

A replication study: Predicting Parkinson's disease trajectory using clinical and functional MRI features (Nguyen et al. 2021)



↑PREPRINT

Germani E.¹, Baghwat N.², Dugré M.³, Gau R.², Montillo A.⁴, Nguyen K.⁴, Sokolowski A.³, Sharp M.², Poline JB.², Glatard T.¹



¹ Univ Rennes, Inria, CNRS, Inserm, France

² Department of Neurology and Neurosurgery, McGill University, Montreal, Canada

³ Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada

⁴ Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, USA



1. Cohort selection

Parkinson's Progression Markers Initiative (PPMI) (Marek & al. 2018)

4 time points: baseline (N=82), 1y (N=53), 2y (N=45) and 4y (N=33).

Baseline selection: fMRI image and MDS-UPDRS score at the same visit, acquisition parameters from the original study (*TR*, *TE*), visit before 2020.

Follow-up selection: Searching for baseline participants that also had a score 1 year, 2 years and 4 years after.

● **Uncertainties about query time of the database**

● **Evolutive database, updated in Sept. 2021**

Selection with criteria → **Replication cohort (A1)**

Authors shared participants & sessions list with us ✓

Missing sessions, different session selected using criteria, etc. → **Closest-to-original cohort (Default)**

3. Features

fALFF (fractional Amplitude at Low Frequency Fluctuation) & **ReHo** (Regional Homogeneity).

Z-scoring and parcellation using 3 atlases.

- Z-scoring step was not found in the code shared by the authors
- ALFF also used by the authors but not reported in the paper
- Homemade Schaefer atlas not available

Reproduction of the Schaefer atlas using publicly available atlases.

Comparison with and without z-scoring (C1) and with ALFF (C2).

2. Image processing pipeline

Homemade pre-processing pipeline using **open source software packages**. Code not available.

- **Missing information on: WM and CSF timeseries estimation, MNI template & denoising method**
- **Non standard pipeline: homemade, multi-software**

Fixed choice for MNI template and denoising.

Implementation of **3 different variations for CSF and WM estimation** using FSL (B1), AFNI (Default) and without using anatomical priors (B2).

4. Machine learning

Nested-cross validation: 10-fold CV inner-loop, Leave-One-Out CV outer-loop.

● **Inconsistencies between code and paper's reporting.**

Results

Code available on Software Heritage

swh:1:snp:ac39cd7495afa754e5d0d298a502cda8684c7eca

With the default reproduction workflow:

- Better than chance performance for all models
- Performance close to the original ones

Multiple variations (A → E), which ones impact the performance?

Baseline: large impact of feature processing choices (C).

At 1 year and after, impact of cohort variations (A) and input features (D - without imaging, without UPDRS)

Small impact of pre-processing pipelines → prediction mainly driven by clinical and demographic features.

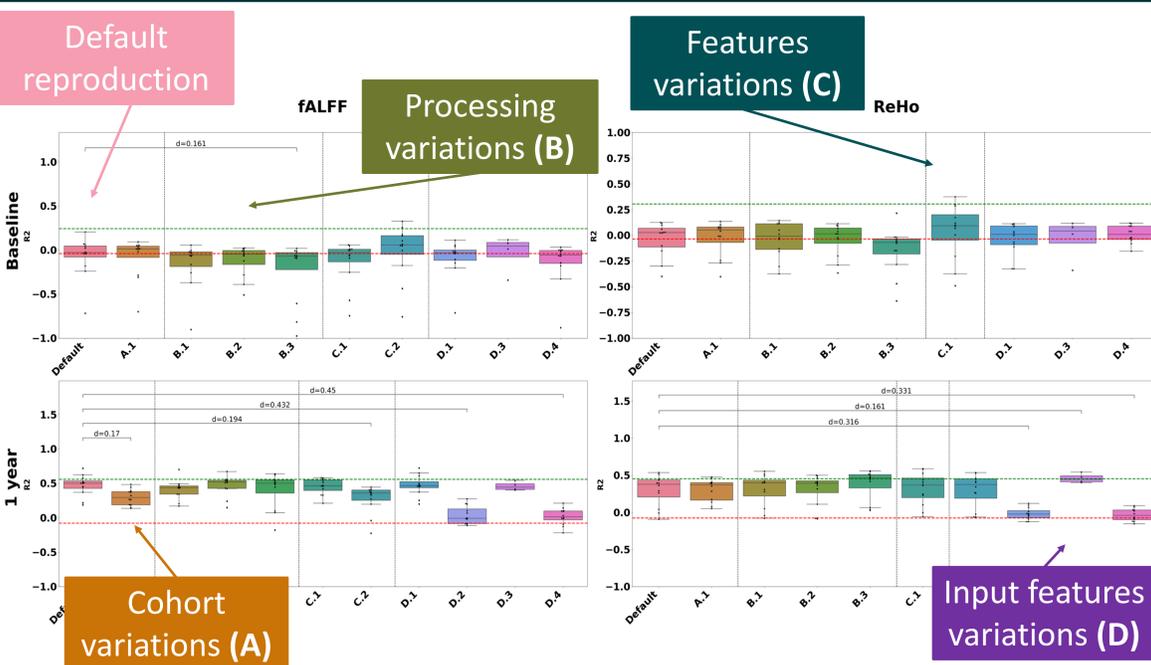


Fig. 1 - Performance of models trained for prediction at each time point, using fALFF or ReHo, with variations in the workflow for prediction at baseline and 1 year. Green line: original performance reported in Nguyen et al. 2021. Red lines: Chance model.

time point	Feature	Type	R2
Baseline	fALFF	Original	0.242
		Default	0.205
		Workflow E.1	0.049
		Workflow E.2	-0.039
1 year	fALFF	Original	0.304
		Default	0.124
		Workflow E.1	-0.164
		Workflow E.2	-0.102

Different choices in machine learning model **optimization** and **performance reporting (E)** → **Large impact**

Fig. 2 - Performance of models with variations in model selection and performance reporting.

Reproducibility associated to clinical data is challenging, in particular for cohort selection and re-use of clinical features.

Code and data sharing is important for reproducibility.

Conclusions

Reproducibility is essential to make trustworthy research.

fMRI studies are highly flexible and a slight change on the workflow can have a large impact on results.

Software package choice and software versions can impact the results.

Machine learning workflows are complex to design, results are often over-estimated due to leakages.

For more info, see the preprint: inserm.hal.science/inserm-04465765

References

Nguyen K., et al. Predicting Parkinson's disease trajectory using clinical and neuroimaging baseline measures. *Parkinsonism & Related Disorders*, 2021.
Marek K., et al. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Annals of clinical and translational neurology*, 2018.

Acknowledgements

This work was funded by the Michael J. Fox Foundation for Parkinson's Research (MJFF-021134). This work was also funded by a MITACS Global Research Award (IT34055). This work was partially funded by Region Bretagne (ARED MAPIS) and Agence Nationale pour la Recherche for the program of doctoral contracts in artificial intelligence (project ANR-20-THIA-0018).