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

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

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
fALFF also used by the authors but not reported in the paper
Hommade Schaefer atlas not available

Reproduction of the Schaefer atlas using publicly available atlases.
Comparison with and without z-scoring (C1) and with ALFF (C2).

4. Machine learning

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

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.

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

Results
Code available on Software Heritage
swb1:anupac3wdf7695575x5/3alp849684670e

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.

References

Acknowledgements
This work was funded by the Michael J. Fox Foundation for Parkinson’s Research (MIFF-011134). This work was also funded by a MITACS Global Research Award (IT14055). 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).