Effective workflow from multi-modal MRI data to model-based prediction

Kyesam Jung[1,2], Kevin J. Wischnewski[1,2,3], Simon B. Eickhoff[1,2], Oleksandr V. Popovych[1,2]

E-mail: k.jung@fz-juelich.de

[1]Institute of Neuroscience and Medicine, Brain and Behaviour (INM-7), Forschungszentrum Jülich, Germany
[2]Institute of Systems Neuroscience, Medical Faculty, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Germany
[3]Institute of Mathematics, Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany

JÜLICH Forschungszentrum

hhu Heinrich Heine Universität Düsseldorf

medRSD Medical Research School Düsseldorf

Introduction

- Comparing structural connectivity (SC) and functional connectivity (FC) led to the **structure-function relationship** as a possible methodological approach to explore the interdependence between structure and function of the human brain. However, this relationship between **empirical SC (eSC)** and **empirical FC (eFC)** is **relatively low**, might depend on many factors, and its mechanism is still unclear **[1]**.
- **Integration of model-based approaches into whole-brain connectome research** can expand the scope of investigation to understand the brain **[2,3]**. The models can be used to generate **simulated FC (sFC)** as **an additional data modality**. Accordingly, it can be suggested as a possible mediator between brain structure and function.
- We suggest a framework that advances the applicability of the model-based approach by applying simulated data to machine-learning analysis.

Methods: A workflow for model-based machine-learning research using multi-modal MRI data

Step 1 **Multi-modal MRI data**

HCP young adults (n=270, 142 females, 28.5± 3.5 years old)

- Structural and functional MRI data: T1-weighted MRI, Diffusion-weighted MRI, Resting-state functional MRI

Step 2 **MRI processing**

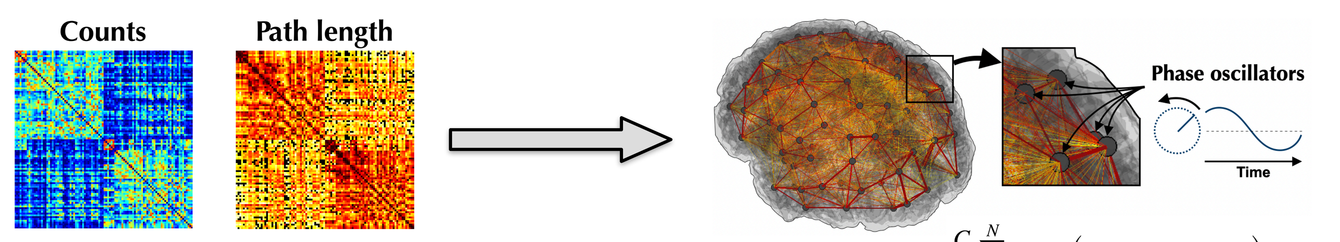
- Correction: bias field, head motion, eddy
- Denoising, tissue segmentation
- Image registration to MNI space

MRI pipeline

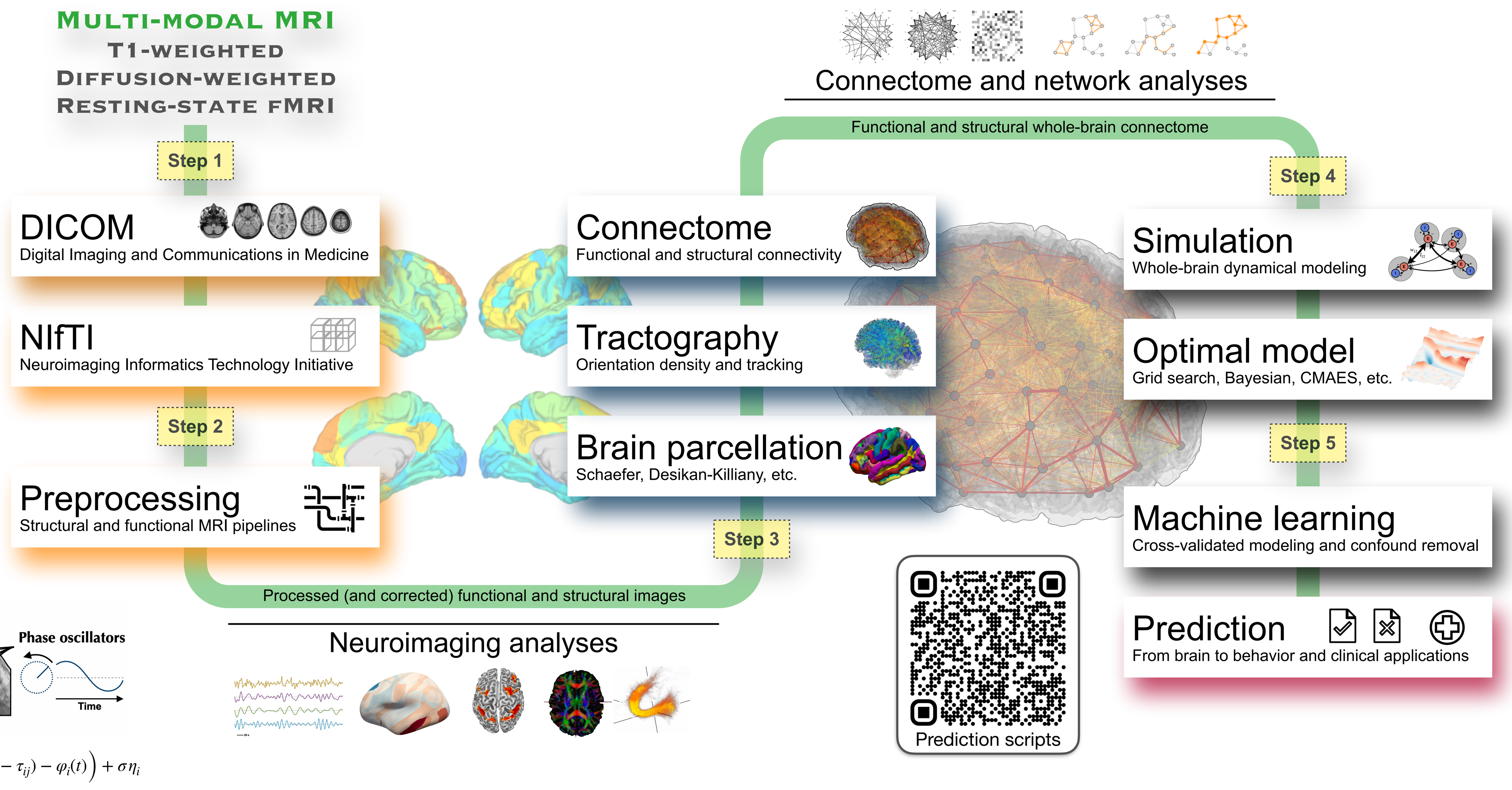
Step 3 **Whole-brain connectome**

- Functional atlas: Schaefer 100 cortical regions
- Structural atlas: Harvard-Oxford 96 cortical regions

✓Tracking streamlines ⇒ whole-brain tractography



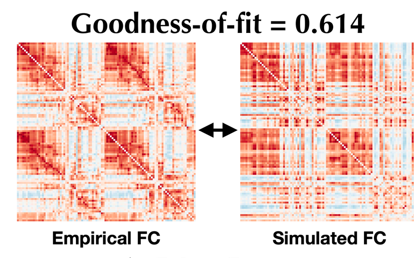
$$\dot{\varphi}_i(t) = 2\pi f_i + \frac{C}{N}\sum_{j=1}^{N} k_{ij}\sin\left(\varphi_j(t-\tau_{ij}) - \varphi_i(t)\right) + \sigma\eta_i$$



Step 4 **Whole-brain dynamical modeling**

- A whole-brain model (coupled phase oscillators with delay)
- Simulated Blood-Oxygen-Level Dependent signals ([0.01,0.1] Hz)

✓Correlation between eFC and sFC (goodness-of-fit)
✓Parameter optimization (Covariance Matrix Adaptation Evolution Strategy) **[4,5]**
Low dimensional (2 parameters), **High dimensional** (around 100 parameters)



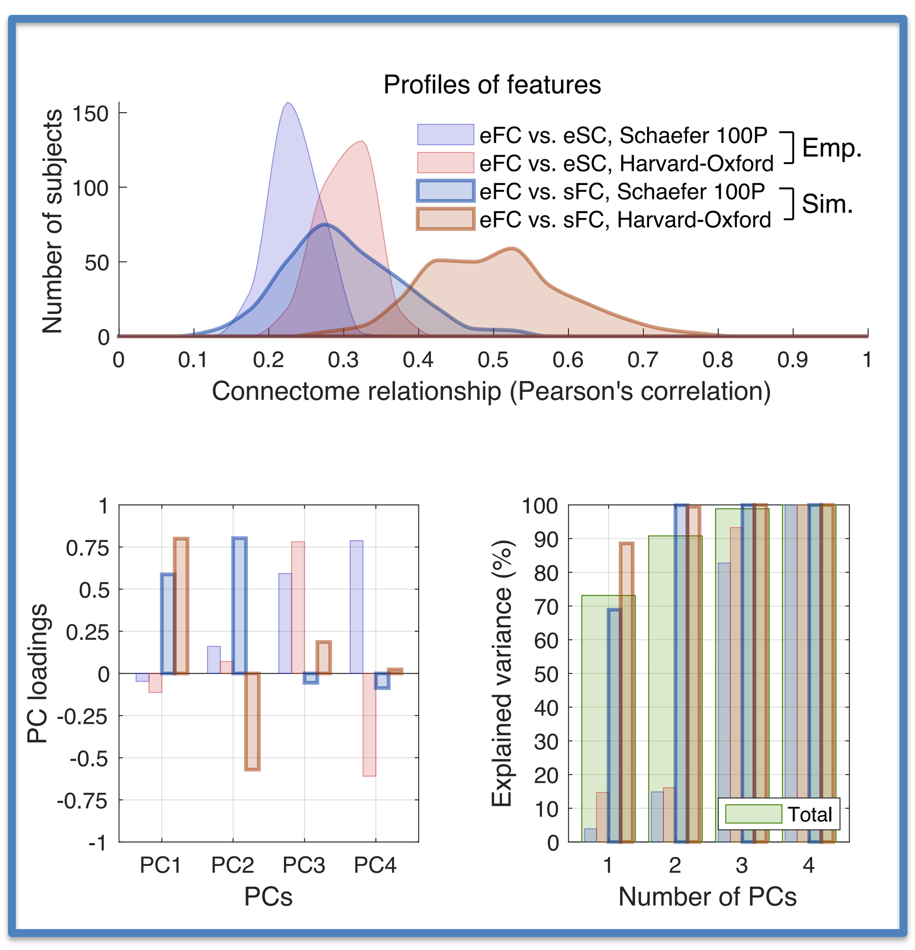
Step 5 **Machine learning for model-based prediction**

- Empirical feature: corr(eFC, eSC) and Simulated feature: corr(eFC, sFC)
- Cross-validated confound removal scheme (5-fold nested cross validation; n=100)

✓Classification of **females** and **males** (confound: brain volume)
✓Prediction of **cognitive composite score** and the **Big-Five personality traits** (confound: brain volume and age)

Results: Simulated features outperforming empirical features in machine-learning analysis

1. Empirical (Emp.) and simulated (Sim.) features



- Feature distributions across individual subjects for two brain parcellations.
- PC1 and PC2 are related to the simulated features and cumulatively explain 90% of the variance of features.
- PC2 and PC4 distinguish the two parcellation schemes.

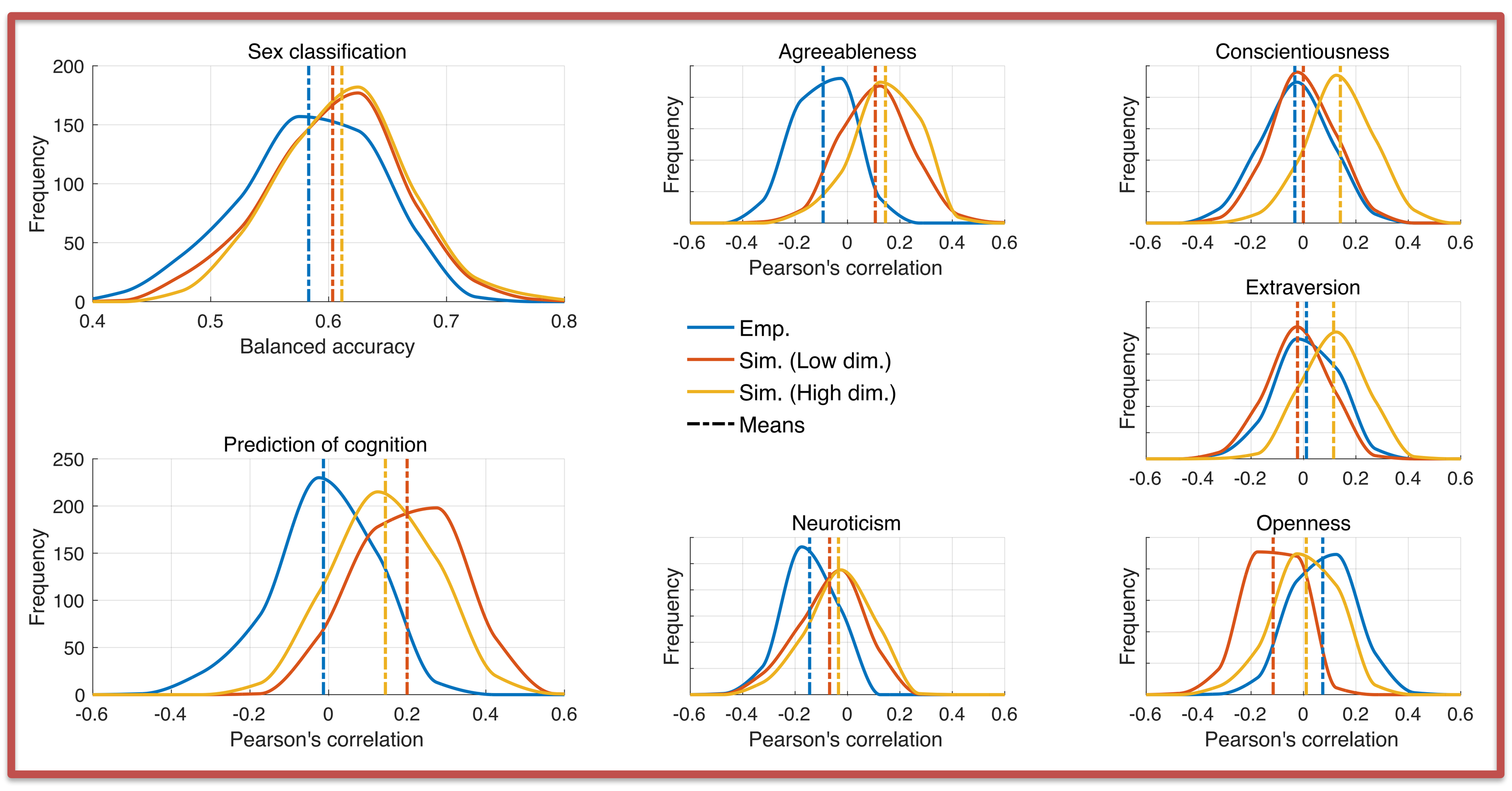
The best predictors in each target	*ES (p value)
Sex classification: Sim. (High dim.)	1.16 (0.00)
Cognition: Sim. (Low dim.)	0.84 (0.00)
Agreeableness: Sim. (High dim.)	0.76 (0.00)
Conscientiousness: Sim. (High dim.)	0.76 (0.00)
Extraversion: Sim. (High dim.)	0.63 (0.00)
Neuroticism: Sim. (High dim.)	-0.21 (0.00)
Openness: Emp.	0.46 (0.00)

*Effect size against the null distribution

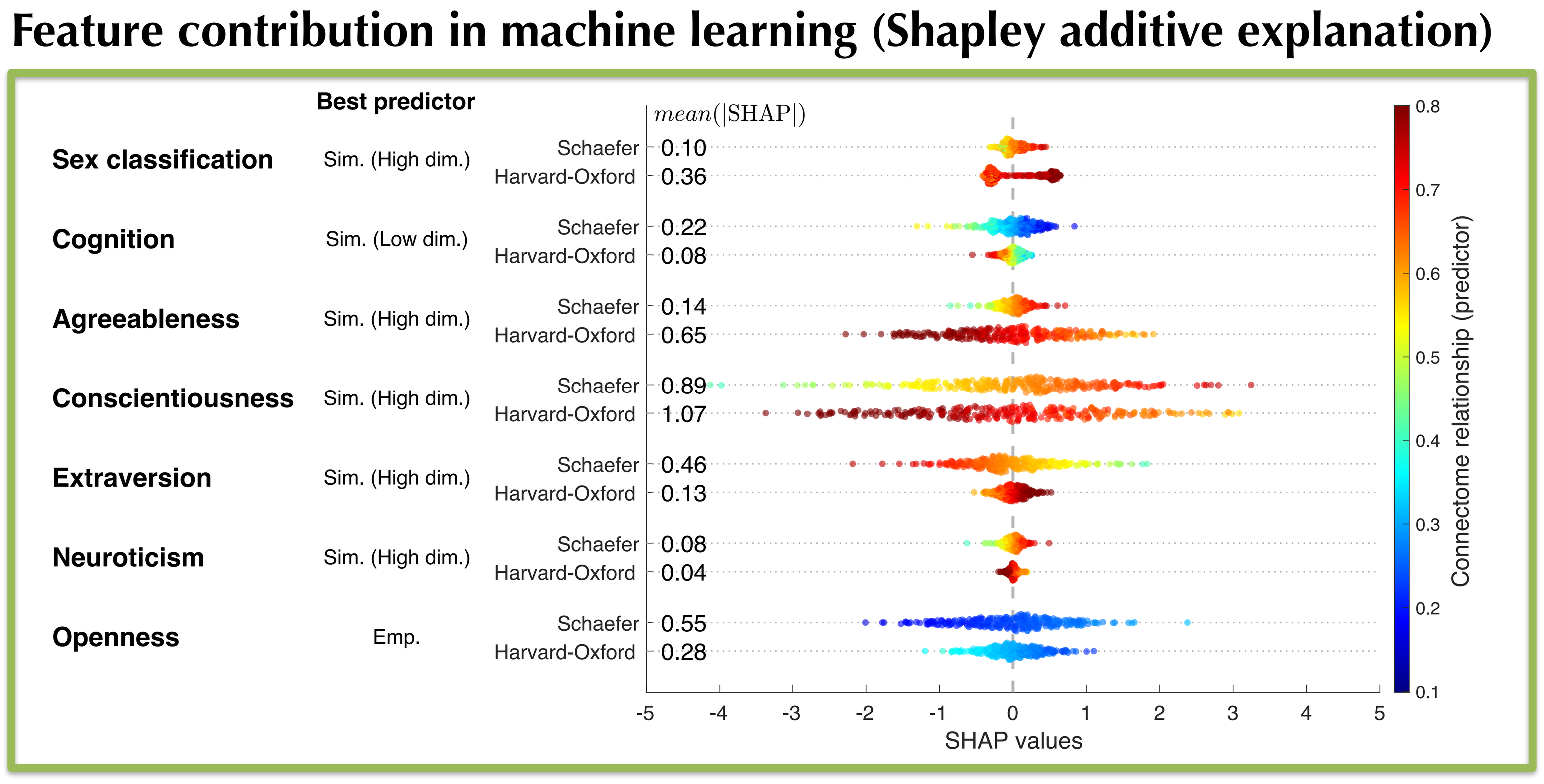
- Distributions of Shapley additive explanation* (SHAP) values **[6]** (data points = individual subjects).
- SHAP value itself does not indicate the magnitude of performance.
- Parcellation schemes can contribute to the prediction in different ways.

*A large |SHAP| means a strong contribution.

2. Machine-learning results



3. Feature contribution in machine learning (Shapley additive explanation)



Conclusion

- By incorporating **model-based features** alongside empirical data, we can explore brain connectomes and their interrelationships, thereby **enhancing performance** and bringing additional benefits in neuroimaging analysis.
- We propose to consider the simulated data as **an additional neuroimaging data modality** that captures distinct properties barely present in empirical data and can be integrated into machine-learning applications.

Acknowledgements and funding

The authors gratefully acknowledge computing time on the supercomputer JURECA at Forschungszentrum Jülich under grant ‘cjinm71’. The authors are also grateful to Vera Komeyer for helpful discussions about machine learning.

This work was supported by the Portfolio Theme Supercomputing and Modeling for the Human Brain by the Helmholtz association, the Human Brain Project and the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreements 785907 (HBP SGA2), 945539 (HBP SGA3) and 826421 (VirtualBrainCloud). Open-access publication was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) – 491111487. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the poster.

References

1. Messe, A., et al., Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities. PLoS Comput Biol, 2014. 10(3): p. e1003530.
2. Jung, K., et al., Whole-brain dynamical modelling for classification of Parkinson's disease. Brain Commun, 2023. 5(1): p. fcac331.
3. Jung, K., et al., Simulated brain networks reflecting progression of Parkinson’s disease. Network Neuroscience, 2024. 8(4): p. 1400-1420.
4. Hansen, N. and A. Ostermeier, Completely derandomized self-adaptation in evolution strategies. Evol Comput, 2001. 9(2): p. 159-95.
5. Wischnewski, K.J., et al., Exploring dynamical whole-brain models in high-dimensional parameter spaces. PLoS One, 2025. 20(5): p. e0322983.
6. Chen, H., et al., Algorithms to estimate Shapley value feature attributions. Nature Machine Intelligence, 2023. 5(6): p. 590-601.