] K. Franke, E. Luders, A. May, M. Wilke, and C. Gaser, "Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI," en, *NeuroImage*, vol. 63, no. 3, pp. 1305–1312, Nov. 2012, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2012.08.001.

[7] C. Gaser, K. Franke, S. Klöppel, N. Koutsouleris, H. Sauer, and A. D. N. Initiative, "BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease," en, *PLOS ONE*, vol. 8, no. 6, e67346, Jun. 2013, Publisher: Public Library of Science, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0067346.

[8] J. H. Cole, S. J. Ritchie, M. E. Bastin, *et al.*, "Brain age predicts mortality," *Molecular Psychiatry*, vol. 23, no. 5, pp. 1385–1392, May 2018, ISSN: 1359-4184. DOI: 10.1038/mp.2017.62.

[9] T. Kaufmann, D. van der Meer, N. T. Doan, *et al.*, "Common brain disorders are associated with heritable patterns of apparent aging of the brain," en, *Nature Neuroscience*, vol. 22, no. 10, pp. 1617–1623, Oct. 2019, Number: 10 Publisher: Nature Publishing Group, ISSN: 1546-1726. DOI: 10.1038/s41593-019-0471-7.

[10] Y. E. Tian, V. Cropley, A. B. Maier, N. T. Lautenschlager, M. Breakspear, and A. Zalesky, "Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality," en, *Nature Medicine*, pp. 1–11, Apr. 2023, Publisher: Nature Publishing Group, ISSN: 1546-170X. DOI: 10.1038/s41591-023-02296-6.

[11] D. A. Wood, S. Kafiabadi, A. A. Busaidi, *et al.*, "Accurate brain-age models for routine clinical MRI examinations," en, *NeuroImage*, vol. 249, p. 118 871, Apr. 2022, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2022.118871.

[12] S. M. Hofmann, F. Beyer, S. Lapuschkin, *et al.*, "Towards the interpretability of deep learning models for multi-modal neuroimaging: Finding structural changes of the ageing brain," en, *NeuroImage*, vol. 261, p. 119 504, Nov. 2022, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2022.119504.

[13] C. Yin, P. Imms, M. Cheng, *et al.*, "Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment," *Proceedings of the National Academy of Sciences*, vol. 120, no. 2, e2214634120, Mar. 2023, Publisher: Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.2214634120.

[14] E. H. Leonardsen, H. Peng, T. Kaufmann, *et al.*, "Deep neural networks learn general and clinically relevant representations of the ageing brain," en, *NeuroImage*, vol. 256, p. 119 210, Aug. 2022, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2022.119210.

[15] V. M. Bashyam, G. Erus, J. Doshi, *et al.*, "MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide," en, *Brain*, vol. 143, no. 7, pp. 2312–2324, Jul. 2020, ISSN: 0006-8950, 1460-2156. DOI: 10.1093/brain/awaa160.

[16] C.-L. Chen, Y.-C. Hsu, L.-Y. Yang, *et al.*, "Generalization of diffusion magnetic resonance imaging–based brain age prediction model through transfer learning," en, *NeuroImage*, vol. 217, p. 116 831, Aug. 2020, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2020.116831.

[17] T. Hahn, J. Ernsting, N. R. Winter, *et al.*, "An uncertainty-aware, shareable, and transparent neural network architecture for brain-age modeling," *Science Advances*, vol. 8, no. 1, eabg9471, Jan. 2022, Publisher: American Association for the Advancement of Science. DOI: 10.1126/sciadv.abg9471.

[18] J. Li, D. Bzdok, J. Chen, *et al.*, "Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity," en, *Science Advances*, vol. 8, no. 11, eabj1812, Mar. 2022, ISSN: 2375-2548. DOI: 10.1126/sciadv.abj1812.

[19] S. Qiu, Y. Lu, Y. Li, *et al.*, "Prevalence of autism spectrum disorder in Asia: A systematic review and meta-analysis," *Psychiatry Research*, vol. 284, p. 112 679, Feb. 2020, ISSN: 0165-1781. DOI: 10.1016/j.psychres.2019.112679.

[20] J.-A. S. Catindig, N. Venketasubramanian, M. K. Ikram, and C. Chen, "Epidemiology of dementia in Asia: Insights on prevalence, trends and novel risk factors," English, *Journal of the Neurological Sciences*, vol. 321, no. 1, pp. 11–16, Oct. 2012, Publisher: Elsevier, ISSN: 0022-510X, 1878-5883. DOI: 10.1016/j.jns.2012.07.023.

[21] R. J. Jirsaraie, A. J. Gorelik, M. M. Gatavins, *et al.*, "A systematic review of multimodal brain age studies: Uncovering a divergence between model accuracy and utility," en, *Patterns*, vol. 4, no. 4, p. 100 712, Apr. 2023, ISSN: 26663899. DOI: 10.1016/j.patter.2023.100712.

[22] J. Wrigglesworth, N. Yaacob, P. Ward, *et al.*, "Brain-predicted age difference is associated with cognitive processing in later-life," en, *Neurobiology of Aging*, vol. 109, pp. 195–203, Jan. 2022, ISSN: 0197-4580. DOI: 10.1016/j.neurobiolaging.2021.10.007.

[23] J. Wrigglesworth, P. Ward, I. H. Harding, *et al.*, "Factors associated with brain ageing - a systematic review," *BMC Neurology*, vol. 21, no. 1, p. 312, Aug. 2021, ISSN: 1471-2377. DOI: 10.1186/s12883-021-02331-4.

[24] L. Baecker, R. Garcia-Dias, S. Vieira, C. Scarpazza, and A. Mechelli, "Machine learning for brain age prediction: Introduction to methods and clinical applications," English, *eBioMedicine*, vol. 72, Oct. 2021, Publisher: Elsevier, ISSN: 2352-3964. DOI: 10.1016/j.ebiom.2021.103600.

[25] D. Vidal-Pineiro, Y. Wang, S. K. Krogsrud, *et al.*, "Individual variations in 'brain age' relate to early-life factors more than to longitudinal brain change," *eLife*, vol. 10, J. Zhou, C. Büchel, and X.-N. Zuo, Eds., e69995, Nov. 2021, Publisher: eLife Sciences Publications, Ltd, ISSN: 2050-084X. DOI: 10.7554/eLife.69995.

[26] M. L. Elliott, D. W. Belsky, A. R. Knodt, *et al.*, "Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort," en, *Molecular Psychiatry*, vol. 26, no. 8, pp. 3829–3838, Aug. 2021, Number: 8 Publisher: Nature Publishing Group, ISSN: 1476-5578. DOI: 10.1038/s41380-019-0626-7.

[27] K. Franke, C. Gaser, T. J. Roseboom, M. Schwab, and S. R. de Rooij, "Premature brain aging in humans exposed to maternal nutrient restriction during early gestation," *NeuroImage*, vol. 173, pp. 460–471, Jun. 2018, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2017.10.047.

[28] K. B. Walhovd, M. Lövden, and A. M. Fjell, "Timing of lifespan influences on brain and cognition," English, *Trends in Cognitive Sciences*, vol. 27, no. 10, pp. 901–915, Oct. 2023, Publisher: Elsevier, ISSN: 1364-6613, 1879-307X. DOI: 10.1016/j.tics.2023.07.001.

[29] R. P. Dörfel, J. M. Arenas-Gomez, P. M. Fisher, *et al.*, "Prediction of brain age using structural magnetic resonance imaging: A comparison of accuracy and test-retest reliability of publicly available software packages," en, Neuroscience, preprint, Jan. 2023. DOI: 10.1101/2023.01.26.525514.

[30] S. Hilal, M. K. Ikram, M. Saini, *et al.*, "Prevalence of cognitive impairment in Chinese: Epidemiology of Dementia in Singapore study," en, *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 84, no. 6, pp. 686–692, Jun. 2013, Publisher: BMJ Publishing Group Ltd Section: Cognition, ISSN: 0022-3050, 1468-330X. DOI: 10.1136/jnnp-2012-304080.

[31] S. Hilal, C. S. Tan, X. Xin, *et al.*, "Prevalence of Cognitive Impairment and Dementia in Malays – Epidemiology of Dementia in Singapore Study," en, *Current Alzheimer Research*, vol. 14, no. 6, pp. 620–627, May 2017, ISSN: 15672050. DOI: 10.2174/1567205012666151002123813.

[32] M. Y. Z. Wong, C. S. Tan, N. Venketasubramanian, *et al.*, "Prevalence and Risk Factors for Cognitive Impairment and Dementia in Indians: A Multiethnic Perspective from a Singaporean Study," en, *Journal of Alzheimer's Disease*, vol. 71, no. 1, pp. 341–351, Jan. 2019, Publisher: IOS Press, ISSN: 1387-2877. DOI: 10.3233/JAD-190610.

[33] M. W. L. Chee, K. H. M. Chen, H. Zheng, *et al.*, "Cognitive function and brain structure correlations in healthy elderly East Asians," en, *NeuroImage*, vol. 46, no. 1, pp. 257–269, May 2009, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2009.01.036.

[34] S.-E. Soh, M. T. Tint, P. D. Gluckman, *et al.*, "Cohort Profile: Growing Up in Singapore Towards healthy Outcomes (GUSTO) birth cohort study," *International Journal of Epidemiology*, vol. 43, no. 5, pp. 1401–1409, Oct. 2014, ISSN: 0300-5771. DOI: 10.1093/ije/dyt125.

[35] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, *Striving for Simplicity: The All Convolutional Net*, arXiv:1412.6806 [cs], Apr. 2015.

[36] R. L. F. Leong, J. C. Lo, S. K. Y. Sim, *et al.*, "Longitudinal brain structure and cognitive changes over 8 years in an East Asian cohort," en, *NeuroImage*, vol. 147, pp. 852–860, Feb. 2017, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2016.10.016.

[37] A.-M. G. de Lange, M. Anatürk, J. Rokicki, *et al.*, "Mind the gap: Performance metric evaluation in brain-age prediction," en, *Human Brain Mapping*, vol. 43, no. 10, pp. 3113–3129, Jul. 2022, ISSN: 1097-0193. DOI: 10.1002/hbm.25837.

[38] M. W. L. Chee, H. Zheng, J. O. S. Goh, D. Park, and B. P. Sutton, "Brain structure in young and old east Asians and Westerners: Comparisons of structural volume and cortical thickness," *Journal of Cognitive Neuroscience*, vol. 23, no. 5, pp. 1065–1079, May 2011, ISSN: 0898-929X. DOI: 10.1162/jocn.2010.21513.

[39] T. T. Brown, J. M. Kuperman, Y. Chung, *et al.*, "Neuroanatomical Assessment of Biological Maturity," en, *Current Biology*, vol. 22, no. 18, pp. 1693–1698, Sep. 2012, ISSN: 0960-9822. DOI: 10.1016/j.cub.2012.07.002.

[40] J. D. Lewis, A. C. Evans, and J. Tohka, "T1 white/gray contrast as a predictor of chronological age, and an index of cognitive performance," en, *NeuroImage*, vol. 173, pp. 341–350, Jun. 2018, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2018.02.050.

[41] R. J. Jirsaraie, T. Kaufmann, V. Bashyam, *et al.*, "Benchmarking the generalizability of brain age models: Challenges posed by scanner variance and prediction bias," en, *Human Brain Mapping*, vol. 44, no. 3, pp. 1118–1128, 2023, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.26144, ISSN: 1097-0193. DOI: 10.1002/hbm.26144.

[42] B. S. Khundrakpam, J. Tohka, and A. C. Evans, "Prediction of brain maturity based on cortical thickness at different spatial resolutions," en, *NeuroImage*, vol. 111, pp. 350–359, May 2015, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2015.02.046.

[43] G. Erus, H. Battapady, T. D. Satterthwaite, *et al.*, "Imaging Patterns of Brain Development and their Relationship to Cognition," *Cerebral Cortex*, vol. 25, no. 6, pp. 1676–1684, Jun. 2015, ISSN: 1047-3211. DOI: 10.1093/cercor/bht425.

[44] B. Tervo-Clemmens, F. J. Calabro, A. C. Parr, J. Fedor, W. Foran, and B. Luna, "A canonical trajectory of executive function maturation from adolescence to adulthood," en, *Nature Communications*, vol. 14, no. 1, p. 6922, Oct. 2023, Number: 1 Publisher: Nature Publishing Group, ISSN: 2041-1723. DOI: 10.1038/s41467-023-42540-8.

[45] K. B. Walhovd, S. K. Krogsrud, I. K. Amlien, *et al.*, "Back to the future: Omnipresence of fetal influence on the human brain through the lifespan," *eLife*, May 2023, Publisher: eLife Sciences Publications, Ltd. DOI: 10.7554/elife.86812.1.

[46] D. J. Simmonds, M. N. Hallquist, M. Asato, and B. Luna, "Developmental stages and sex differences of white matter and behavioral development through adolescence: A longitudinal diffusion tensor imaging (DTI) study," *NeuroImage*, vol. 92, pp. 356–368, May 2014, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.12.044.

[47] A. M. Fjell and K. B. Walhovd, "Structural brain changes in aging: Courses, causes and cognitive consequences," eng, *Reviews in the Neurosciences*, vol. 21, no. 3, pp. 187–221, 2010, ISSN: 0334-1763.

[48] M. Korbmacher, A. M. de Lange, D. van der Meer, *et al.*, "Brain-wide associations between white matter and age highlight the role of fornix microstructure in brain ageing," en, *Human Brain Mapping*, vol. 44, no. 10, pp. 4101–4119, 2023, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.26333, ISSN: 1097-0193. DOI: 10.1002/hbm.26333.

[49] B. Mwangi, K. M. Hasan, and J. C. Soares, "Prediction of individual subject's age across the human lifespan using diffusion tensor imaging: A machine learning approach," en, *NeuroImage*, vol. 75, pp. 58–67, Jul. 2013, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.02.055.

[50] G. Rodríguez-Nieto, C. Seer, J. Sidlauskaite, *et al.*, "Inhibition, Shifting and Updating: Inter and intra-domain commonalities and differences from an executive functions activation likelihood estimation meta-analysis," *NeuroImage*, vol. 264, p. 119 665, Dec. 2022, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2022.119665.

[51] N. Gogtay, J. N. Giedd, L. Lusk, *et al.*, "Dynamic mapping of human cortical development during childhood through early adulthood," *Proceedings of the National Academy of Sciences*, vol. 101, no. 21, pp. 8174–8179, May 2004, Publisher: Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.0402680101.

[52] J. Bagautdinova, J. Bourque, V. J. Sydnor, *et al.*, *Development of White Matter Fiber Covariance Networks Supports Executive Function in Youth*, en, Pages: 2023.02.09.527696 Section: New Results, Feb. 2023. DOI: 10.1101/2023.02.09.527696.

[53] A.-L. Goddings, D. Roalf, C. Lebel, and C. K. Tamnes, "Development of white matter microstructure and executive functions during childhood and adolescence: A review of diffusion MRI studies," *Developmental Cognitive Neuroscience*, vol. 51, p. 101 008, Aug. 2021, ISSN: 1878-9293. DOI: 10.1016/j.dcn.2021.101008.

[54] B. Cao, B. Mwangi, K. M. Hasan, *et al.*, "Development and validation of a brain maturation index using longitudinal neuroanatomical scans," en, *NeuroImage*, vol. 117, pp. 311–318, Aug. 2015, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2015.05.071.

[55] F. Ségonne, A. M. Dale, E. Busa, *et al.*, "A hybrid approach to the skull stripping problem in MRI," en, *NeuroImage*, vol. 22, no. 3, pp. 1060–1075, Jul. 2004, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2004.03.032.

[56] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," en, *NeuroImage*, 20 YEARS OF fMRI, vol. 62, no. 2, pp. 782–790, Aug. 2012, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2011.09.015.

[57] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," eng, *Medical image analysis*, vol. 5, no. 2, pp. 143–156, Jun. 2001, ISSN: 1361-8423. DOI: 10.1016/s1361-8415(01)00036-6.

[58] F. Chollet *et al.*, *Keras*, https://keras.io, 2015.

[59] M. Abadi, A. Agarwal, P. Barham, *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, 2015.

[60] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," en, *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[61] R Core Team, *R: A language and environment for statistical computing*, Vienna, Austria, 2022.

[62] RStudio Team, *RStudio: Integrated Development Environment for R*, Boston, MA, 2022.

[63] H. Wickham, R. François, L. Henry, K. Müller, and RStudio, *Dplyr: A Grammar of Data Manipulation*, Sep. 2022.

[64] H. Wickham, M. Girlich, and RStudio, *Tidyr: Tidy Messy Data*, Sep. 2022.

[65] J. Fox, S. Weisberg, B. Price, *et al.*, *Car: Companion to Applied Regression*, Oct. 2022.

[66] H. Wickham, D. Seidel, and RStudio, *Scales: Scale Functions for Visualization*, Aug. 2022.

[67] B. B. Avants, N. Tustison, and H. Johnson, "Advanced Normalization Tools (ANTS)," en,

[68] A. Schaefer, R. Kong, E. M. Gordon, *et al.*, "Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, Sep. 2018, ISSN: 1047-3211. DOI: 10.1093/cercor/bhx179.

[69] B. T. Thomas Yeo, F. M. Krienen, J. Sepulcre, *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, Sep. 2011, ISSN: 0022-3077. DOI: 10.1152/jn.00338.2011.

[70] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, "Automated anatomical labelling atlas 3," en, *NeuroImage*, vol. 206, p. 116 189, Feb. 2020, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2019.116189.

[71] S. Mori, K. Oishi, H. Jiang, *et al.*, "Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template," *NeuroImage*, vol. 40, no. 2, pp. 570–582, Apr. 2008, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2007.12.035.