Brain-Age Prediction: Systematic Evaluation of Site Effects, and Sample Age Range and Size

Yuetong (Jordan) Yu[1], Hao-Qi Cui[1], Shalaila Haas[2], Faye New[2], Nicole Sanford[1], Kevin Yu[1], Denghuang Zhan[1], Guoyuan Yang[3], Jia-hong Gao[4], Dongtao Wei[5], Jiang Qiu[5], Boris Bernhardt[6], Paul Thompson[7], Sophia Frangou[1,2], Ruiyang Ge[1], ENIGMA World Aging Center

[1]University of British Columbia, Vancouver, Canada [2]Icahn School of Medicine at Mount Sinai, New York, USA [3]Beijing Institute of Technology, Beijing, China [4]Peking University, Beijing, China [5]Southwest University, Chongqing, China [6]Montreal Neurological Institute and Hospital, Montreal, Canada [7]Keck School of Medicine, University of Southern California, Marina del Rey, CA, USA

Background

Brain age prediction using machine learning and neuroimaging techniques has recently gained research momentum, and brain age has emerged as a promising biomarker of cognitive and clinical outcomes in both healthy and clinical populations. The ongoing research interest in brain age has highlighted the need for robust and publicly available brain age models pre-trained on data from large samples of healthy individuals. In this work, we develop, empirically validate, and disseminate a pre-trained brain-age model to cover most of the human lifespan [1].

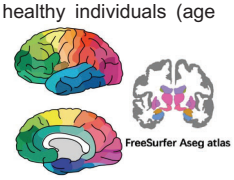
Objectives

- Assess the effects of age range on brain age models across the human lifespan
- Assess the effects of site harmonization methods
- Test longitudinal consistency of pre-trained brain age models
- Establish a minimum sample size requirement for a stable brain age model

Methods

We developed sex-specific brain age models across the human lifespan by using brain regional morphometric data [2]. We selected the best-performing model after systematically examining the impact of site harmonization, age range, and sample size on brain-age prediction in a discovery sample of brain morphometric measures from 35,683 healthy individuals (age range: 5-90 years; 53.59% female). The pre-trained models were tested for cross-dataset generalizability in an independent sample comprising 2,101 healthy individuals (age range: 8-80 years; 55.35% female) and for longitudinal consistency in a further sample comprising 377 healthy individuals (age range: 9-25 years; 49.87% female).

The regional morphometric features include cortical thickness, cortical surface area, and subcortical volume; and the model development procedures were based on our prior evidence [3]. Each model was trained by using Support Vector Regression with Radial Basis Function Kernel. Hyperparameter tuning was performed by using a grid search in a 10-fold cross-validation scheme across five repeats. The mean absolute error (MAE), which is the average of the absolute differences between the predicted and the chronological age, served as the main measure of model performance, supplemented by the correlation coefficient (CORR) between the predicted and the chronological age.



Desikan-Killiany atlas

Regional morphometric features include cortical thickness of 68 brain parcels, surface area of 68 brain parcels, and 14 subcortical volumes.

Results

Elements common to all results. (a) For simplicity, all figures display the results averaged across sexes as the sex-specific results were identical for males and females. (b) In this presentation, results of MAE were presented, and the same pattern was observed for all results of CORR [1].

- For both sexes, omitting any site correction consistently demonstrated superior performance in terms of attaining the lowest MAE values and highest CORR values across the age bins (Figure 1).

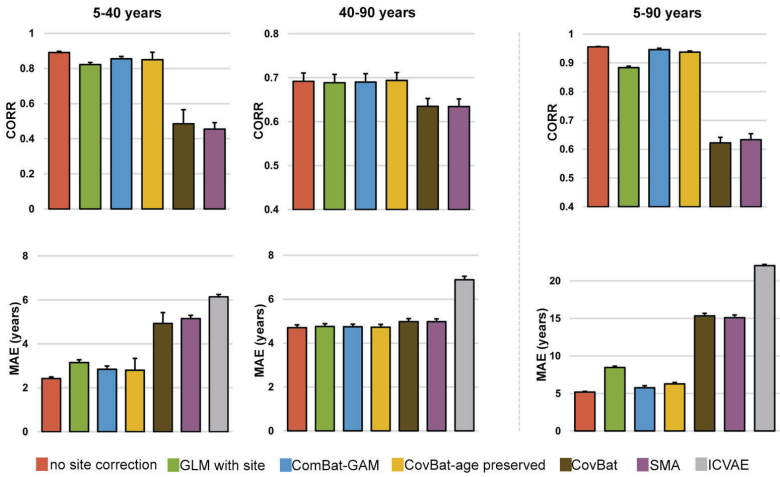


Figure 1. Performance metrics derived from repeated cross-validation in different age bins of the discovery sample

GLM: Generalized Linear Model; **ComBat**: combating batch effects when combining Batches method; **CovBat**: Correcting Covariance Batch Effects; **SMA**: Subsampling Maximum-mean-distance based distribution shift correction Algorithm; **ICVAE**: Invariant Conditional Variational AutoEncoder

- Models based on a wider age range accounted for more of the variance in age but were less accurate (Figure 2); we selected the two-bin partition with sequential 40-year intervals (i.e., 5-40 years and 40-90 years).

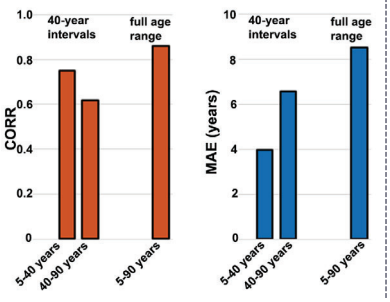


Figure 2. Performance metrics derived from the application of the models pre-trained on different age bins of the discovery sample to the corresponding age bins of the replication sample

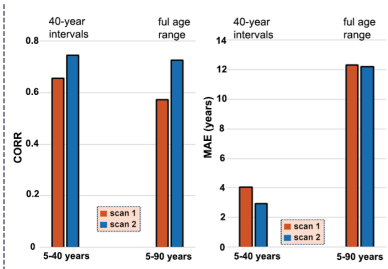


Figure 3. Model performance as a function of the sample size in the two age bins (5-40 years and 40-90 years) of the discovery sample. Model performance in longitudinal data. The results were generated by employing models that had been trained on discovery samples from each age range division and applied to the longitudinal consistency sample.

- The results (Figure 3) indicated that models utilizing the two-bin partition (i.e., 5-40 years and 40-90 years) achieved optimal consistency on the longitudinal data.
- The CORR (correlation coefficient between the predicted and the chronological age) and MAE (mean absolute error) exhibited smaller variation across varying sample sizes and stabilized around 1,600 participants (Figure 4).

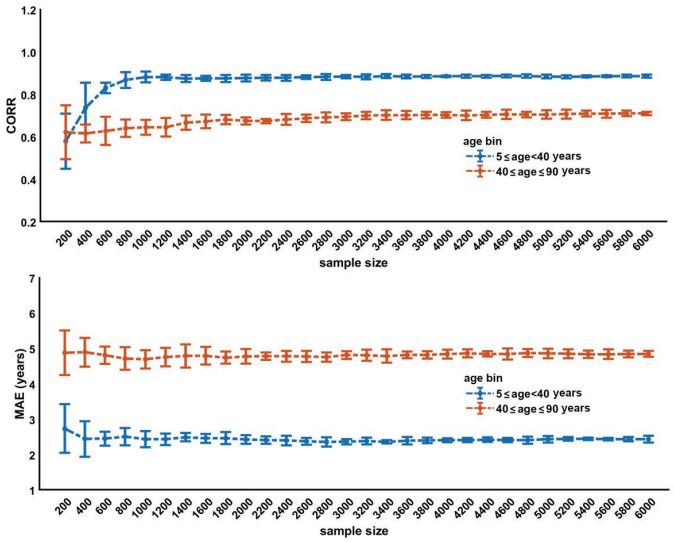


Figure 4. Model performance as a function of the sample size in the two age bins (5-40 years and 40-90 years) of the discovery sample.

Conclusions

- The accuracy of age prediction from morphometry data was higher when no site harmonization was applied.
- Dividing the discovery sample into two age-bins (5-40 years and 40-90 years) provided a better balance between model accuracy and explained age variance than other alternatives.
- Model accuracy for brain-age prediction plateaued at a sample size exceeding 1,600 participants.
- These findings have been incorporated into CentileBrain (https://centilebrain.org/#/brainAGE2), an open-science, web-based platform for individualized neuroimaging metrics.

References

[1] Yuetong Yu, Haoqi Cui, Shalaila Haas, Faye New, Nicole Sanford, Kevin Yu, Denghuang Zhan, Guoyuan Yang, Jia-Hong Gao, Dongtao Wei, Jiang Qiu, Boris Bernhardt, Paul Thompson, Sophia Frangou, Ruiyang Ge, ENIGMA World Aging Center. Brain-Age Prediction: Systematic Evaluation of Site Effects, and Sample Age Range and Size. bioRxiv, 2024. doi: https://doi.org/10.1101/2023.11.06.565917

[2] Ruiyang Ge, Yuetong Yu, et al., Paul M. Thompson, Sophia Frangou. Normative modelling of brain morphometry across the lifespan with CentileBrain: algorithm benchmarking and model optimisation. The Lancet Digital Health, 2024:6(3), e211-e221.

[3] Amirhossein Modabbernia, Heather C. Whalley, David C. Glahn, Paul M. Thompson, Rene S. Kahn, Sophia Frangou. Systematic evaluation of machine learning algorithms for neuroanatomically-based age prediction in youth. Human Brain Mapping, 2022:43(17), 5126-5140.