Benchmarking federated learning approaches against siloed and mega-analysis regimes

Michelle Wang, Nikhil Bhagwat, Alyssa Dai, Arman Jahanpour,
Brent McPherson, Sebastian Urchs, Jean-Baptiste Poline

McGill University, Montreal, Canada

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

Introduction

Many clinical neuroscience datasets **cannot be openly shared** due to institutions and countries adopting strong data privacy frameworks (Marelli & Testa, 2018): these dataset remain in so-called **"data silos"**.

Decentralized data processing tools and **federated** analysis methods could enable large-scale, multi-site studies that make use of these datasets while **respecting data sharing constraints**.

We compare a simple **federated** machine learning setup (i.e. sharing only fitted models) with two traditional experimental setups – **siloed** (no sharing) and **mega-analysis** (sharing data).

Hypothesis: model generalizability improves as we go from siloed to federated to mega-analysis setups.

Neuroinformatics tool for multicentric analyses

Nipoppy framework for the organization and **decentralized processing** of datasets.

Neurobagel ecosystem for distributed data **harmonization** and **search**.

Methods

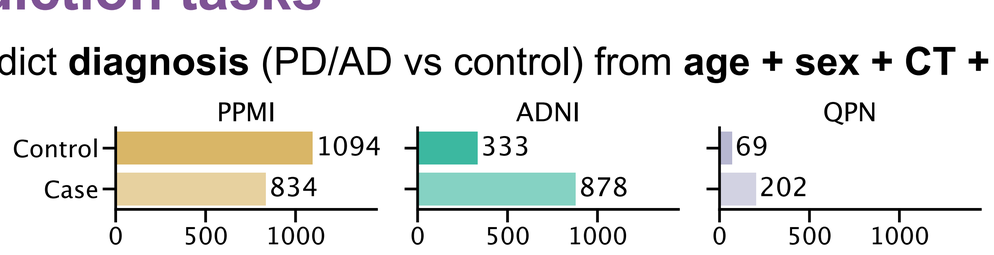
Datasets

We used **demographic** information, structural **neuroimaging** features (cortical thickness [CT] and subcortical volumes [SV]) and **cognitive** assessment scores (Montreal Cognitive Assessment [MoCA] or Mini-Mental State Examination [MMSE]) from three Parkinson's disease (PD) or Alzheimer's disease (AD) datasets:

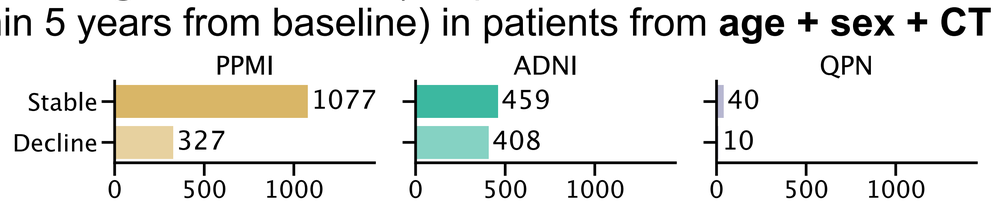
- Parkinson's Progression Markers Initative (**PPMI**) (Marek et al., 2018)
- Alzheimer's Disease Neuroimaging Initative (**ADNI**) (Jack et al., 2008)
- Quebec Parkinson Network (**QPN**) (Gan-Or et al., 2020)

Prediction tasks

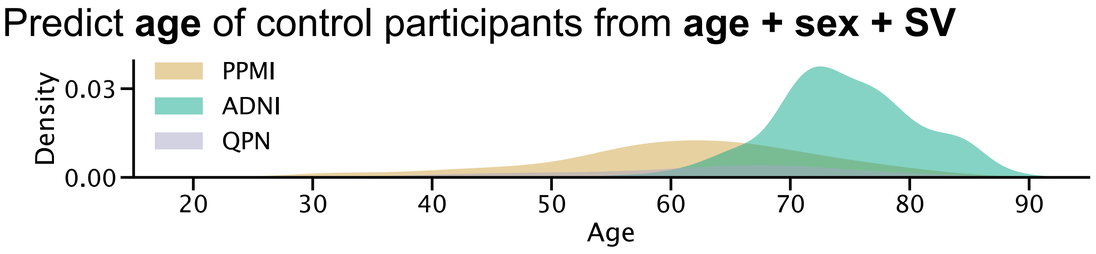
1) Predict **diagnosis** (PD/AD vs control) from **age + sex + CT + SV**



2) Predict **cognitive decline** (≥ 3-point loss in MoCA or MMSE scores within 5 years from baseline) in patients from **age + sex + CT**



3) Predict **age** of control participants from **age + sex + SV**



Model training setups

Model: Z-scoring + logistic/linear regression

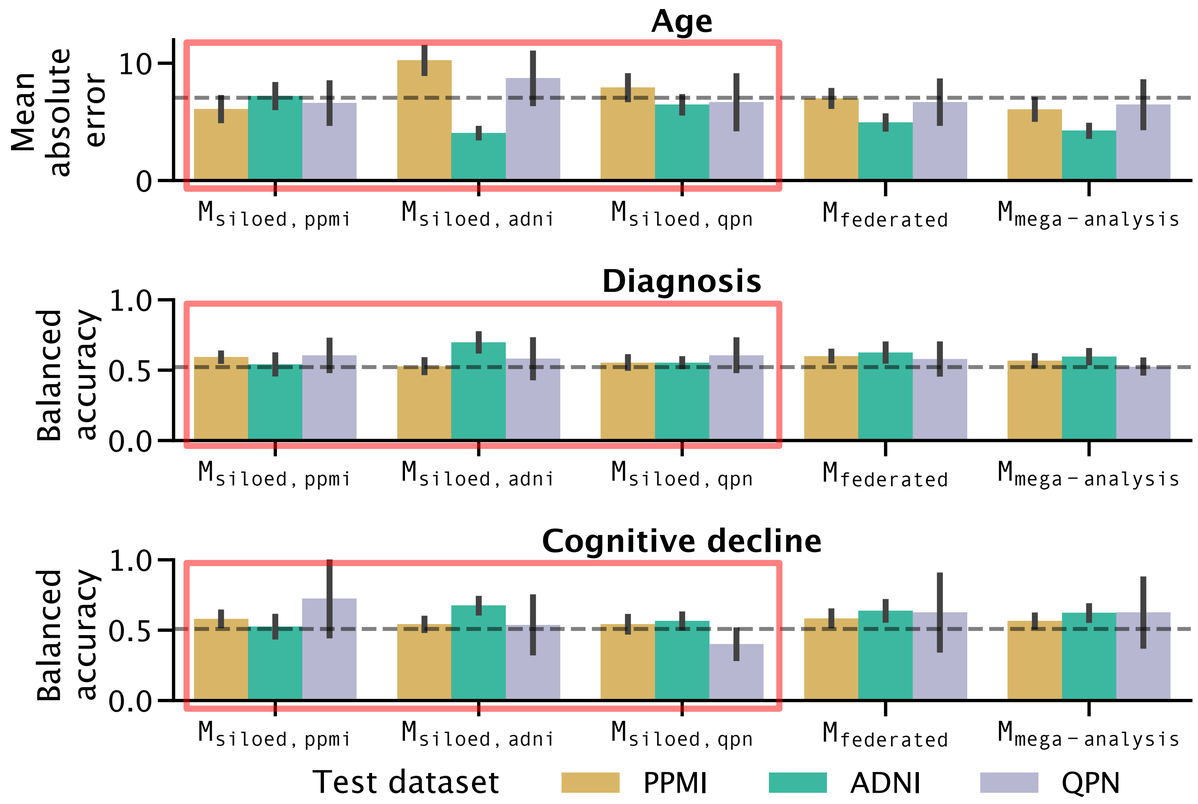
Setup	Train data	Model
Siloed (PPMI)	PPMI	$M_{siloed,\,ppmi}$
Siloed (ADNI)	ADNI	$M_{siloed,\,adni}$
Siloed (QPN)	QPN	$M_{siloed,\,qpn}$
Federated	PPMI	M_{ppmi}
	ADNI	M_{adni}
	QPN	M_{qpn}
	(PPMI, ADNI, QPN)	$M_{federated}$ (weighted avg. of params.)
Mega-analysis	PPMI, ADNI, QPN	$M_{mega-analysis}$

Model performance estimated using **10-fold cross-validation**

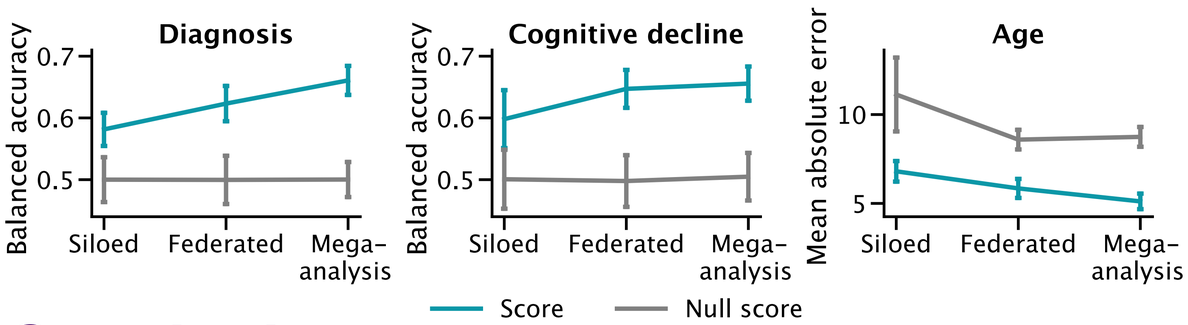
Test datasets: **PPMI**, **ADNI**, **QPN**, **PPMI+ADNI+QPN**

Results

Poor cross-dataset generalizability in Siloed setups (dataset shift)



Test performance scores on **PPMI+ADNI+QPN** show **improvement from Siloed to Federated setup** in all three prediction tasks



Conclusion

Preliminary results suggest that, for the datasets and use-cases investigated, **federated setup shows generalizable and comparable performance to the mega-analysis approach**, opening the way to a change in analysis paradigm for many studies.

Neuroinformatics infrastructure is needed for handling the practical aspects of this type of analysis. **Nipoppy** and **Neurobagel** can help build standardized cohorts for federated analyses.

References

Gan-Or, Z. et al. (2020). The Quebec Parkinson Network: A Researcher-Patient Matching Platform and Multimodal Biorepository. *Journal of Parkinson's Disease*, *10*(1), 301–313.

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.

Marek, K. et al. (2018). The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, *5*(12), 1460–1477.

Marelli, L., & Testa, G. (2018). Scrutinizing the EU General Data Protection Regulation. *Science*, *360*(6388), 496–498.

Acknowledgements

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada
Québec – Fonds de recherche – Nature et technologies; Fonds de recherche – Santé; Fonds de recherche – Société et culture
PARKINSON QUÉBEC
UNIQUE CENTRE
NIH National Institutes of Health
THE MICHAEL J. FOX FOUNDATION FOR PARKINSON'S RESEARCH
Chan Zuckerberg Initiative
HBHL NEUROHUB
QUEBEC PARKINSON NETWORK
Fondation Brain Canada Foundation