Stacking ensemble for age-prediction improves performance and privacy

G. Antonopoulos[1,2], S. More[1,2], F. Raimondo[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

Brain-age prediction (BAP) using structural MRI has shown great potential for studying healthy aging and disease. Two major desirable properties for BAP are high accuracy and data privacy. We propose a stacking ensemble model (SEM) [1] which takes the advantage of the most informative voxels and improves both properties compared to current standard implementations [2].

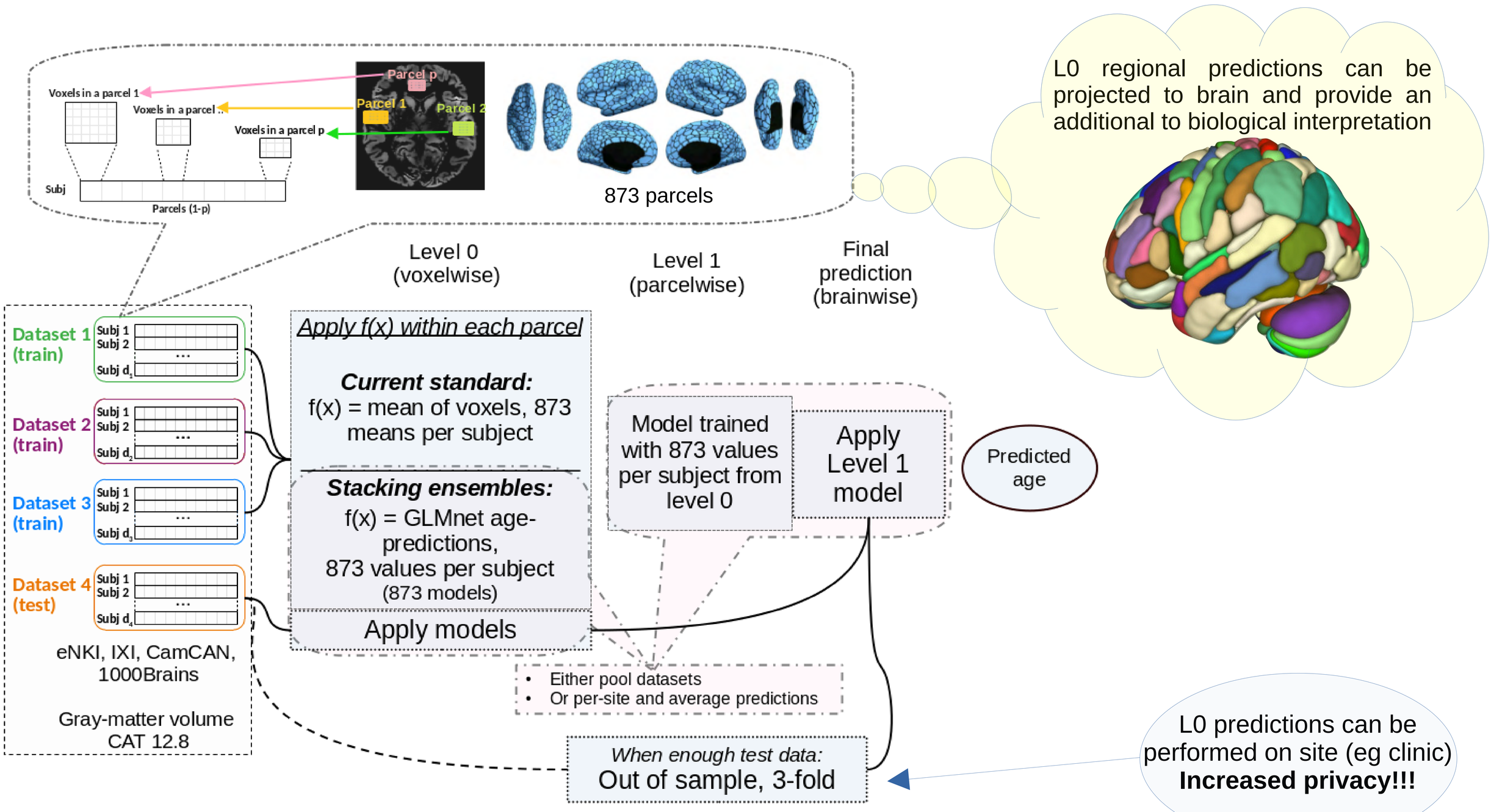
Material

- T1w MRI scans of healthy subjects from IXI [3], eNKI [4], CamCAN [5] and 1000Gehirne [6].
  (each n>500, total N=3103, 18-90 age range).
- Preprocessing: Voxel Based Morphometry CAT 12.8→ gray-matter volume (GMV).
- 873-parcel atlas to group GMV voxels

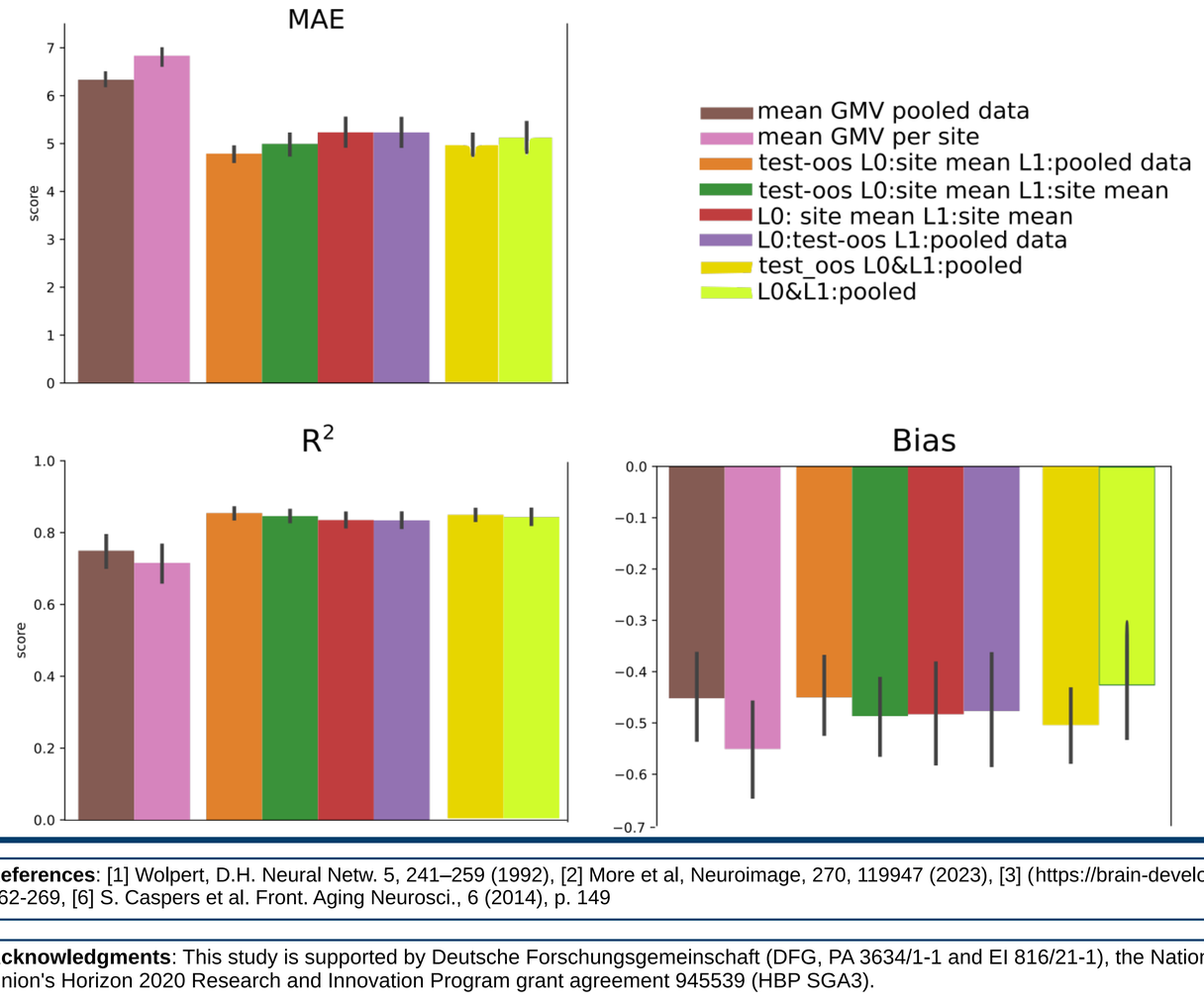
Methods

- Two-level (L0 and L1) SEM.
- Level 0 → trained a GLMnet model for each parcel. The out-of-sample (OOS, using 3-fold cross-validation). Final L0 models were trained on the whole train datasets
- Predictions from all L0 models were used as features to train a GLMnet model at L1. L1 provides the final age-prediction.

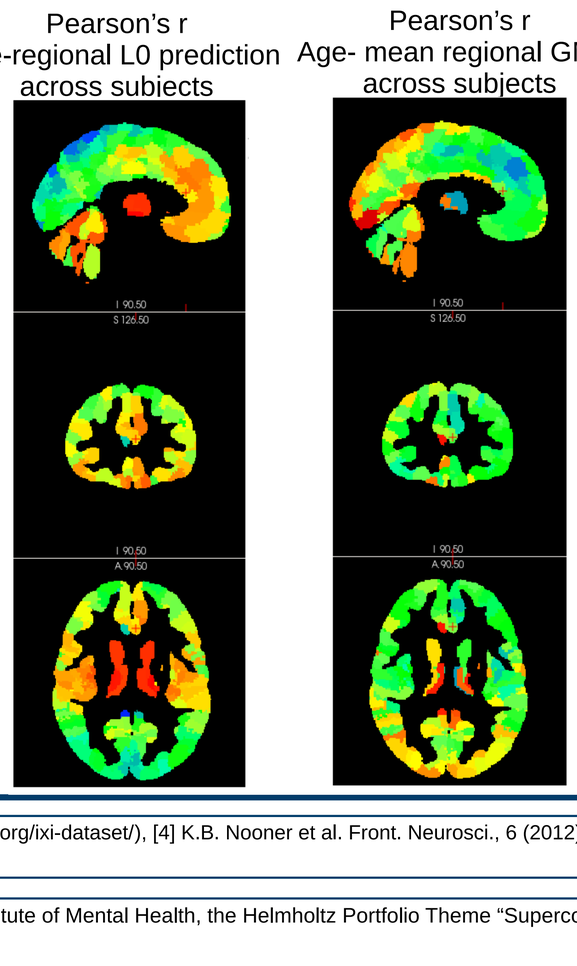
We explored two different ways to train models at L0 and L1 and all combinations between them, i.e.,
i) using pooled data from different sites, and
ii) treating each site separately and then averaging their outcomes.
Additionally, to test the case where enough data is available at the test site, we estimated L0-level OOS predictions on the test data.
These were then used to obtain predictions using L1 models.
The former schemes provide different levels and types of privacy advantage.
To compare with current standards we also tested models using average GMV in each parcel as inputs of L1.
Performance was estimated using leave-one-site-out analysis.



Results & Conclusions



Brain maps



- **Improved performance in all metrics!**
- **Improved data privacy**
- **An additional biological interpretation**

References: [1] Wolpert, D.H. Neural Netw. 5, 241–259 (1992), [2] More et al, Neuroimage, 270, 119947 (2023), [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

Acknowledgments: This study is supported by Deutsche Forschungsgemeinschaft (DFG, PA 3634/1-1 and EI 816/21-1), the National Institute of Mental Health, the Helmholtz Portfolio Theme "Supercomputing and Modelling for the Human Brain" and the European Union's Horizon 2020 Research and Innovation Program grant agreement 945539 (HBP SGA3).