Detecting Structural Brain Aberrations in Dementia Patients with Explainable Artificial Intelligence

Esten H. Leonardsen *University of Oslo*

Karin Persson *National Centre for Ageing and Health*

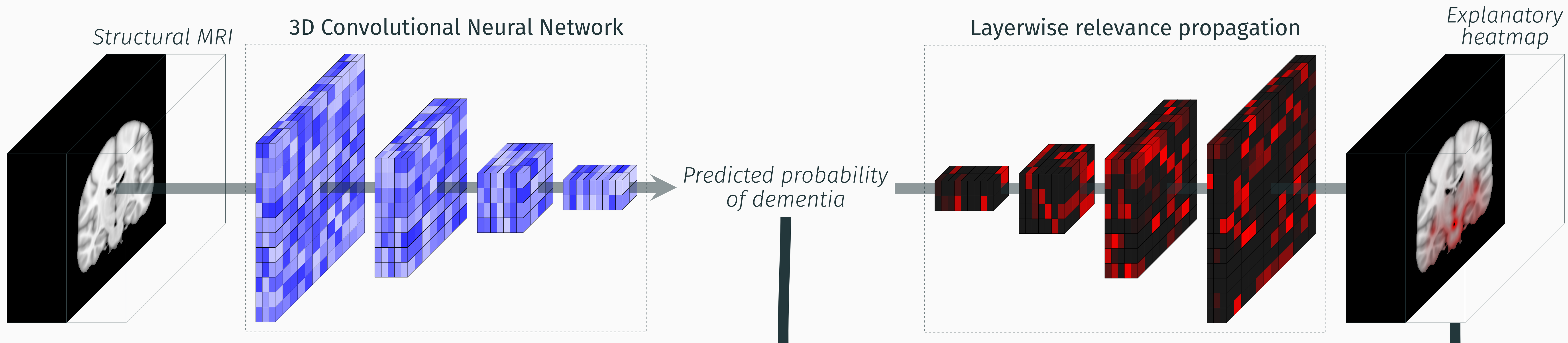
Geir Selbæk *National Centre for Ageing and Health*

Ole A. Andreassen *University of Oslo*

Lars T. Westlye *University of Oslo*

Thomas Wolfers *University of Tübingen*

Yunpeng Wang *University of Oslo*



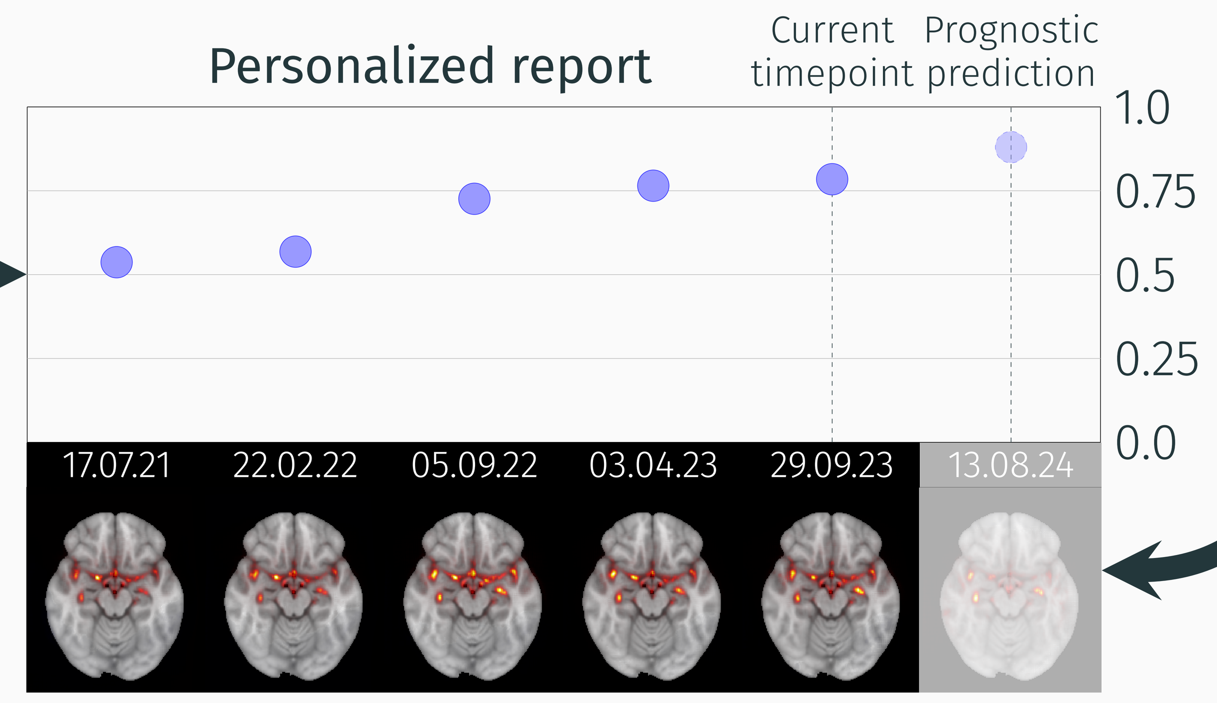


Figure 1: An overview of the proposed analytical pipeline. A structural T1-weighted magnetic resonance imaging (MRI) scan from a potential patient is given as input to a convolutional neural network (CNN). The CNN, trained to predict whether the scan originates from a patient or a healthy control, computes the probability that the patient has dementia at the current point in time. Next, layerwise relevance propagation is applied to procure a heatmap explaining this prediction, highlighting regions of the brain that contribute towards a predicted diagnosis. When compiled across visits, the predictions and heatmaps can be assembled into a personalized, longitudinal report, encoding trajectories of the absolute quantity, and localization, of dementia-related structural changes in the brain of the patient over time. By analyzing these trajectories, a prognostic prediction can be made about future clinical status. Furthermore, they contains precise information about which brain regions are afflicted, and how this is expected to change. Overall, this pipeline enables personalized diagnostics and prognostics in patients with dementia by providing detailed information about structural, disease-related changes in individual brains.

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. One of the arising needs is for technology and clinical methods to accurately characterize the disease in individual patients from a heterogeneous patient group. Deep learning applied to magnetic resonance imaging (MRI) scans have shown great promise for diagnosis and prognosis in dementia, but its clinical adoption is still limited. This can partially be 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 by computing heatmaps that highlight regions of an image contributing to a prediction, potentially ameliorating concerns impeding their clinical use. Furthermore, the explanations procured by LRP are highly individualized and can shed light on the specific manifestation of the disease in the individual brain, information that could prove crucial for accurate diagnostics and treatment.

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, to accompany and complement its predictions. We validated the heatmaps by comparing an average map compiled from all correctly predicted dementia patients with a statistical reference map., constructed by a GingerALE meta-analysis to reveal spatial locations with known structural changes in dementia patients based on 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 towards two clinically relevant tasks: to predict progression to dementia in the 5 years following the scan, and to investigate associations between spatial variability in the heatmaps and the structural abberations they encode, and impairments in specific cognitive domains.

Results

- The best performing classifier was able to differentiate dementia patients from controls with an out-of-sample AUC of 0.9.
- The average heatmap highly resembled the statistical reference map, yielding a normalized cross-correlation of 0.64 (Figure 2).
- In MCI patients, information derived from the pipeline allowed us to predict progression within 5 years with an AUC of 0.9 (Figure 3).
- Spatial patterns in the heatmaps were associated with distinct patterns of performance on neuropsychological tests.

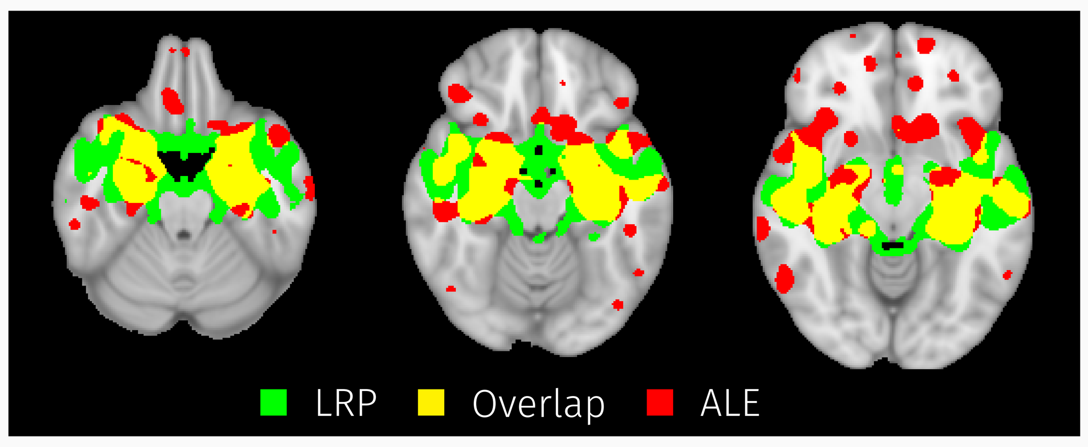


Figure 2: Overlap between heatmaps and reference map.

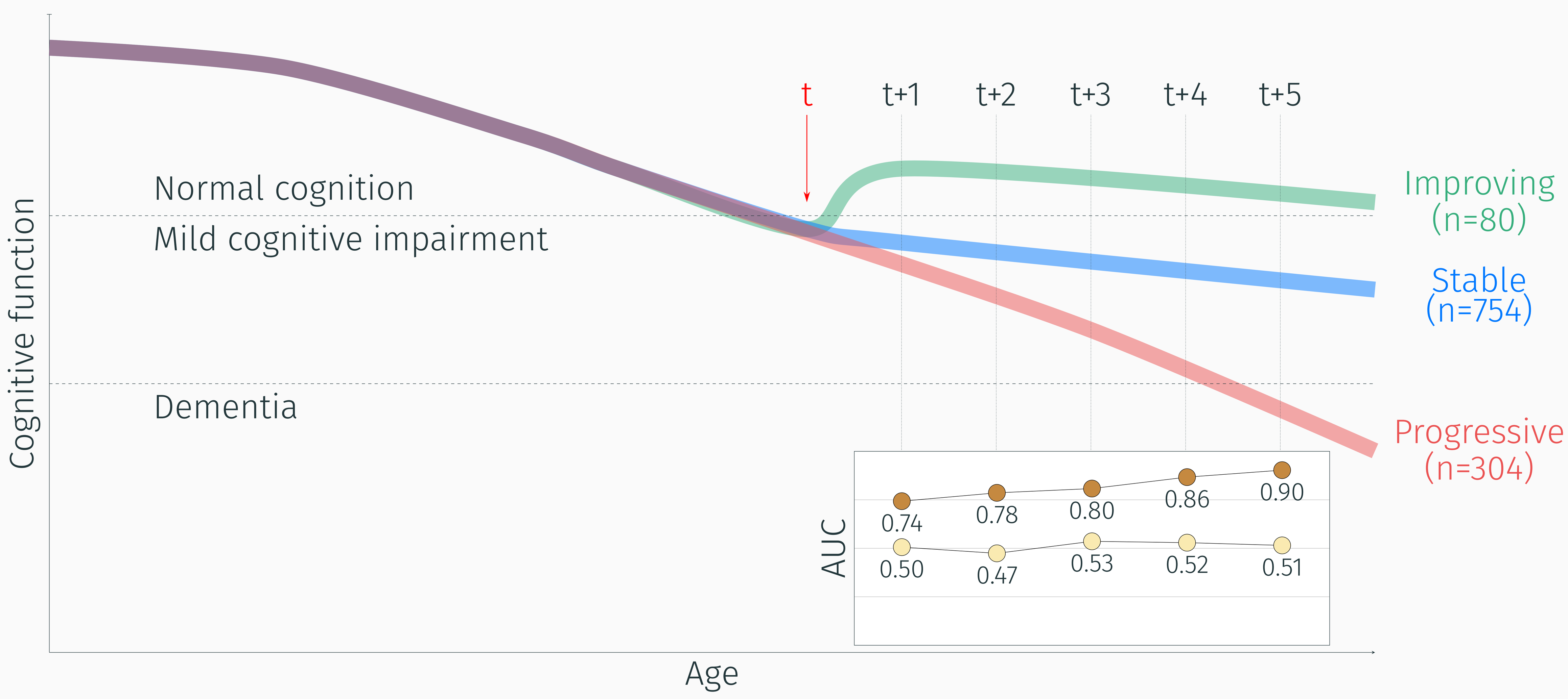


Figure 3: Conceptual overview and results from the prognostic predictive analysis.

Conclusion

Our explainable pipeline for dementia-prediction allowed us to **precisely characterize the manifestation of dementia in the brains of individual patients**. Employing the pipeline in a sample of patients with MCI, we used information derived from the pipeline to **accurately predict progression of the disease**, and **reveal associations between brain heterogeneity and impairments in cognitive domains**. Our study presents an empirical foundation for how explainable artificial intelligence can support personalized diagnostics and prognostics of dementia.

Competing interests

EHL is the CTO and a major shareholder in baba.vision, a company developing clinial decision support systems for neurological disorders. LTW, TW and YW are shareholders in baba.vision.

UNIVERSITETET I OSLO