

Dominant Trajectories of Aging-related Brain Atrophy Identified in 48,949 Individuals via a coupled Cross-sectional and Longitudinal Non-negative Matrix Factorization

Ioanna Skampardoni¹ | Guray Erus¹ | Ilya M. Nasrallah¹ | Zhijian Yang¹ |
Kyunglok Baik¹ | Haochang Shou¹ | Konstantina Nikita² | Christos Davatzikos¹

¹Centre for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, USA

²School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Attiki, Greece

Correspondence

Ioanna Skampardoni, Centre for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, USA.
Email: ioannaki2010@gmail.com

Abstract

Background: Understanding brain aging heterogeneity is critical for early detection of neurodegeneration related to underlying neuropathologies and effective recruitment for clinical trials. Machine learning techniques have shown potential in this domain but often rely on cross-sectional data, neglecting dynamic observations of pathological changes. This study applies Coupled Cross-sectional and Longitudinal Non-negative Matrix Factorization (CCL-NMF), a novel framework integrating static and dynamic brain changes, to analyze aging-related brain atrophy heterogeneity in a large, diverse dataset from 12 neuroimaging studies consolidated by the iSTAGING consortium (Skampardoni et al., 2024).

Method: CCL-NMF uses a mutually constrained NMF framework to identify brain aging components, combining population-level aging effects from cross-sectional data with individual-specific dynamics from longitudinal data. Individual expression levels (loadings) for each component are estimated, optimizing the reconstruction of both data types to capture the interplay between static and dynamic aspects of brain alterations.

Structural MRI data from 48,949 individuals ≥ 50 years were analyzed. Comparative analyses were conducted against a weakly-supervised generative adversarial network model, Surreal-GAN (Yang et al., 2024), which relies solely on cross-sectional data. Predictive performance for biomarkers, clinical variables, and disease progression was assessed using regression and Cox proportional hazards models with 5-fold stratified cross-validation.

To facilitate application to external datasets without retraining or harmonization, out-of-sample regression-based estimation of CCL-NMF loadings was implemented via

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Alzheimer's Association. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

NiChart (https://cbica.github.io/NiChart_Project/) with reliability validated through Pearson correlations with original loadings.

Result: CCL-NMF identified seven distinct brain atrophy components associated with Alzheimer's disease (AD), cognitive decline, and cardiovascular risk factors. Comparative analyses with the five Surreal-GAN components revealed consistent representations between the two models, elucidating well-reproducible brain atrophy components (Figure 1A). Importantly, CCL-NMF provided a richer representation and demonstrated improved predictive performance across various outcomes, including AD, cardiovascular disease markers, and disease progression (Figure 1B). Furthermore, Spearman correlations between original and approximated CCL-NMF loadings demonstrated high reliability, enabling seamless out-of-sample usage (Figure 2).

Conclusion: CCL-NMF offers a robust, interpretable framework for understanding brain aging and neurodegeneration by integrating cross-sectional and longitudinal data. It outperforms cross-sectional approaches, delivering richer representations with superior predictive accuracy, and supports easy application to external datasets through regression-based loadings estimation under a web-accessible server (neuroimagingchart.com).

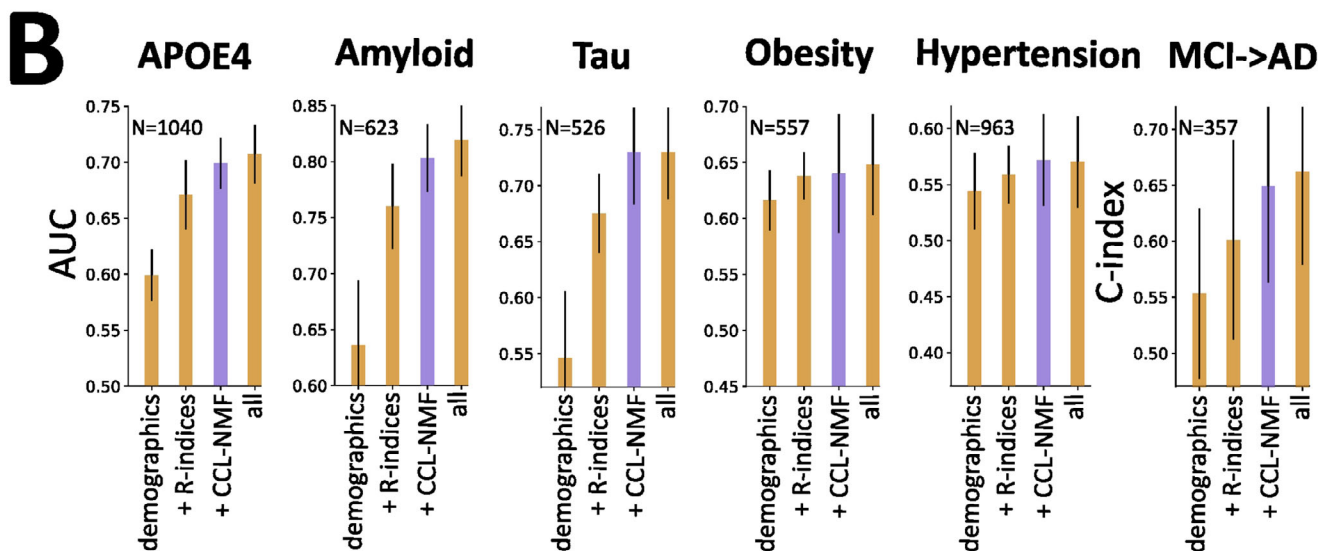
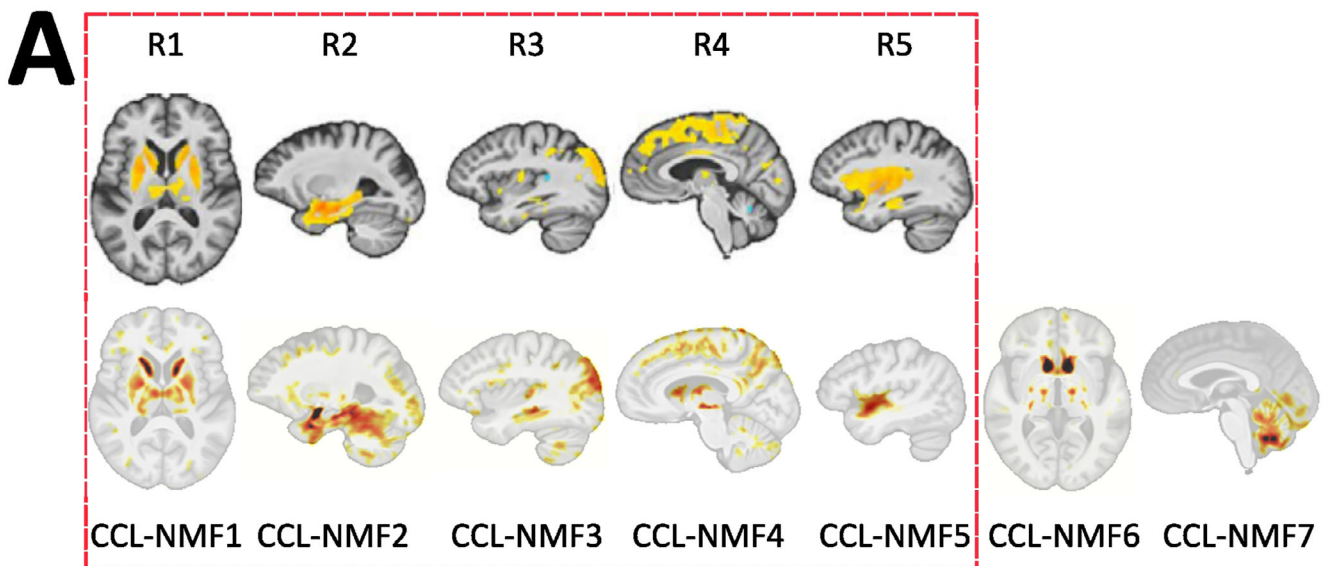


Figure 1: A) Atrophy patterns for CCL-NMF vs. Surreal-GAN components shown via voxel-wise t-tests performed for each CCL-NMF (Surreal-GAN) component while adjusting for age, sex, intracranial volume, and the remaining CCL-NMF (Surreal-GAN) components. False discovery rate correction was conducted to adjust multiple comparisons with a p-value threshold of 0.001. Increasing voxel redness indicates stronger association with a specific component. The first five CCL-NMF components show consistency with the five Surreal-GAN components. B) Estimation of Alzheimer's disease biomarkers, cardiovascular disease markers, and disease progression from the imaging components. Area under the curve (AUC) and concordance index (C-index) for models using 1) only demographics, 2) demographics and R-indices, 3) demographics and CCL-NMF loadings, and 4) all previous variables as predictors. Age, sex, and study were included as covariates in all models.

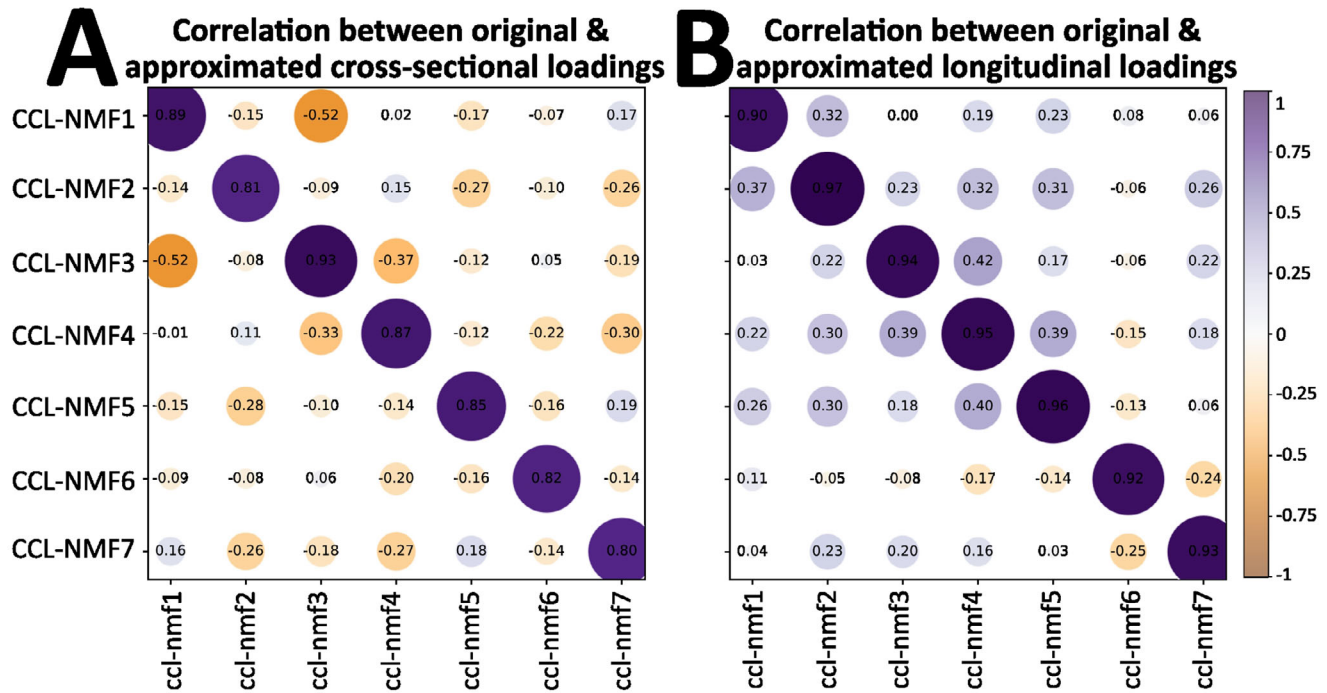


Figure 2: How well can these components' loadings be approximated using simple regression? Spearman correlations between original and approximated A) cross-sectional and B) longitudinal loading coefficients. Uppercase (lowercase) letters stand for original (approximated) loading coefficients.