Federated learning of neurodegenerative disease progression

Michelle Wang[1], Francesco Cremonesi[2], Lucie Chambon[2], Nikhil Bhagwat[1], Sergen Cansiz[2], Marco Lorenzi[2], Jean-Baptiste Poline[1]

[1] McGill University, Montreal, Canada
[2] Inria Center of Université Côte d'Azur, Sophia Antipolis, France

Email: michelle.wang6@mail.mcgill.ca, jean-baptiste.poline@mcgill.ca

Introduction

The study of the **progression of neurodegenerative diseases** such as Alzheimer's disease (AD) or Parkinson's disease (PD) **requires large, longitudinal sample sizes**.

Due to the clinical nature of these datasets, traditional data-sharing approaches may face **barriers related to data privacy and governance concerns**.

Federated learning is a method that enables the decentralized fitting of machine learning models, avoiding the need to share participant-level information.

We present a **federated disease progression model** based on the Disease Progression Modeling and Stratification (DPMoSt) model by Viani et al. (2025). We apply the model to synthetic data and to the Alzheimer's disease neuroimaging initiative (ADNI) dataset (Jack et al., 2008).

Neuroinformatics infrastructure for federated analyses

Nipoppy — Organize Process
Neurobagel — Harmonize Discover
Fed-BioMed — Federate Analyze

Want to learn more about this ecosystem? **Check out our preprint!**

Methods

Model

We model each participant i's standardized level for biomarker j over time as a logistic function:

$$b_{i,j}(t) = \frac{1}{1 + e^{-k_j(t+\tau_i-m_j)}} + \epsilon$$ where

k_j is the biomarker-specific **steepness**
m_j is the biomarker-specific **midpoint**
τ_i is the participant-specific **time shift**
$\epsilon \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise

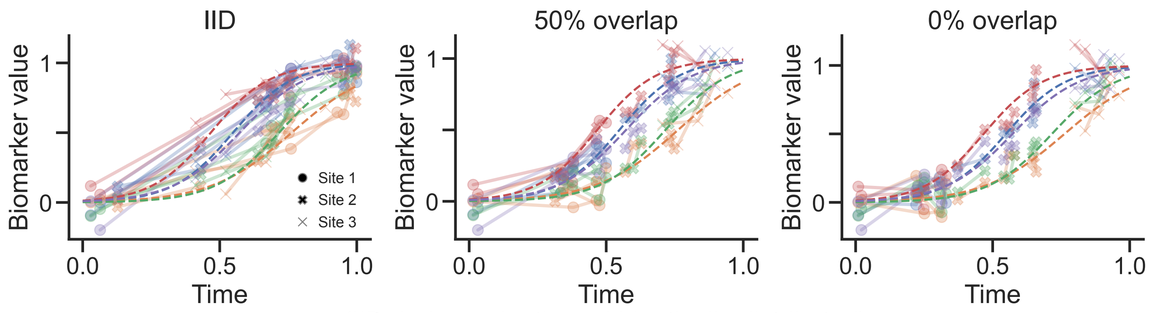
Federated learning framework

Fed-BioMed (Cremonesi et al., 2025)

- Client nodes fit local models independently at each site
- Local parameters (excluding time shifts) are sent to server and aggregated using the **FedAvg algorithm** (McMahan et al., 2023)
- Global parameters are sent back to clients for further training

Synthetic data experiments

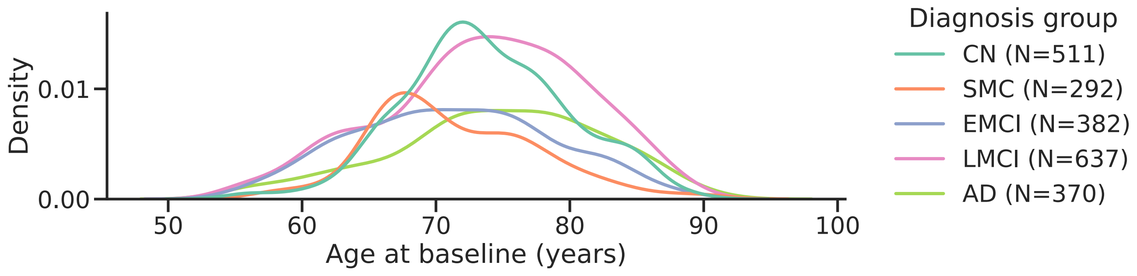
- 3 sites, 5 biomarkers, 50 participants per site (150 total)
- Up to 3 timepoints per participant
- $k_j \in [5, 10)$ and $m_j \in [0, 1)$ from uniform distributions
- 3 conditions with various degrees of site heterogeneity (τ_i range)



(IID: independently and identically distributed)

ADNI experiments

- **2192 participants**
- T1w image-derived FreeSurfer 7.1.1 **cortical thickness** measures for **5 regions**: parahippocampal, entorhinal, middletemporal, inferior parietal, and precuneus.
- Divided into **5 sites**, either randomly (**IID**) or based on ADNI site information (**non-IID**)

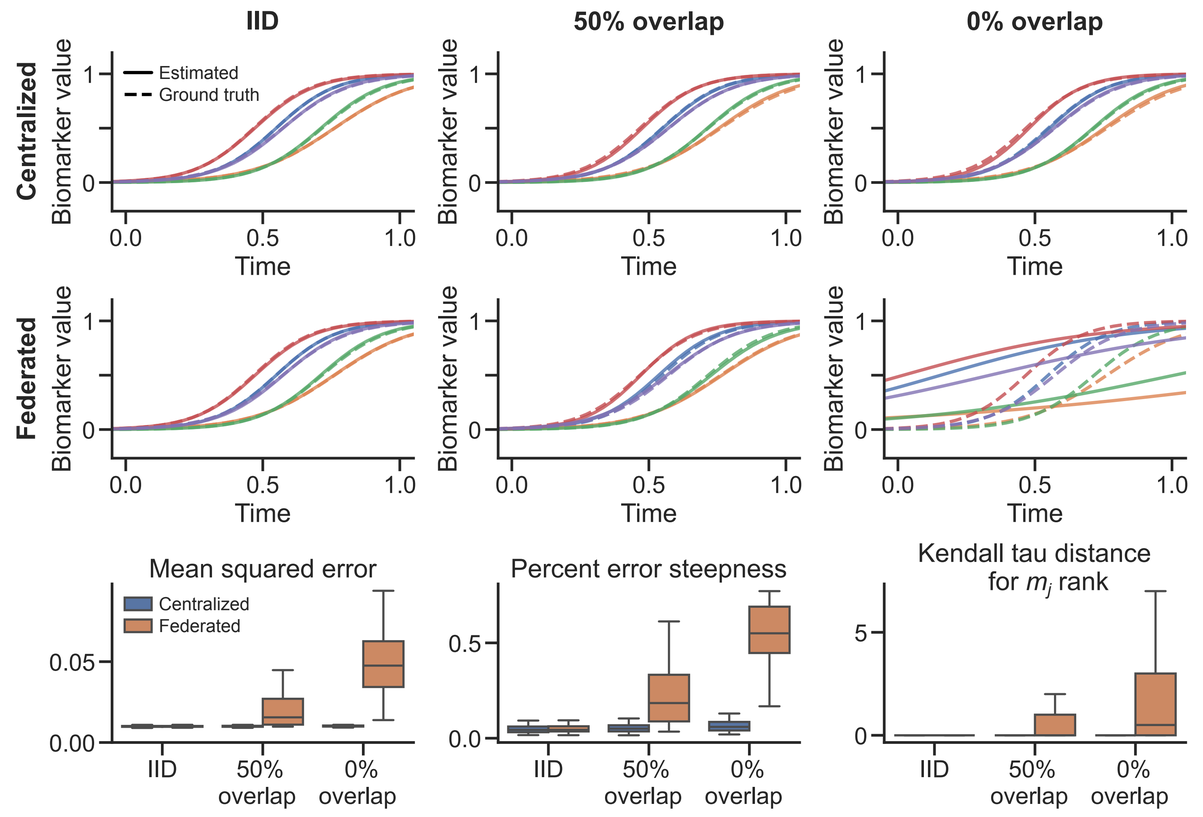


(CN: control, SMC: subjective memory complaint, EMCI: early mild cognitive impairment, LMCI: late mild cognitive impairment, AD: Alzheimer's disease)

Results

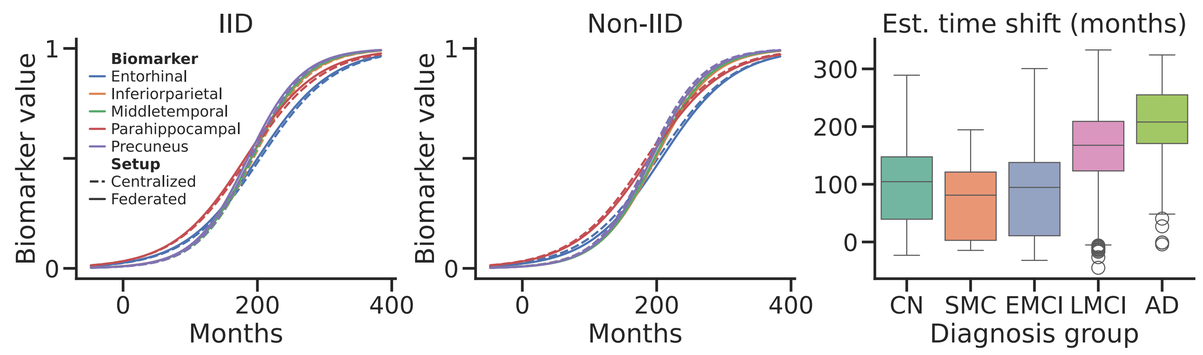
Synthetic data experiments

Federated model fits are worse than centralized model fits for non-IID distributions, but relative biomarker changes are still largely captured.



ADNI experiments

Federated model fits match centralized model fits (IID and non-IID). Estimated time shift values track with diagnosis group.



Conclusion and next steps

Preliminary results suggest that the **federated model performs similiarly to the centralized model** except in extreme conditions.

When applied to brain biomarkers extracted from magnetic resonance imaging (MRI), the participant-specific time shift values estimated by the model **track with disease status**.

Next steps include incorporating additional information (e.g. age) in the model and extending the simulation and ADNI scenarios.

References

Cremonesi, F. et al. (2025). Fed-BioMed: Open, Transparent and Trusted Federated Learning for Real-world Healthcare Applications. In M. H. ur Rehman & M. M. Gaber (Eds.), Federated Learning Systems: Towards Privacy-Preserving Distributed AI (pp. 19–41). Springer Nature Switzerland.

Jack, C. R. et al. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging: JMRI, 27(4), 685–691.

McMahan, H. B. et al. (2023). Communication-Efficient Learning of Deep Networks from Decentralized Data (arXiv:1602.05629). arXiv.

Viani, A. et al. (2025). Disease Progression Modeling and Stratification for detecting sub-trajectories in the natural history of pathologies: Application to Alzheimer's Disease trajectory modeling. Imaging Neuroscience.

Acknowledgements

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu)