Deep neural networks interpret white matter lesions as a signature of higher brain-age

Goltermann O.[1,2,3,4*], Hofmann S.M.[2,5,6,*], Villringer A.[1,2,7,8,9], Witte A.V.[2,7], Beyer F.[2]

[1] Max Planck School of Cognition, Leipzig, Germany, [2] Max Planck Institute of Human Cognitive and Brain Sciences, Department of Neurology Leipzig, Germany, [3] Max Planck Institute for Psycholinguistics, Language and Genetics Department, Nijmegen, Netherlands, [4] Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Germany, [5] Fraunhofer Institute Heinrich Hertz, Department of Artificial Intelligence, Berlin, Germany, [6] Technical University Berlin, Machine Learning Group, Berlin, Germany, [7] Clinic for Cognitive Neurology, University of Leipzig Medical Center, Leipzig, Germany, [8] Charité – Universitätsmedizin Berlin, Germany, [9] Berlin School of Mind and Brain, Berlin, Germany

* equal contribution

Introduction

- Relatively **little is known** which **specific brain features contribute** to deep neural network based brain-age estimations.
- To address this issue, Hofmann et al. (2021) used **Layer-wise Relevance Propagation** (LRP; Lapuschkin et al., 2019) on brain-age predicting multi-level ensembles (MLENS) of 3D-convolutional neural networks (Fig. 1) to **identify which brain features contribute** to brain-age estimations.
- Here, we computed binarized **white matter lesion** (WML) **probabilistic maps** of **1290 participants** from a population-based cohort study (LIFE-Adult; Loeffler et al., 2015; age range 18-82 years) using the Lesion segmentation toolbox (Schmidt et al., 2021).
- We hypothesized that the MLENS capture WMLs and **use them as a information source to predict higher brain-age**.

Methods

To **test whether the FLAIR sub-ensemble** of the MLENS (Fig.1) **uses WML as an information source for higher brain-age** we performed the following analyses steps:

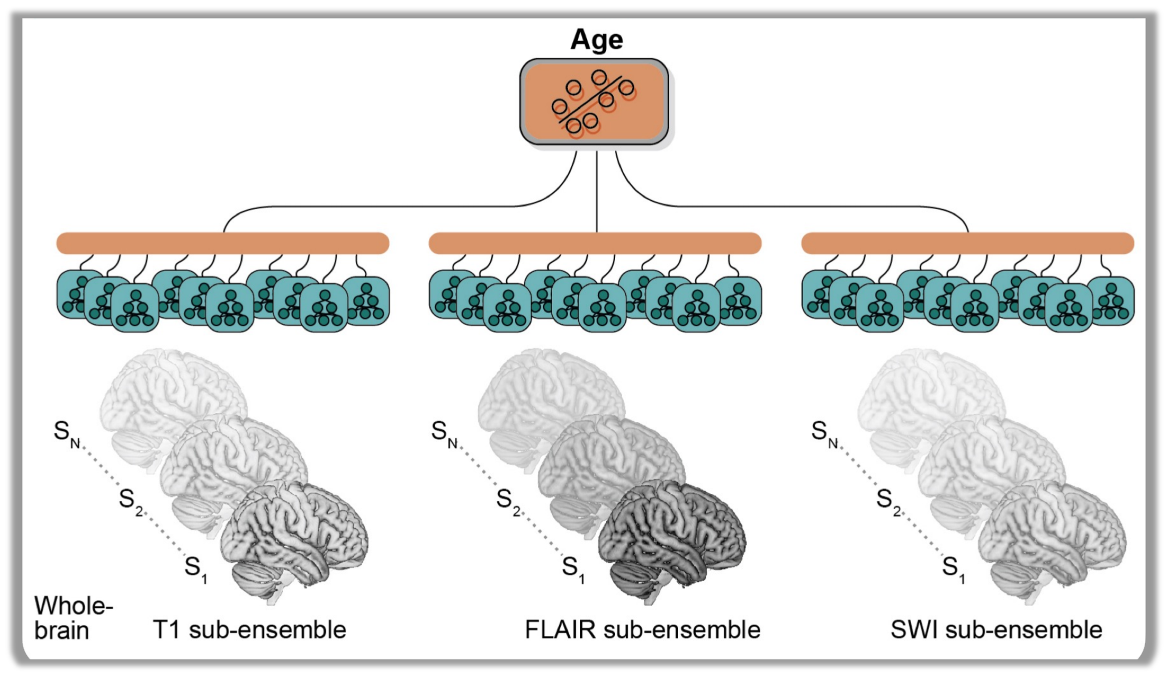

Fig. 1 Model architecture for brain-age estimations. (*Figure from Hofmann et al., 2021*)

1 LRP heatmaps (relevance maps) were computed on FLAIR images of all subjects and warped to the MNI152 space (2mm resolution).

2 For each subject the **WML probabilistic map** and the **relevance map** were aligned and **overlaid**.

3 In subjects with more than 30 WML voxels, **relevance values were averaged over WML voxels** and **compared to** the **expected relevance** per voxel [=average relevance of all brain-voxels with positive relevance].

Results

1 Identifying input features that drove the decision process

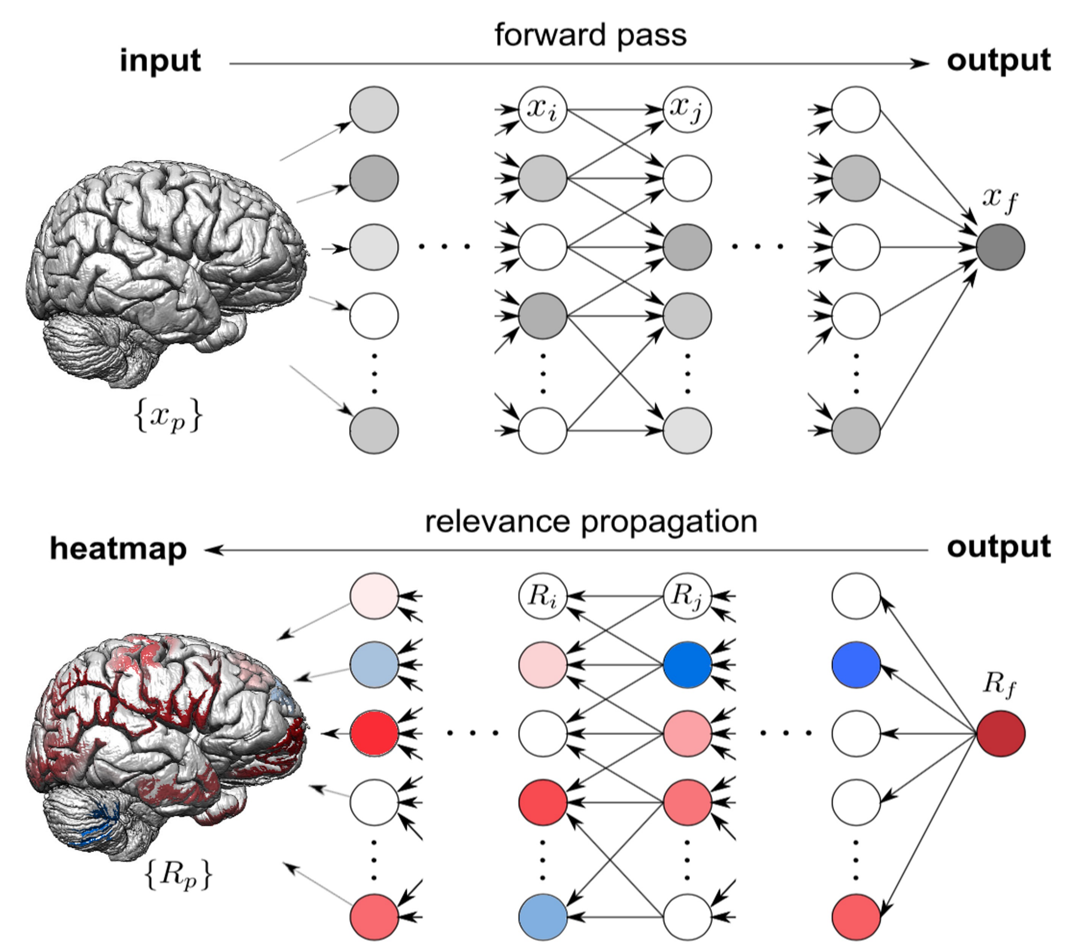

Fig. 2 Prediction step (*top*) and relevance propagation (*bottom*) highlighting individual brain areas that are relevant for the age estimation. *Figure adaptation from Montavon et al. (2017).*

2 Probabilistic WML and relevance map

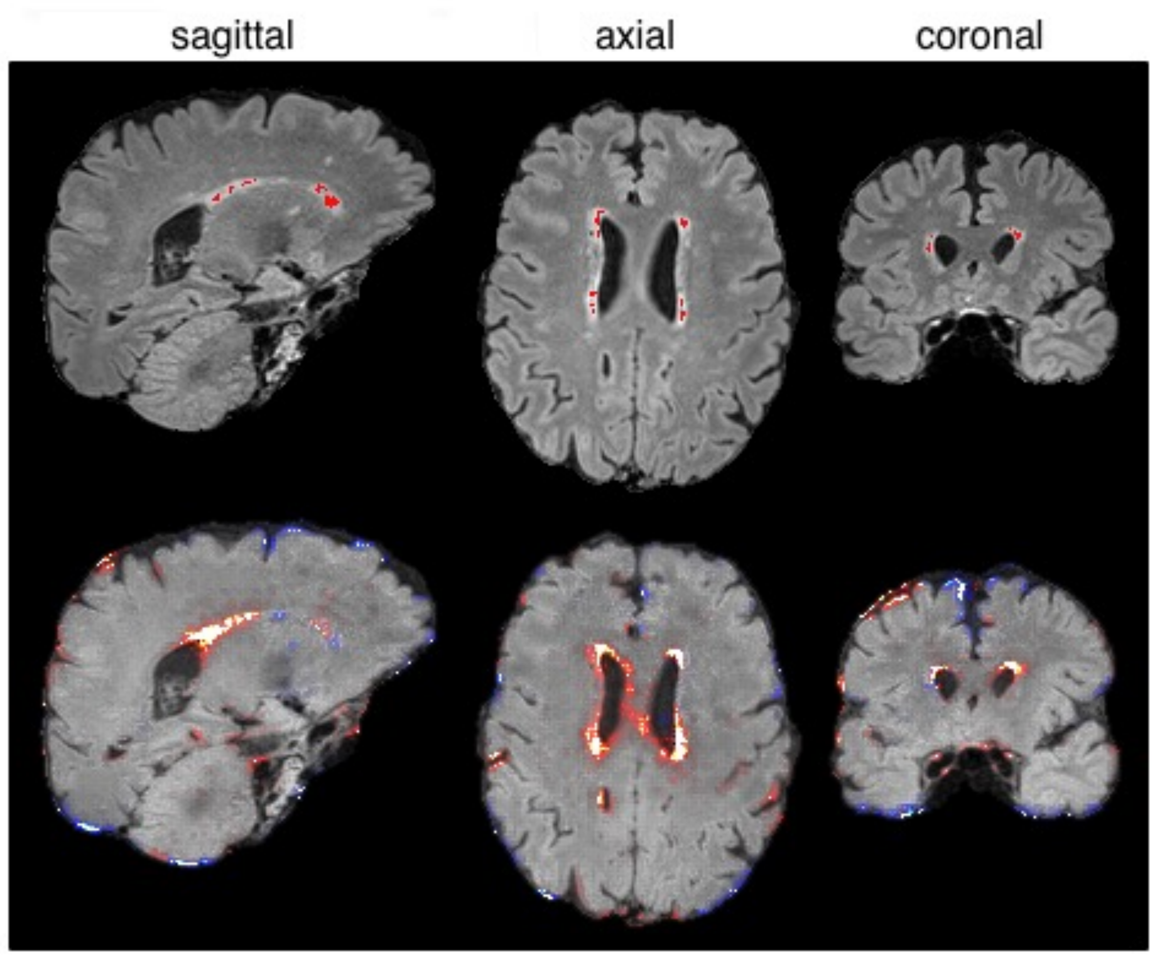

Fig. 3 *Top*: WML map (*red* voxels). *Bottom*: LRP map of a single subject (age = 68, predicted brain-age = 67.16). Here, *red-yellow* voxels indicate information indicating a higher age, and vice versa. Sum over all relevance represents the age estimate by the model.

3 Average relevance of WML voxels

Number of WML voxels > 30, n = 654

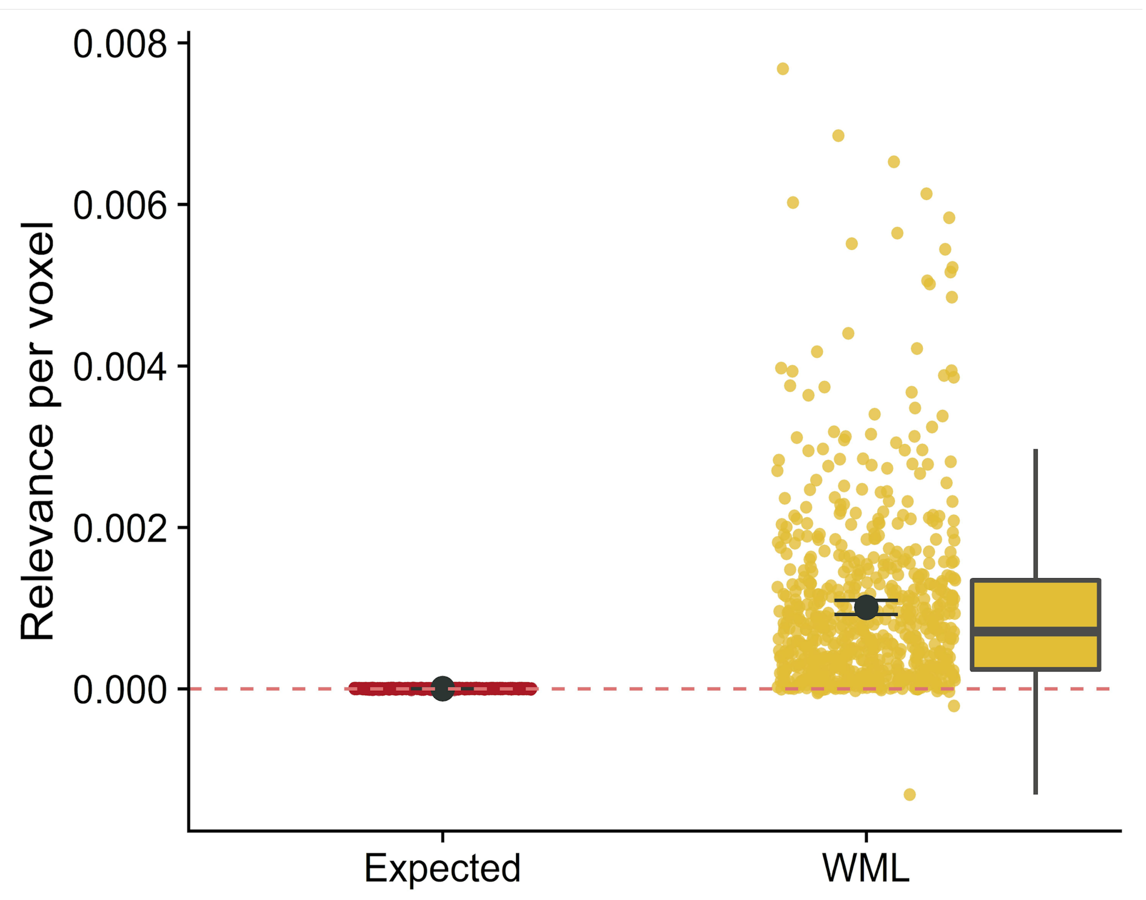

Fig. 4 Average relevance in WML voxels (*yellow*) and expected relevance per voxel (*red*). WML voxels contributed to higher brain-age estimates significantly more than the average brain voxel (M_{diff} = 0.001, *d* = 0.90, $t(653)$ = 22.95, p<.001).

Conclusion

- **Deep learning models capture WMLs** and associate higher brain age with a higher lesion load, underlining that these models are **capable of learning biologically relevant age-associated structural brain changes** while being trained end-to-end, that is on *relatively* raw MR images**.**
- However, we also found that brain-age estimates **do not exclusively rely on WML**.
- In future studies further known brain features, such as **gray matter volume** and **sulcal widening** should be studied with respect to the relevance maps.
- This would allow us to **test how much deep learning models rely on brain features that are known** to be related to aging. Conversely, this approach also **numerically and visually indicates unexplained variance** that could be studied further, given the relevance maps of our study.

References

Hofmann, S. M., Beyer, F., Lapuschkin, S., Loeffler, M., Müller, K.-R., Villringer, A. Samek W., & Witte, A. V. (2021). *Towards the Interpretability of Deep Learning Models for Human Neuroimaging. bioRxiv* https://doi.org/10.1101/2021.06.25.449906

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K.-R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*, *10*(1), 1096. https://doi.org/10.1038/s41467-019-08987-4

Loeffler, M., Engel, C., Ahnert, P., Alfermann, D., Arelin, K., Baber, R., Beutner, F., Binder, H., Brähler, E., Burkhardt, R., Ceglarek, U., Enzenbach, C., Fuchs, M., Glaesmer, H., Girlich, F., Hagendorff, A., Häntzsch, M., Hegerl, U., Henger, S., … Thiery, J. (2015). The LIFE-Adult-Study: Objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health*, *15*, 691. https://doi.org/10.1186/s12889-015-1983-z

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., & Müller, K.-R. (2017). Explaining nonlinear classification decisions with deep Taylor decomposition. Pattern Recognition, 65, 211–222. https://doi.org/10.1016/j.patcog.2016.11.008

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., Zimmer, C., Hemmer, B., & Mühlau, M. (2012). An automated tool for detection of FLAIR -hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage*, *59*(4), 3774–3783. https://doi.org/10.1016/j.neuroimage.2011.11.032

MAX PLANCK INSTITUTE
FOR HUMAN COGNITIVE AND BRAIN SCIENCES