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

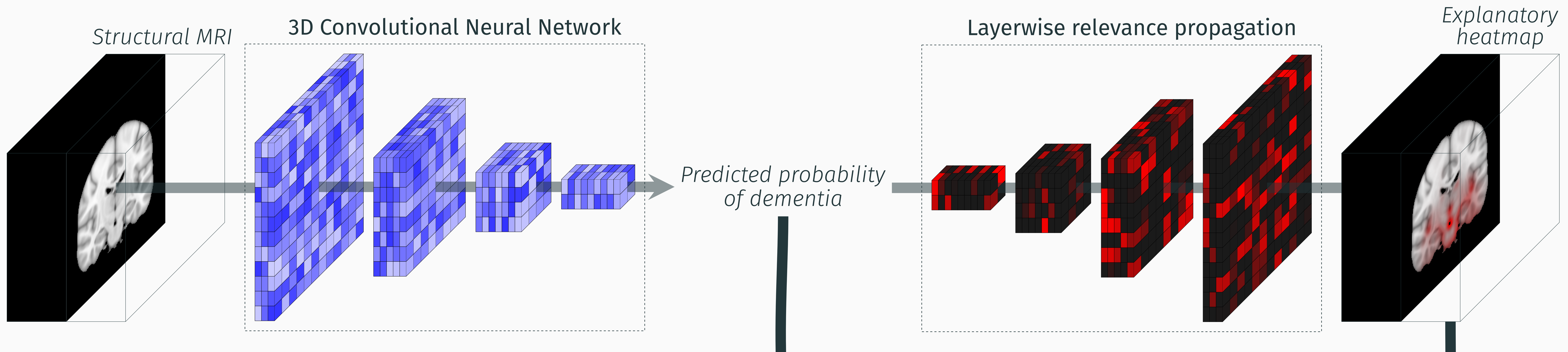
Geir Selbæk
National Centre for Ageing and Health

Ole A. Andreassen
University of Oslo

Lars T. Westlye
University of Oslo

Thomas Wolfers
University of Tübingen

Yunpeng Wang
University of Oslo



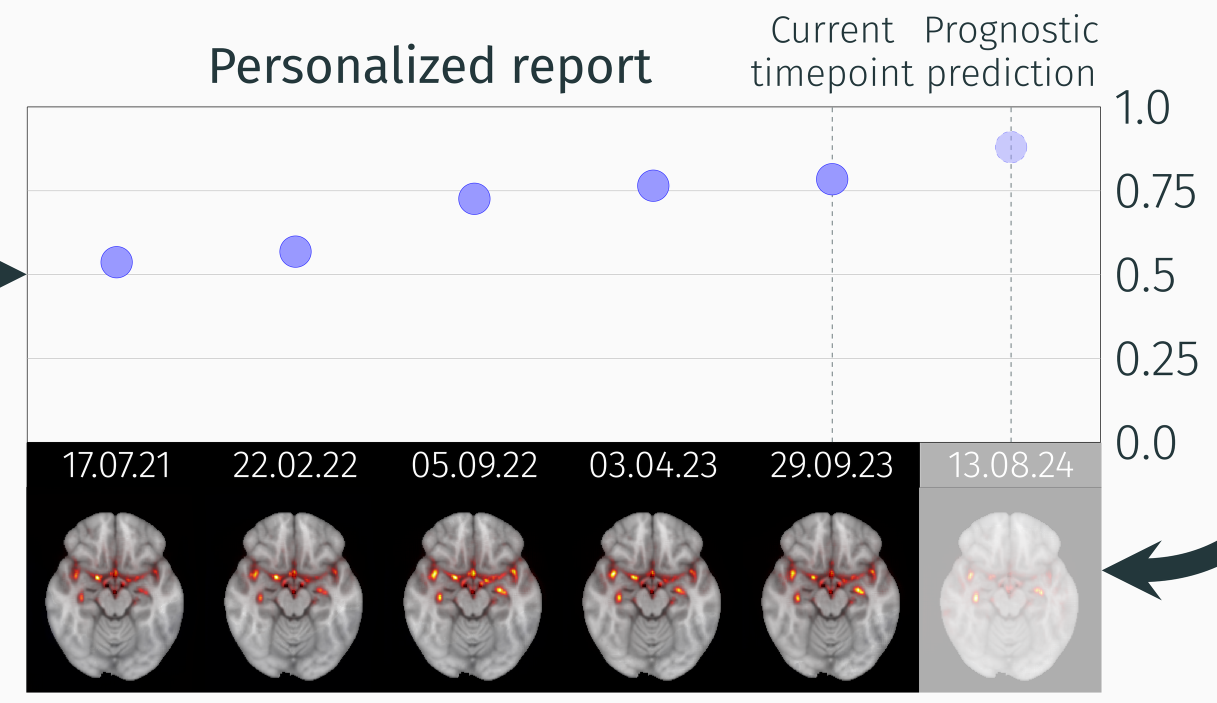


Figure 1: An overview of the proposed analytical pipeline. A structural T1-weighted magnetic resonance imaging (MRI) scan from a potential patient is given as input to a convolutional neural network (CNN). The CNN, trained to predict whether the scan originates from a patient or a healthy control, computes 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 contain precise information about which brain regions are afflicted and how this is expected to change. Overall, the pipeline enables personalized diagnostics and prognostics in patients with dementia by providing detailed information about structural, disease-related changes in the brains of individual patients.

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 emerging 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 has 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 matched healthy controls, and fit a simple fully convolutional network (SFCN) to differentiate between them. Next, we constructed a comprehensive analytical pipeline by implementing 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 to a statistical reference map, composed using 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 perform two clinically relevant tasks: predict progression to dementia within 5 years following the scan, and investigate associations between spatial variability in the heatmaps 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 variability in the heatmaps was associated with distinct performance profiles on neuropsychological tests.

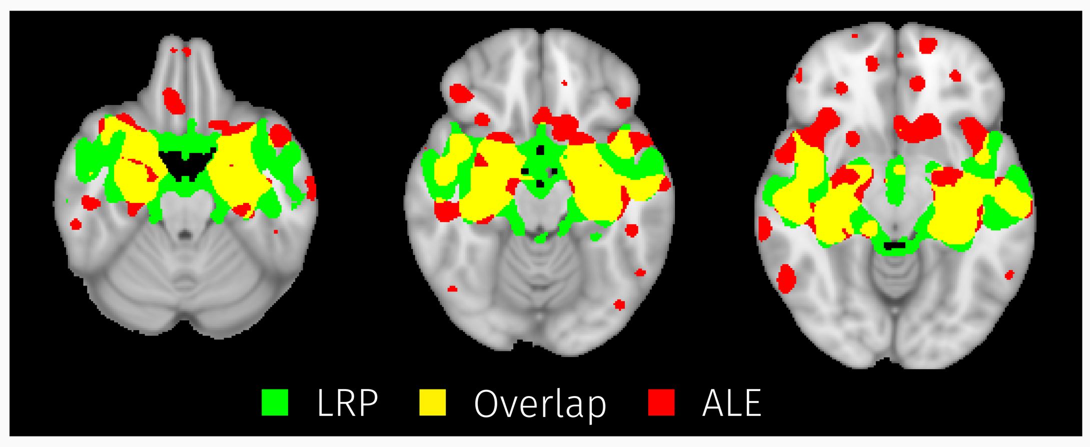


Figure 2: Overlap between the average and reference heatmaps.

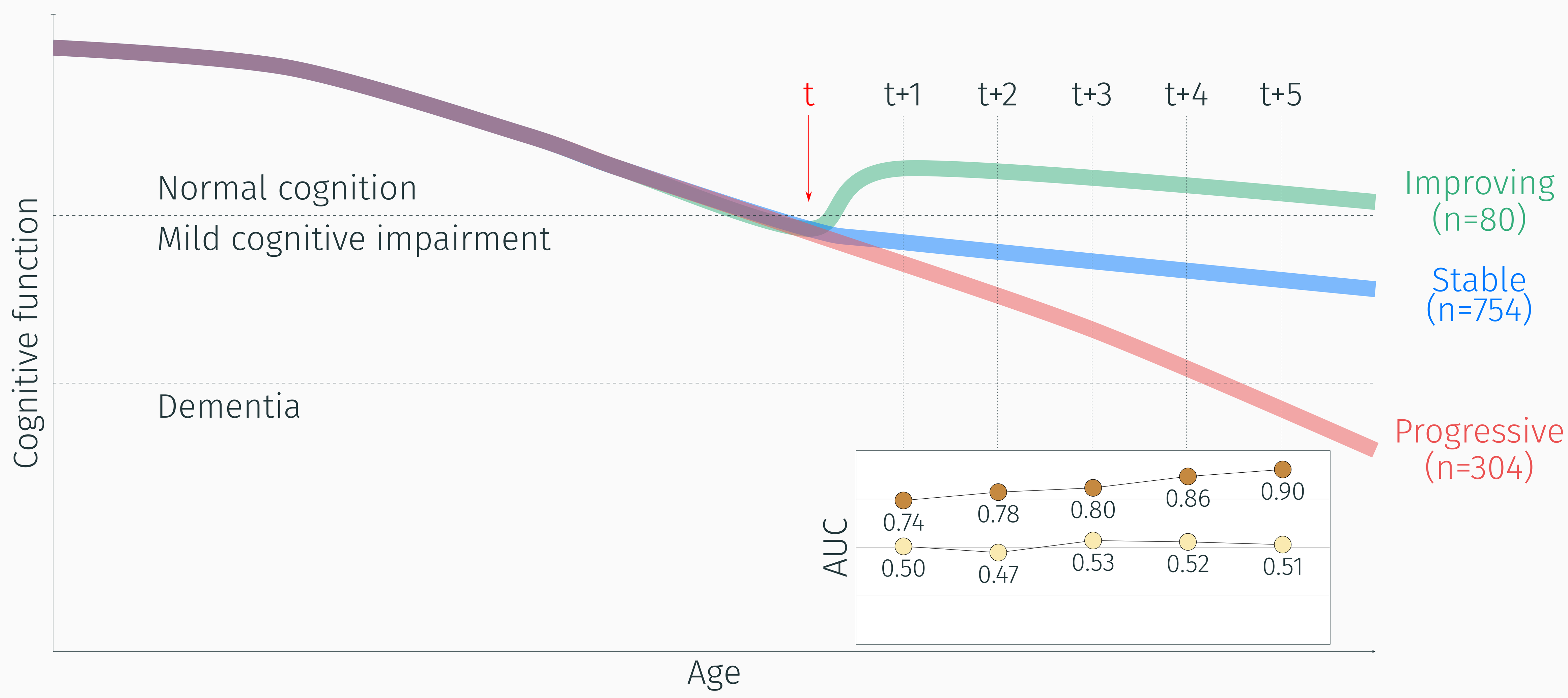


Figure 3: Conceptual overview and results from the prognostic predictive analysis, predicting progression from MCI to dementia.

Conclusion

Our explainable pipeline for dementia-prediction allowed us to **accurately characterize the manifestation of dementia in the brains of individual patients**. In a sample consisting of MCI patients, we used information derived from the pipeline to **predict progression of the disease**, with a predictive performance that surpassed common thresholds recommended for technologies to provide clinical value. Furthermore, the rich spatial information in the heatmaps procured by the pipeline revealed **associations between brain heterogeneity and impairments in cognitive domains**. Our study represents an empirical foundation for how explainable artificial intelligence can support clinical personnel to realize personalized diagnostics and prognostics in dementia.

Disclosures

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

UNIVERSITETET I OSLO