Batch effects in sex classification analyses

Lisa Wiersch[1,2], Kaustubh R. Patil[1,2], Simon B. Eickhoff[1,2], Susanne Weis[1,2]

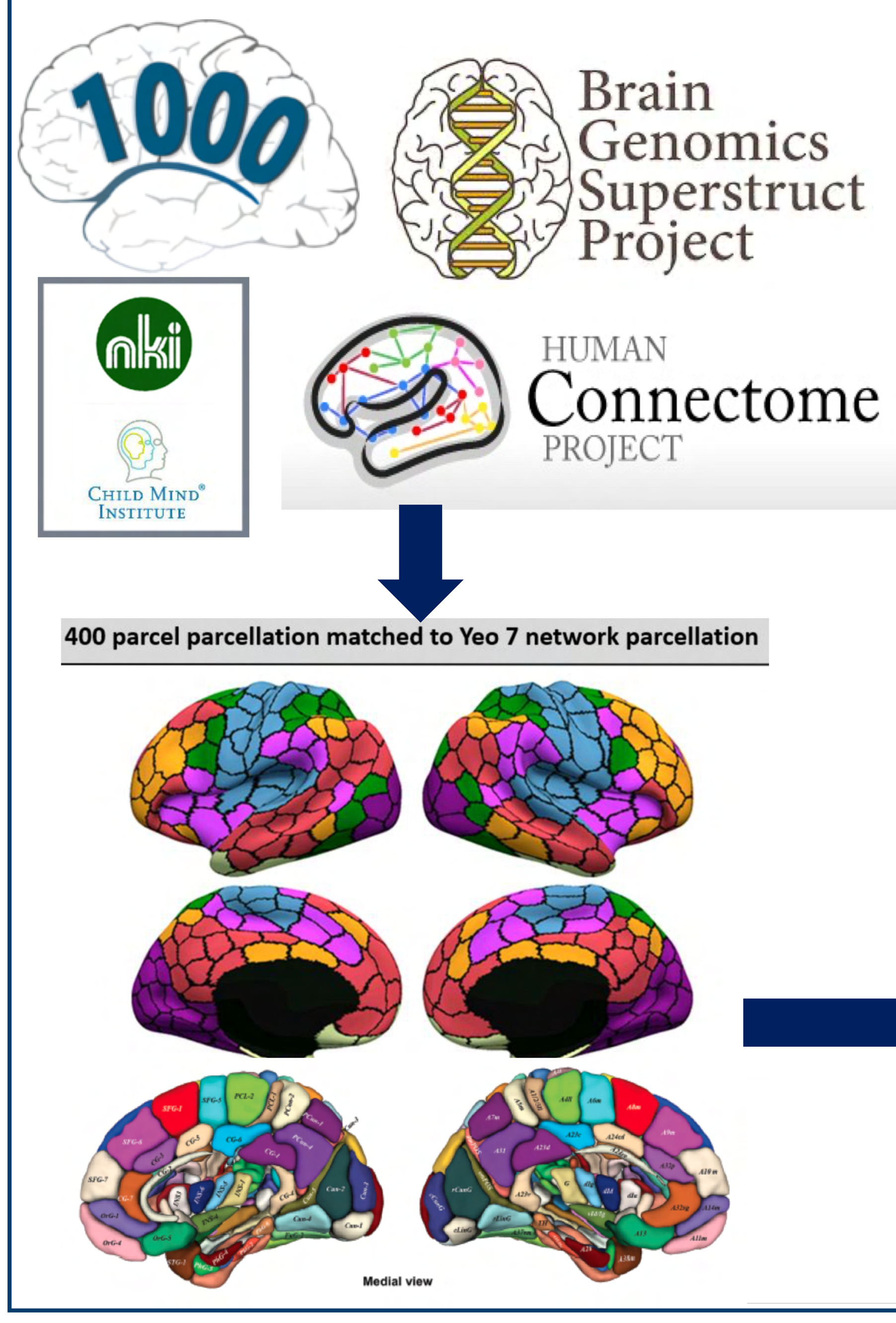
[1]Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Centre Jülich, Jülich, Germany
[2]Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany,
l.wiersch@fz-juelich.de

ID # 51

Introduction

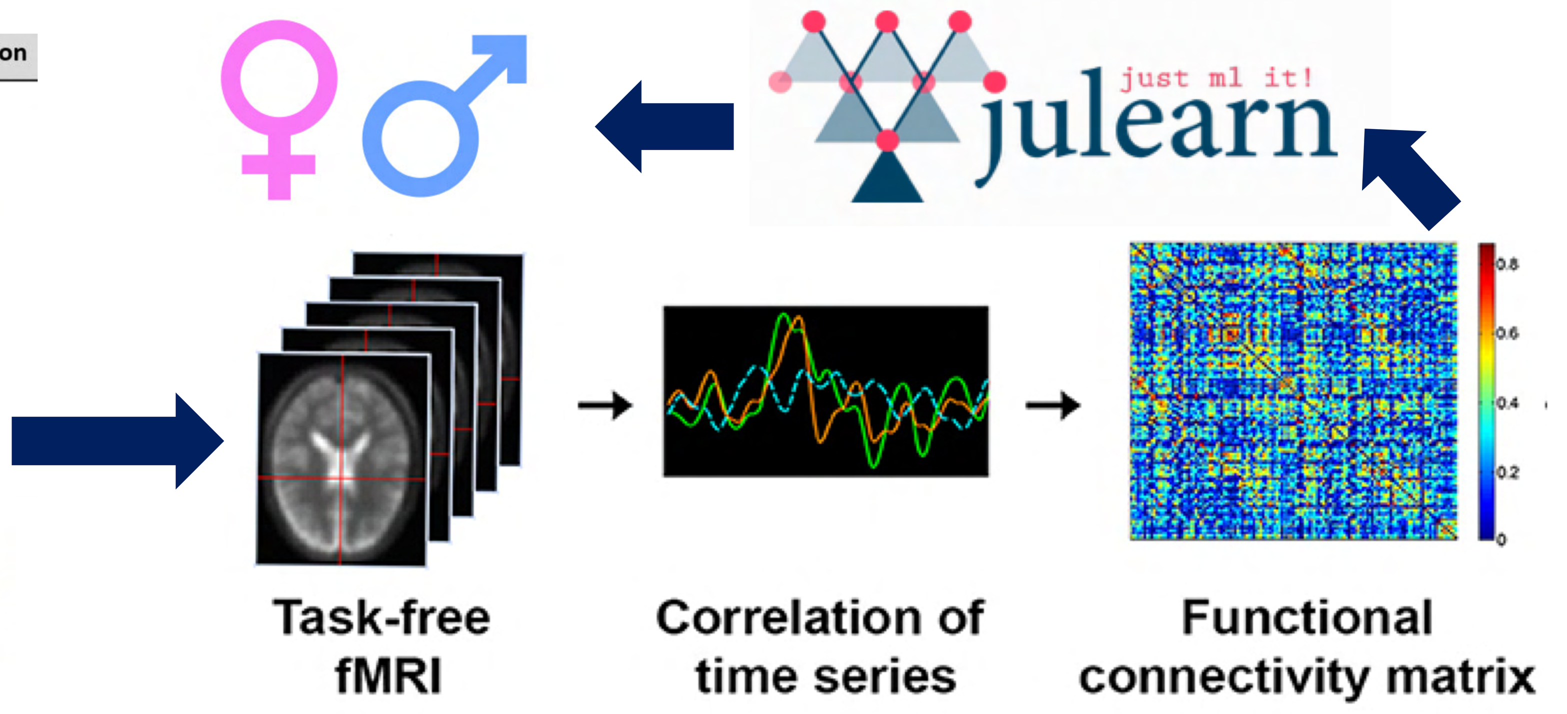
- Machine-learning analyses allow for the prediction of phenotypes from neuroimaging data (e.g. sex of a person)
- Which sample characteristics provide highest model performance for within- but also between-sample predictions?
- The present study adresses this question for sex classification analyses based on the resting-state functional connectivity
- 3 out of 4 datasets respectively were combined for training a sex classification model and the final model was applied to the remaining dataset
- Do the different models provide differences in the spatial pattern and total accuracies due to differences in the samples → Batch effects?
- Which brain regions classify consistenly on a high level?

Methods



- 1000Brains: N = 1000, age range: 21-85 (M = 61.18)
- eNKI: N = 190, age range: 20-83 (M = 46.02)
- GSP: N = 854, age range: 21-35 (M = 22.92)
- HCP: N = 878, age range 22-37 (M = 28.49)

- Parcelwise approach as in Weis et al. (2020, 5): 400 parcels from the Schaefer Atlas (6); 36 parcels from the Brainnetome Atlas (7)
- Support vector machine (SVM) classifier with 5 repetitions of a 10-fold cross-validation was built using Julearn



Results

1 Classification accuracies per model

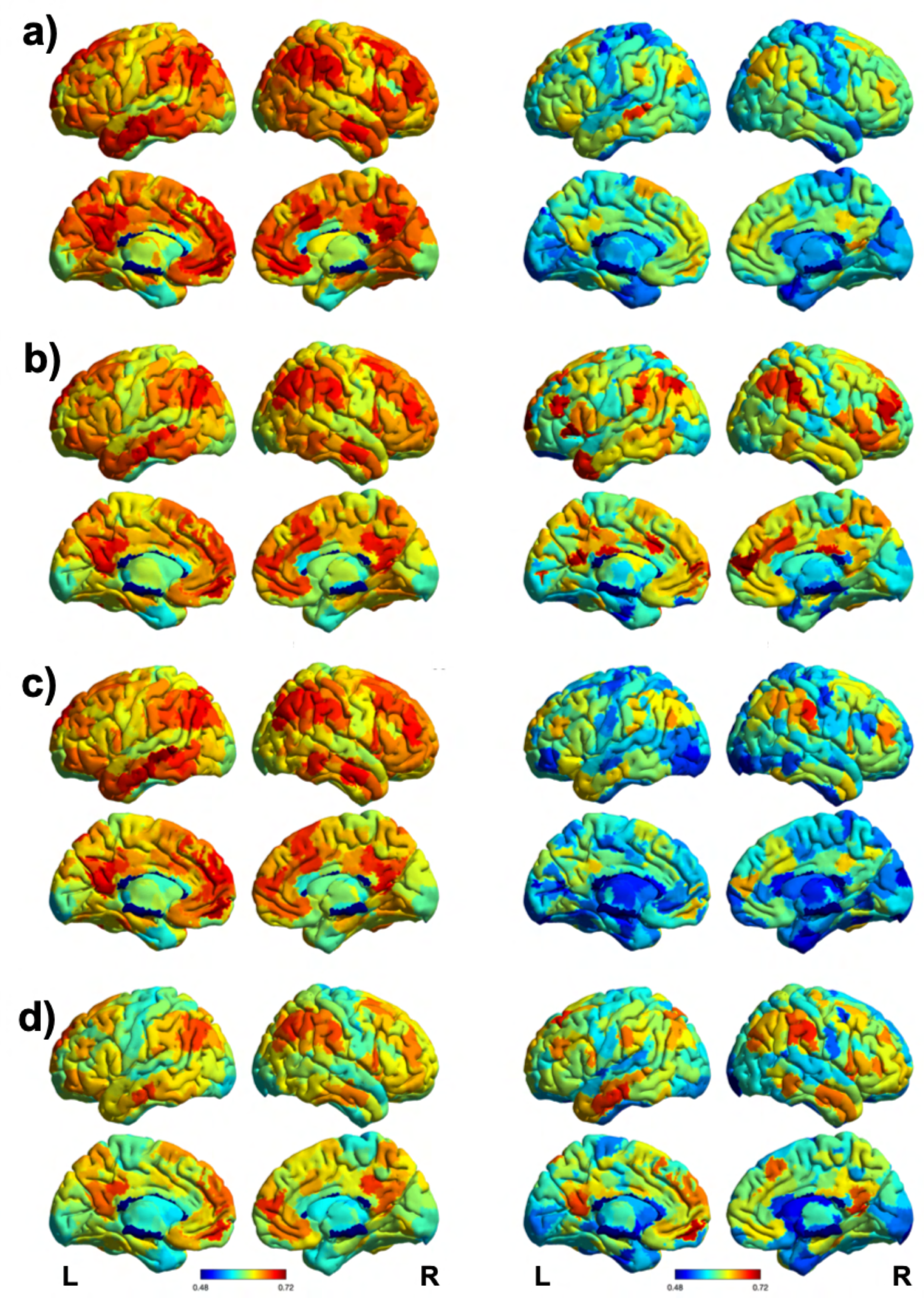


Figure 1. Spatial pattern of sex classification accuracies for within-sample predictions (first column) and between-sample predictions (second column) a) applied to 1000Brains; b) applied to eNKI; c) applied to GSP; d) applied to HCP

- Model trained on eNKI, GSP & HCP (N = 1922): M = 64.9% (53.0-71.3%), Model applied to 1000Brains: M = 58.0% (49.5-68.0%)
- Model trained on 1000Brains, GSP & HCP (N = 2732): M = 63.8% (53.0-70.5%), Model applied to eNKI: M = 61.4% (50.0-72.6%)
- Model trained on 1000Brains, eNKI, HCP (N = 2068): M = 63.9% (52.8-71.7%), Model applied to GSP: M = 57.3% (48.7-68.6%)
- Model trained on 1000Brains, eNKI, GSP (N = 2044): M = 61.5% (54.1-68.4%), Model applied to HCP: M = 59.4% (48.8-69.9%)
- Within- and between sample predictions displayed similar spatial patterns in classification accuracies as assessed by spearman rank correlations
- Highly classifying parcels are consistenly located in areas of the cingulate cortex, temporal gyrus, Precuneus, parietal lobule and inferior frontal gyrus

2 Consistent highly classifying parcels

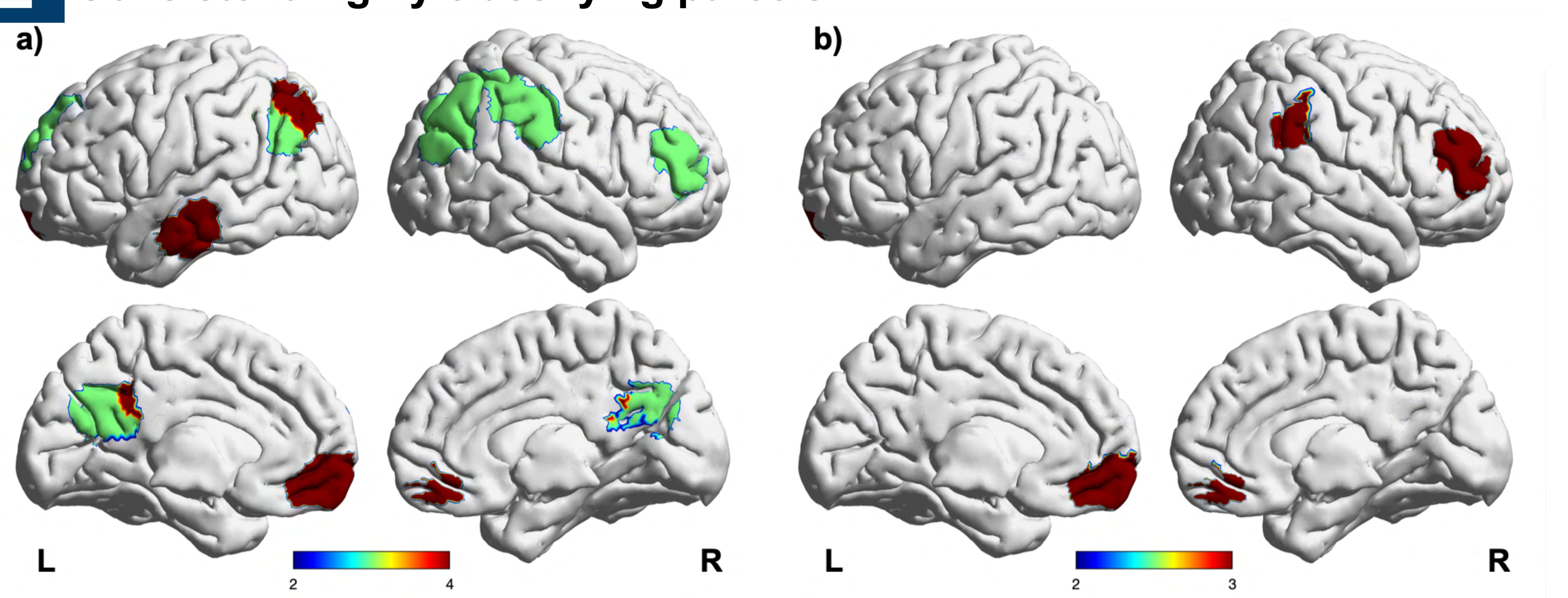


Figure 2. Frequencies of highly classying parcels across a) within- and b) between-sample predictions

Discussion

- Within-sample predictions were higher for those training samples that included the HCP dataset, possible due to the high imaging quality of this dataest
- Within-sample predictions were not driven by sample size, as the smallest dataset had the highest mean classification accuracy
- Model application worked best on eNKI in achieving highest mean and maximum classification accuracy whereas model application to GSP provided lowest between-sample prediction
- High accuracies of model application to eNKI were likely driven by sample size or the representation of age range of the application set
- Even if HCP has a comparative age range to GSP, the imaging quality of HCP in training may not be representative enough for a highly classifying model application to the GSP datasets, resulting in the lower model performance
- Even though we see differences in the models due to the differences in the samples, regions belonging to the default mode network are consistently the highest classifying parcels (5,8)

- Matching characteristics of training- and test-sample (e.g. age range) are essential for accurate model applications
- We observed batch effects in the models due to the differences in the samples, but the spatial patterns of highly classifying parcels were consistent

→ Males and females differ in the brain functional organization as captured by resting-state functional connectivity consistenly across differently trained models

References

1. Caspers, S., et al., *Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000BRAINS.* Front Aging Neurosci, 2014. **6**: p. 149.
2. Nooner, K.B., et al., *The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry.* Frontiers in neuroscience, 2012. **6**: p. 152.
3. Holmes, A.J., et al., *Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures.* Sci Data, 2015. **2**: p. 150031.
4. Van Essen, D.C., et al., *The Human Connectome Project: a data acquisition perspective.* Neuroimage, 2012. **62**(4): p. 2222-2231.
5. Weis, S., et al., *Sex Classification by Resting State Brain Connectivity.* Cereb Cortex, 2020. **30**(2): p. 824-835.
6. Schaefer, A., et al., *Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI.* Cereb Cortex, 2018. **28**(9): p. 3095-3114.
7. Fan, L., et al., *The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture.* Cereb Cortex, 2016. **26**(8): p. 3508-26.
8. Zhang, C., et al., *Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity.* Human brain mapping, 2018. **39**(4): p. 1765–1776.

https://juaml.github.io/julearn/main/index.html
https://www.nitrc.org/projects/bnv/

Acknowledgments: This study was supported by the Deutsche Forschungsgemeinschaft (DFG)