Characterizing Personalized Neuropathology in Dementia and Mild Cognitive Impairment with Explainable Artificial Intelligence

Esten H. Leonardsen
University of Oslo

Yunpeng Wang
University of Oslo

Lars T. Westlye
University of Oslo

Thomas Wolfers
University of Tübingen

Introduction

With over 55 million people affected globally and a projected threefold increase in prevalence by 2050, dementia presents a paramount public health challenge for the coming decades. Deep learning applied to magnetic resonance imaging (MRI) scans have shown great promise for diagnosis and prognosis in dementia, but its clinical adoption is limited. This is partially attributed to the opaqueness of deep neural networks (DNNs), causing insufficient understanding of what underlies their decisions. Layerwise relevance propagation (LRP) is a technique for explaining the decision of DNNs via heatmaps highlighting regions of an image contributing to the prediction, potentially ameliorating the distrust impeding their clinical use. Furthermore, the explanations procured by LRP are highly individualized and could shed light on the specific manifestation of the disease in the brain, information which could prove crucial for accurate diagnosis and treatment.

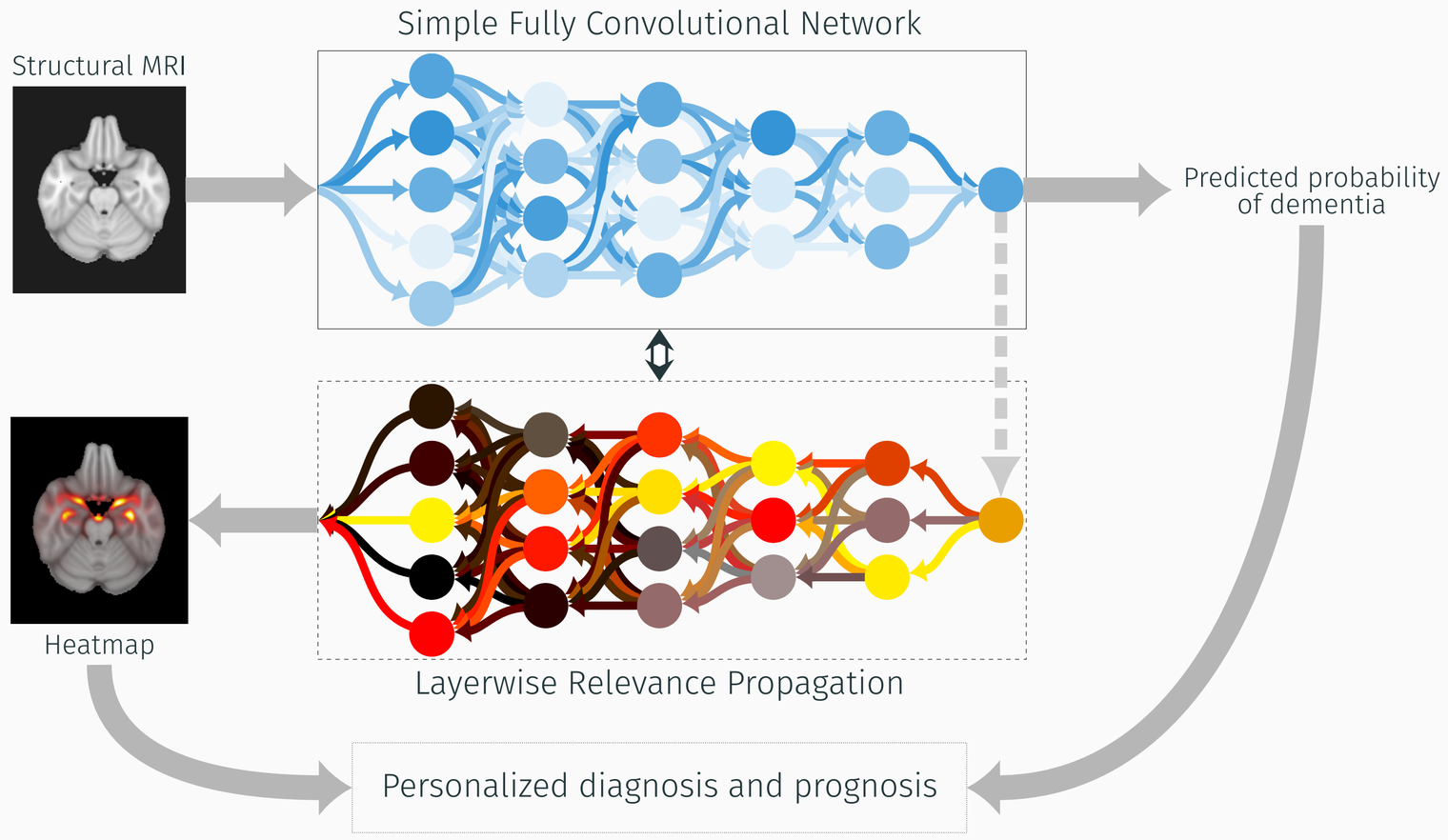

Figure 1: An overview of the explainable pipeline applied to a single individual.

Methods

We compiled structural MRI scans from a balanced set of 1708 dementia patients and healthy controls, and fit a simple fully convolutional network (SFCN) to differentiate between them. Next, we implemented LRP on top of the trained model to generate explanations in the form of heatmaps, accompanying its predictions. We validated the heatmaps by comparing an average map compiled from all true positives to a statistical reference map constructed with a GingerALE meta-analysis, containing spatial locations with observed deviations in dementia from 124 relevant publications. Following the validation, we employed the explainable pipeline in an exploratory analysis of 1256 patients with mild cognitive impairment (MCI). Here, we utilized its predictions and heatmaps to predict progression to dementia in the 5 years following the scan, and to investigate associations between spatial variability in the heatmaps and impairments in specific cognitive domains (Figure 1).

Results

- The best performing classifier was able to differentiate dementia patients from controls with an out-of-sample AUC of 0.9.

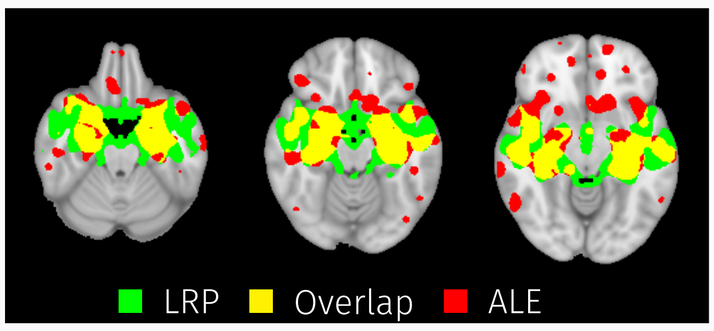

Figure 2: Three axial brain slices, showing the concordance between the average heatmap from our pipeline and the statistical reference map from GingerALE.

- The average heatmap for dementia patients highly resembled the statistical reference map (Figure 2), yielding a normalized cross-correlation of 0.64.
- In MCI patients, we predicted progression within 5 years with an AUC of 0.9 (Figure 4).
- Inter-individual variation in the heatmaps were associated with distinct patterns of performance on neuropsychological tests (Figure 3).

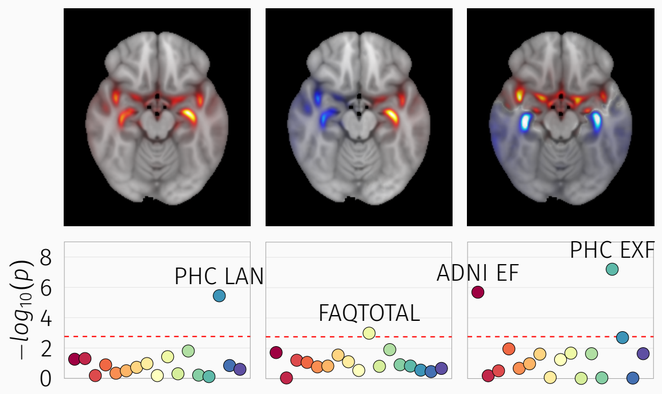

Figure 3: Three prototypical heatmaps and the strength of their associations with performance on cognitive tests. PHC LAN: Composite language score; FAQTOTAL: Functional Activities Questionnaire; ADNI EF/PHC EXF: Executive function scores

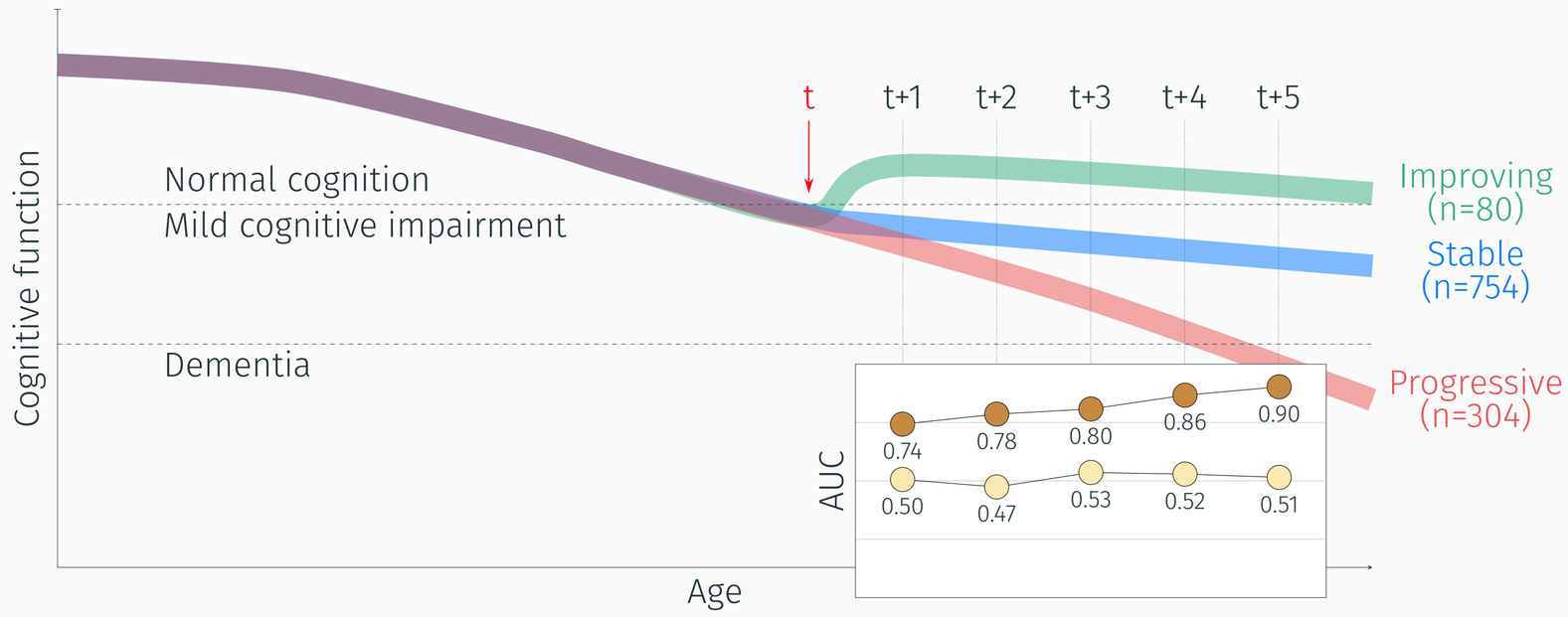

Figure 4: Clinical trajectories observed in the MCI sample. Embedded is the performance of the prognostic models for each year, the baseline model (○) and the model employing information from the pipeline (●).

Conclusion

Our explainable pipeline for dementia prediction allowed us to accurately **characterize the manifestation of dementia in individual patients**. When employing the pipeline in a sample of patients with MCI, information derived from it allowed us to **predict progression of the disease**, and revealed **associations between heterogeneity in the brain and impairments in distinct cognitive domains**. Our study presents an empirical foundation for further investigations into how explainable artificial intelligence can play an important role in precise personalized diagnosis of heterogeneous neurological disorders.

UNIVERSITETET I OSLO