Enhancing Brain Age Prediction

Backes Lucas[1,2], Eickhoff Simon[1,2], Rubbert Christian[3], Phillips Christophe[4], Antonopoulos Georgios[1,2] & Patil Kaustubh[1,2]

[1]Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany;
[2]Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Centre Jülich, Jülich, Germany
[3]Department of Diagnostic and Interventional Radiology, Medical Faculty and University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, Germany
[4]GIGA CRC Human Imaging, University of Liège, Belgium
l.backes@fz-juelich.de

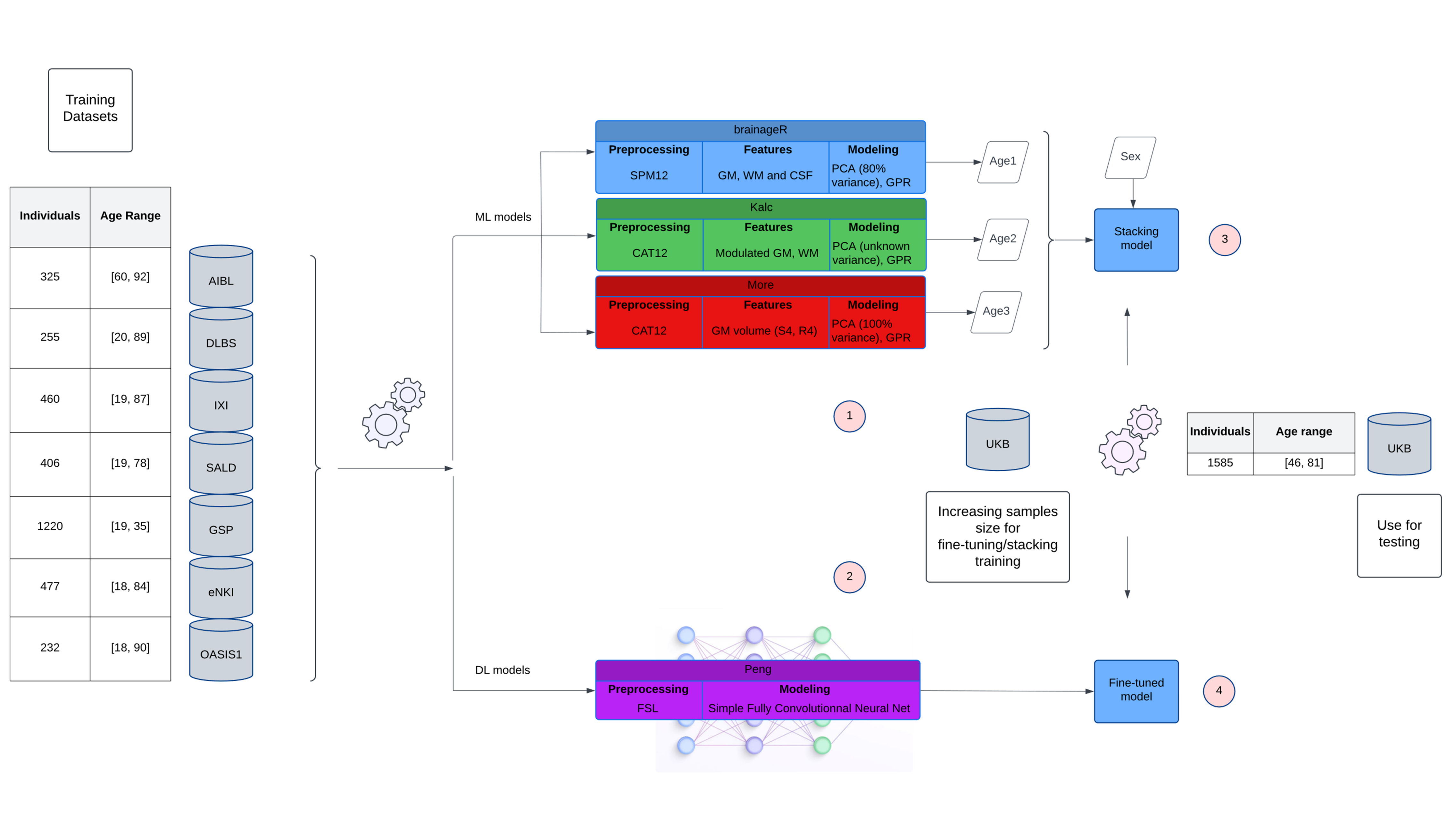
Introduction

- MRI can help us produce a biomarker for an overall ‘brain age’ such that it more accurately measures disease and mortality risks than chronological age.
- However, due to site and scanner differences brain age models do not usually work well on new datasets that were not used for training. This drawback prevents clinical use of ‘brain age’ as an informative and relevant biomarker.

Methods

- Train different algorithms the on same seven datasets
- UK Biobank as a new test site
- Evaluate stacking of ML models and fine-tuning of a DL model using increasing size of UKB training samples

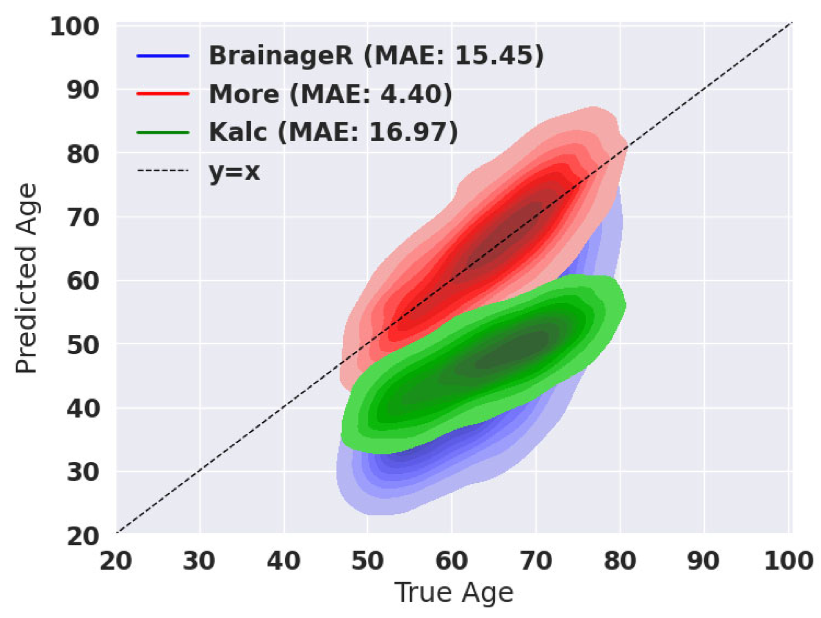
Schematic



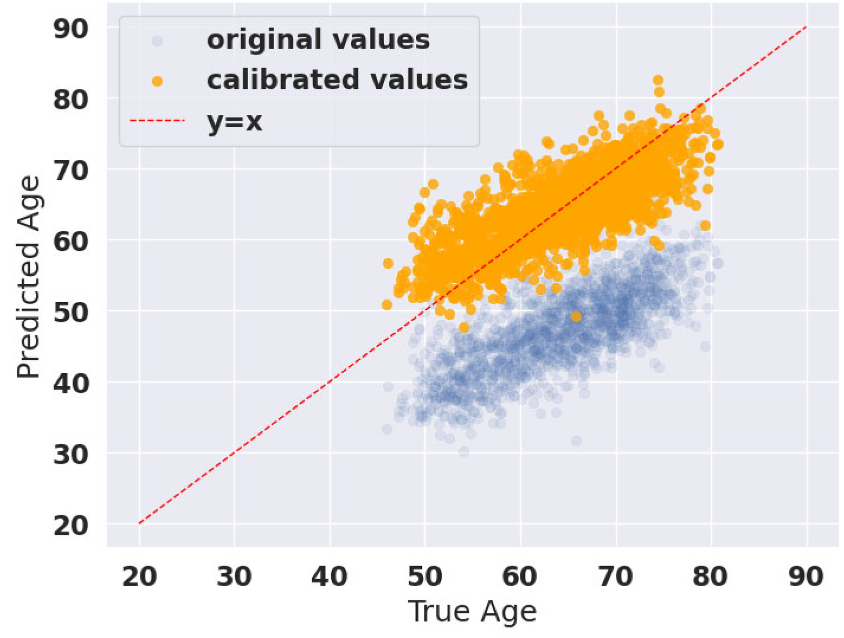
Results

Out-of-box

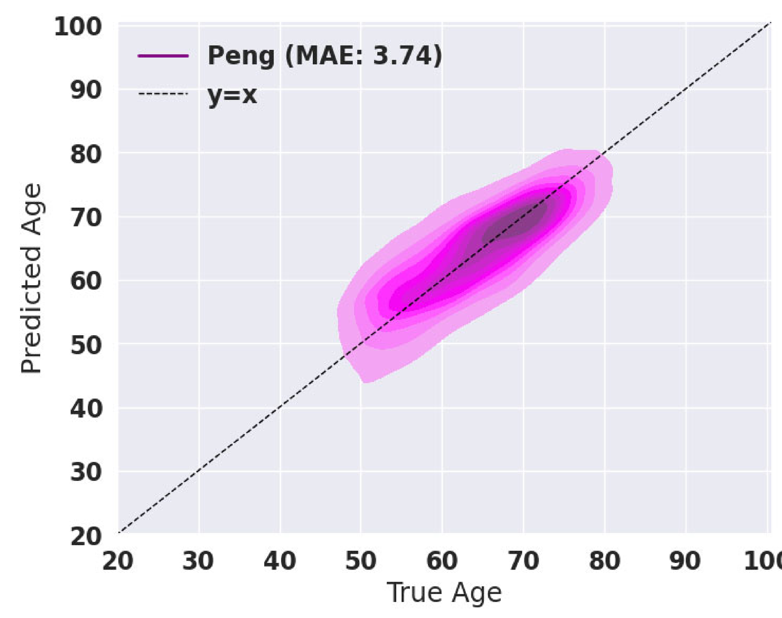
1



Contour plot of true age vs predicted age for the three ML models



True age vs predicted age of the Kalc model (blue) and calibrated (orange) using linear regression



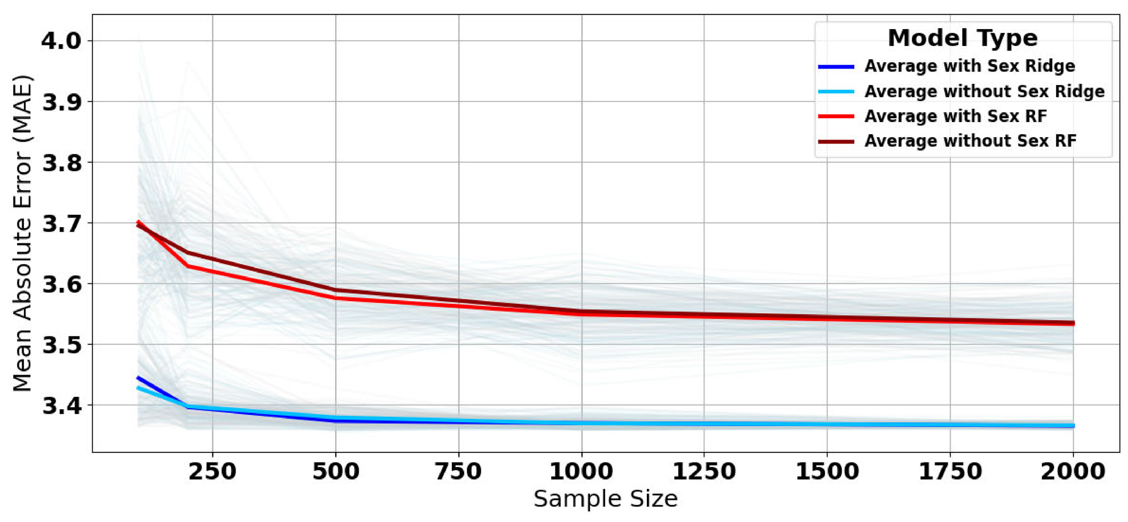
Contour plot of true age vs predicted age for the DL model from Peng

2

Enhanced

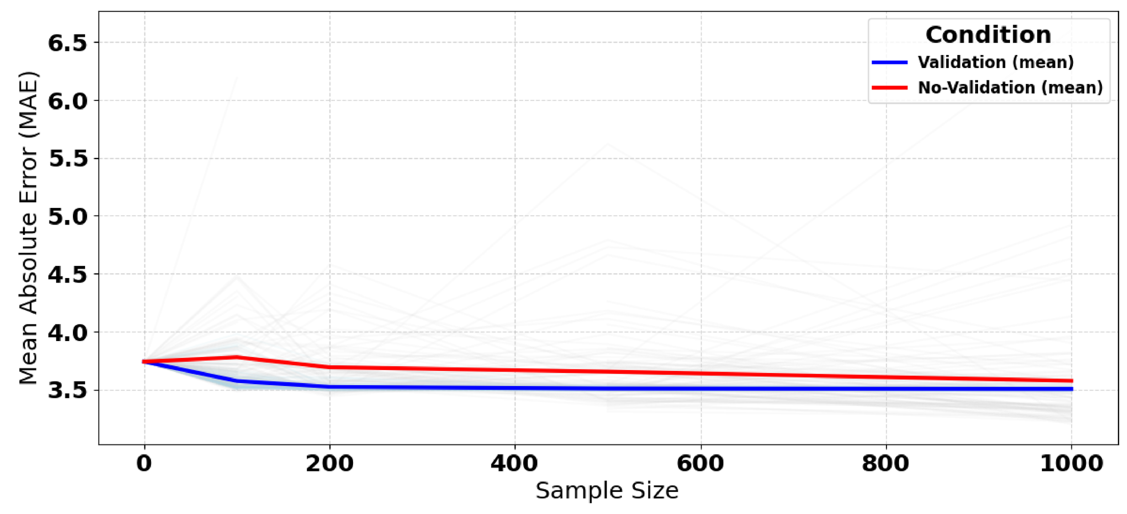
- Same UKB samples were used for model enhancement (stacking and fine tuning)
- Samples stratified by age (using age bins) and sex to preserve the actual data distribition.
- All samples were from healthy, white individuals.
- For stacking, Ridge Regression and Random Forest were compared, with and without **sex** as a feature.
- For fine-tuning, we compared validation approach and full training approach.

3



Evolution of MAE with respect to the training sample size for Random Forest and Ridge Regression

4



Evolution of MAE with respect to the training sample size for DL fine tuning using Validation and no Validation

Discussion

1 and 2

- Among the three classical ML models, *More* yielded the best results (MAE = 4.40), outperforming *brainageR* (MAE = 15.45) and *Kalc* (MAE = 16.97).
- Calibrating each ML model using a linear regression led to improvement suggesting that a linear shift can improve the models (example displayed for calibration of the Kalc model).
- The Simple Fully Convolutional Network (SFCN) model *Peng* generalized best as an "out-of-the-box" model, achieving lowest MAE of 3.74.

3 and 4

- Ridge regression (MAE = 3.37) outperformed Random Forests as a stacking method (MAE = 3.53), especially with increasing training sample size.
- Including sex as a predictive feature in the stacking models did not lead to substantial improvement suggesting that either the input features already encode information related to sex, or that sex itself does not contribute meaningful information for predicting age in this specific task.
- Fine-tuning the Peng model yielded only marginal gains.

- When fine tuning, the validation method (training on 80% of the data and stopping when validation error on 20% data is at its lowest) seems to be on average better than the full training method (same validation as before but followed by retraining on 100% of the data stopping at same epochs).
- The latter approach seems to give rise to more jumps in error therefore compromising the average.
- Fine tuning plateaued rather early (N=200) compared to stacking, RF (N=1500) and Ridge (N=500), and performed in between the two methods (MAE=3.5).

References: Kalc, Polona et al. “BrainAGE: Revisited and reframed machine learning workflow.” Human brain mapping vol. 45,3 (2024): e26632. doi:10.1002/hbm.26632; More, Shammi et al. Brain-age prediction: A systematic comparison of machine learning workflows, NeuroImage, Volume 270, 2023, 119947, ISSN 1053-8119, doi.org/10.1016/j.neuroimage.2023.119947; Peng, Han et al., Accurate brain age prediction with lightweight deep neural networks, Medical Image Analysis, Volume 68, 2021, 101871, ISSN 1361-8415, doi.org/10.1016/j.media.2020.101871. Sudlow, Cathie et al. “UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.” PLoS medicine vol. 12,3 e1001779. 31 Mar. 2015, doi:10.1371/journal.pmed.1001779

Acknowledgments: This study was partly supported by Helmholtz-AI project BrainAge4AD and private funding from TVG