FASTGPR: DIVIDE-AND-CONQUER TECHNIQUE IN NEUROIMAGING DATA SHORTENS TRAINING TIME AND IMPROVES ACCURACY

F. Raimondo[1,2], G. Antonopoulos[1,2], S. B. Eickhoff [1,2], K. R. Patil[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
g.antonopoulos@fz-juelich.de

Introduction

- **Gaussian process regression (GPR)** has shown great potential for studying healthy aging and disease via brain-age prediction (BAP) using structural MRI[1].
- A big drawback of GPR is the **training complexity** which is an **O(N^3) operation** (N=number of data points).
- The need for expansive datasets and the **high dimensionality of MRI data**, renders the training of GPR impractical with conventional computing resources.
- We investigated whether a divide-and-conquer approach can be used together with the GPR model.

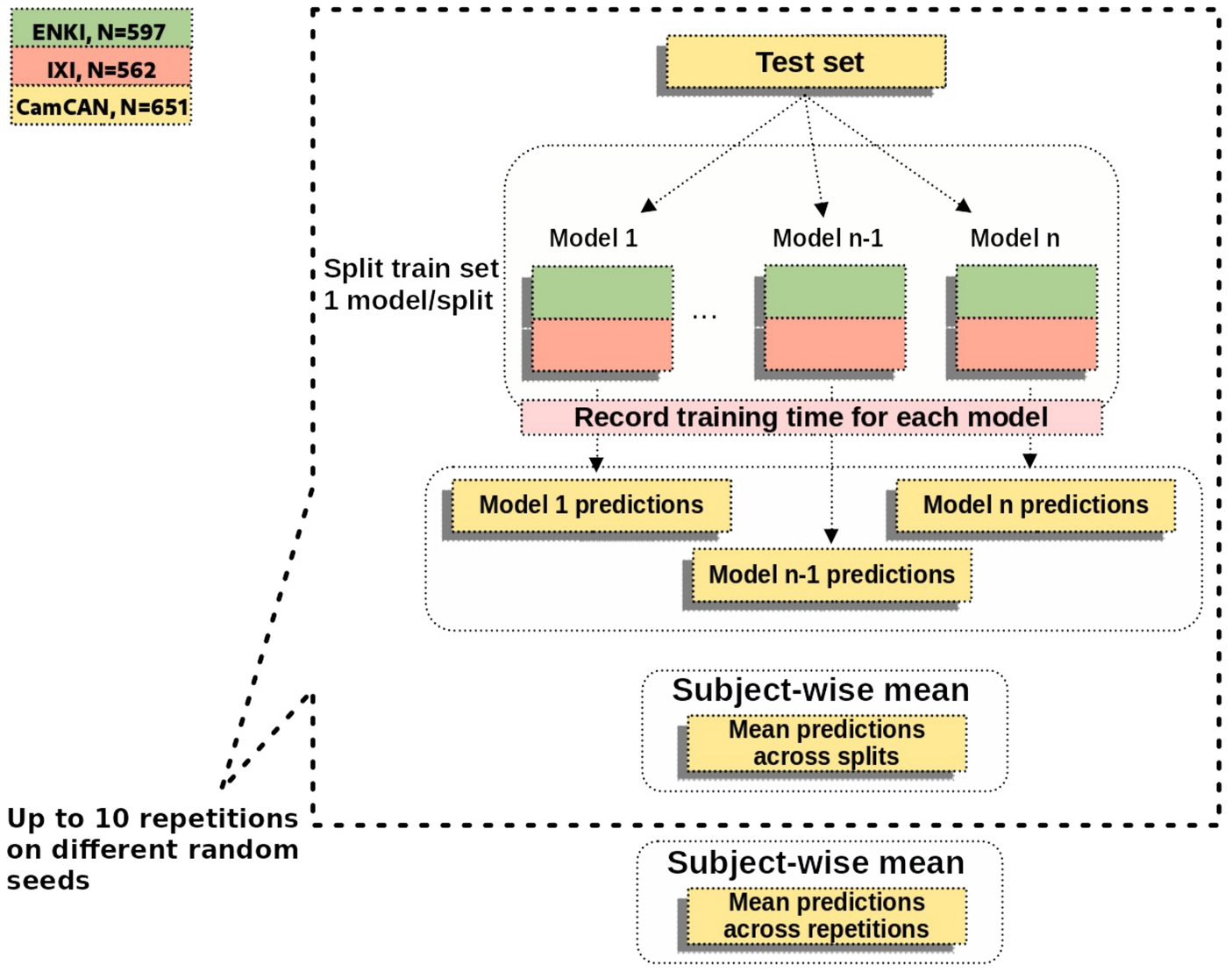
Material

- **T1w MRI** scans of healthy subjects from IXI [2], eNKI [3], CamCAN [4] (each n>500, **total N=1810**, 18-88 age range).
- All analyses were run on an Apple M2 pro (12-core) processor with 32GB RAM, under the same conditions.

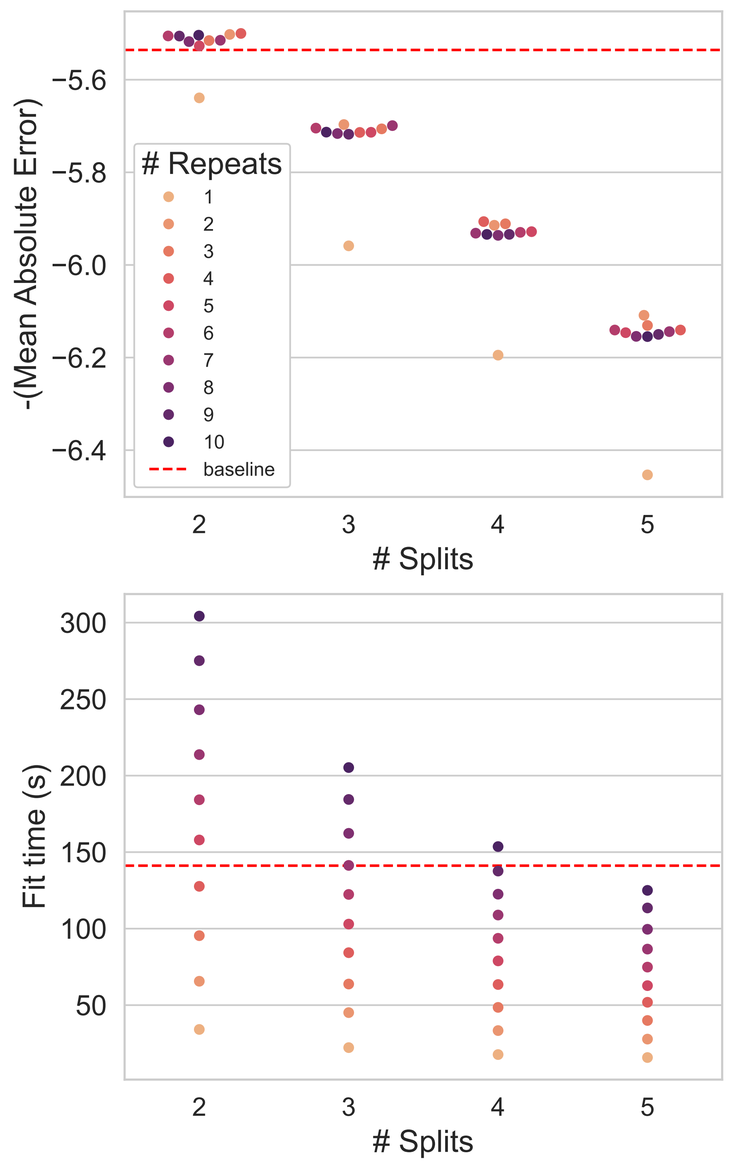
Methods

MRI preprocessing: CAT 12.8 [5] → linear and non-linear spatial normalization, tissue segmentation and modulation.
238,955 voxels per subject representing voxel-wise gray matter volume.
Smoothing → 4 mm FWHM Gaussian kernel,
Resampling (linear interpolation) → 8 mm spatial resolution Finally **n=3747** features per subject.

- **Performance estimated in terms of mean absolute error (MAE) using Leave-One-Site-Out.**
- **Randomly divided training data into non-overlapping subsets while stratifying the splits over age.**
- **One GPR model was trained on each subset**
- **Final prediction was obtained by averaging the predictions of all the models.**
- **We implemented this process with two-, three-, four- and a five-way split of the training data.**
- **Repeated up to ten times with distinct random seeds. The prediction for a test sample was obtained by averaging predictions across the splits.**



Results & Conclusions



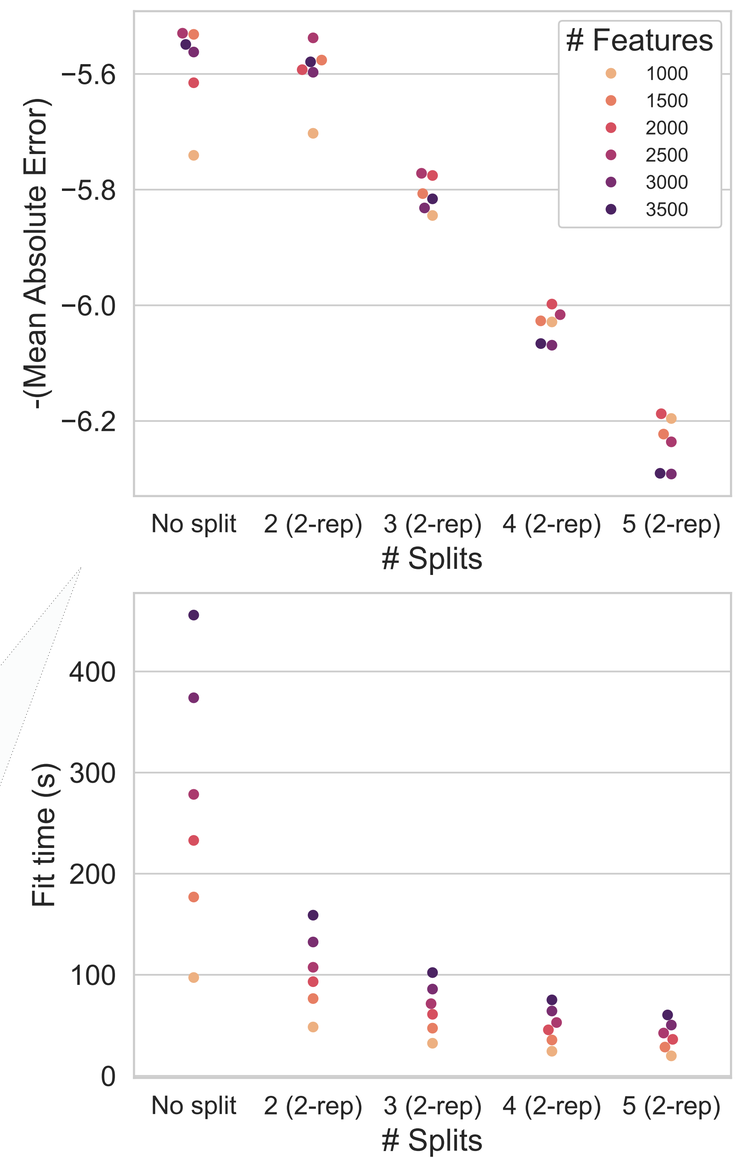
** training time is estimated as
** no-split model's performance (red line)

- 2-splits MAE=5.63 – total compute time=24.2% of the no-split model's time
- The 3-, 4- and 5-split configurations manifest a reduction in training time and a pronounced drop in predictive performance
- 2 repetitions (2 random seeds) of 2-splits MAE=5.5, training time 46.4%

The 2-split configuration with at least two iterations as the optimal trade-off between training time and prediction accuracy when using GPR for brain age prediction.

No-split model, with 1000 features: total training time represented 42.1% of the training time of 2000 features, & 26.4% of the training time of using 3000 features.
Similar time reductions were obtained within the split setups

The non-split, 2x2-splits and 2x3-splits models exhibited a less pronounced impact on MAE, except for 1000 features.
The 2x4- and 2x5-splits models displayed a decrease in predictive performance for high number of features (3000 and 3500).
Low sample/feature ratio might cause a drop in predictive performance.



References: [1] More et al, Neuroimage, 270, 119947 (2023), [2] Wolpert, D.H. Neural Netw. 5, 241–259 (1992), [3] (https://brain-development.org/ixi-dataset/), [4] K.B. Nooner et al. Front. Neurosci., 6 (2012), p. 152, [5] J.R. Taylor et al. Neuroimage, 144 (Pt B) (2017), pp. 262-269, [6] S. Caspers et al. Front. Aging Neurosci., 6 (2014), p. 149
[2] 'IXI CAT Preprocessed data', doi: 10.5281/zenodo.7716839.
[3] J. R. Taylor et al., 'The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample', NeuroImage, vol. 144, pp. 262–269, Jan. 2017, doi: 10.1016/j.neuroimage.2015.09.018.
[8] K. Nooner et al., 'The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry', Frontiers in Neuroscience, vol. 6, 2012, Accessed: Nov. 10, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2012.00152
[5] C. Gaser, R. Dahnke, K. Kurth, E. Luders, and Alzheimer's Disease Neuroimaging Initiative, 'A Computational Anatomy Toolbox for the Analysis of Structural MRI Data'.

Acknowledgments: This work was partly supported by the Helmholtz-AI project BrainAge4AD, The HelmholtzPortfolio Theme "Supercomputing and Modeling for the Human Brain", the Helmholtz Imaging Platform (project NimRLS), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project PA 3634/1-1, EI 816/21-1.