Machine learning applied to higher order functional representations of omics data reveals biological pathways associated with Parkinson's Disease

Elisa Gómez de Lope[1], Enrico Glaab[1] from the NCER-PD consortium

[1] Biomedical Data Science Group, **LCSB, University of Luxembourg**

Background

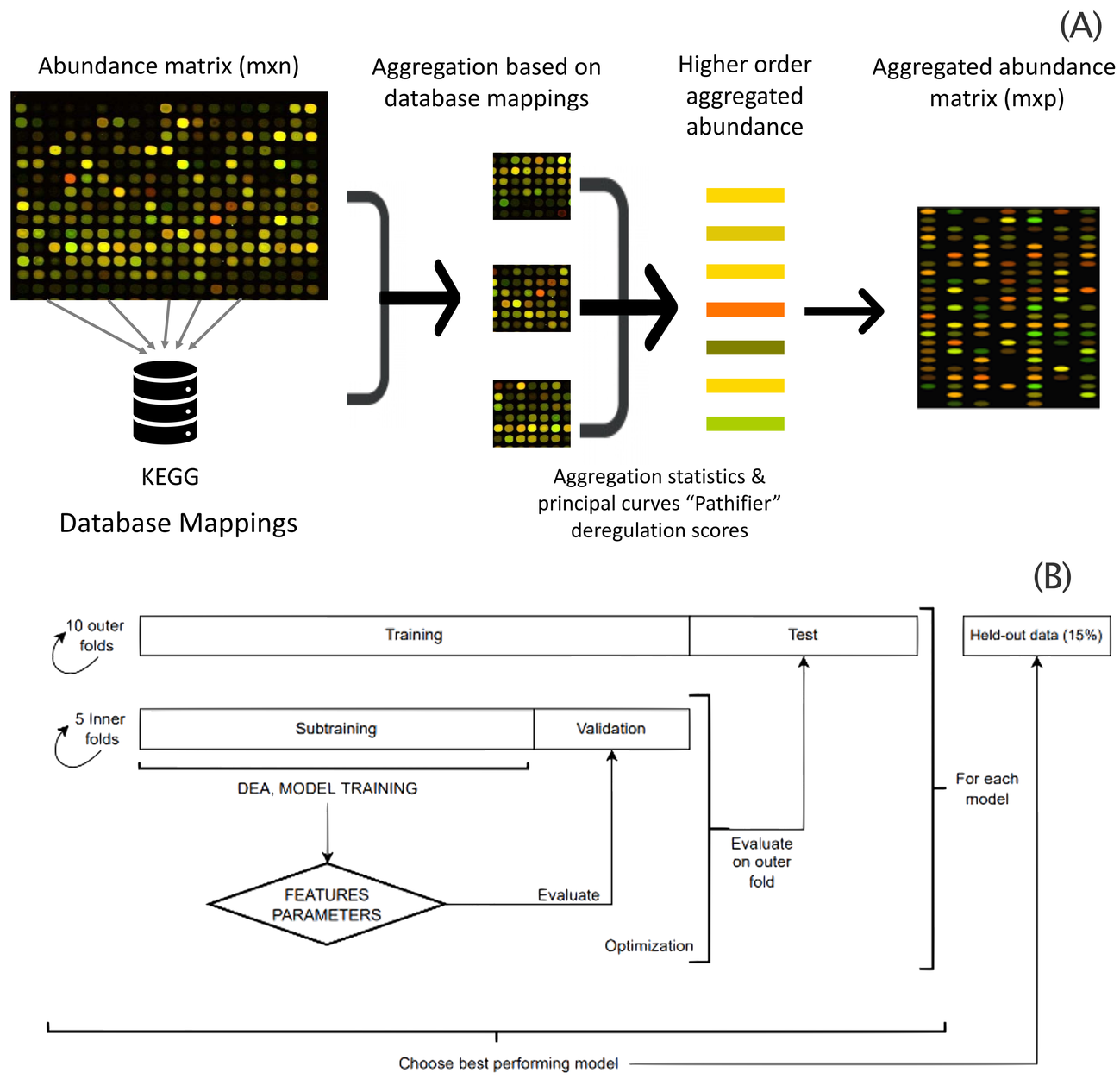
Despite the increasing prevalence of Parkinson's Disease (PD) and research efforts to understand its underlying molecular pathogenesis, early diagnosis of PD remains a challenge.[1]

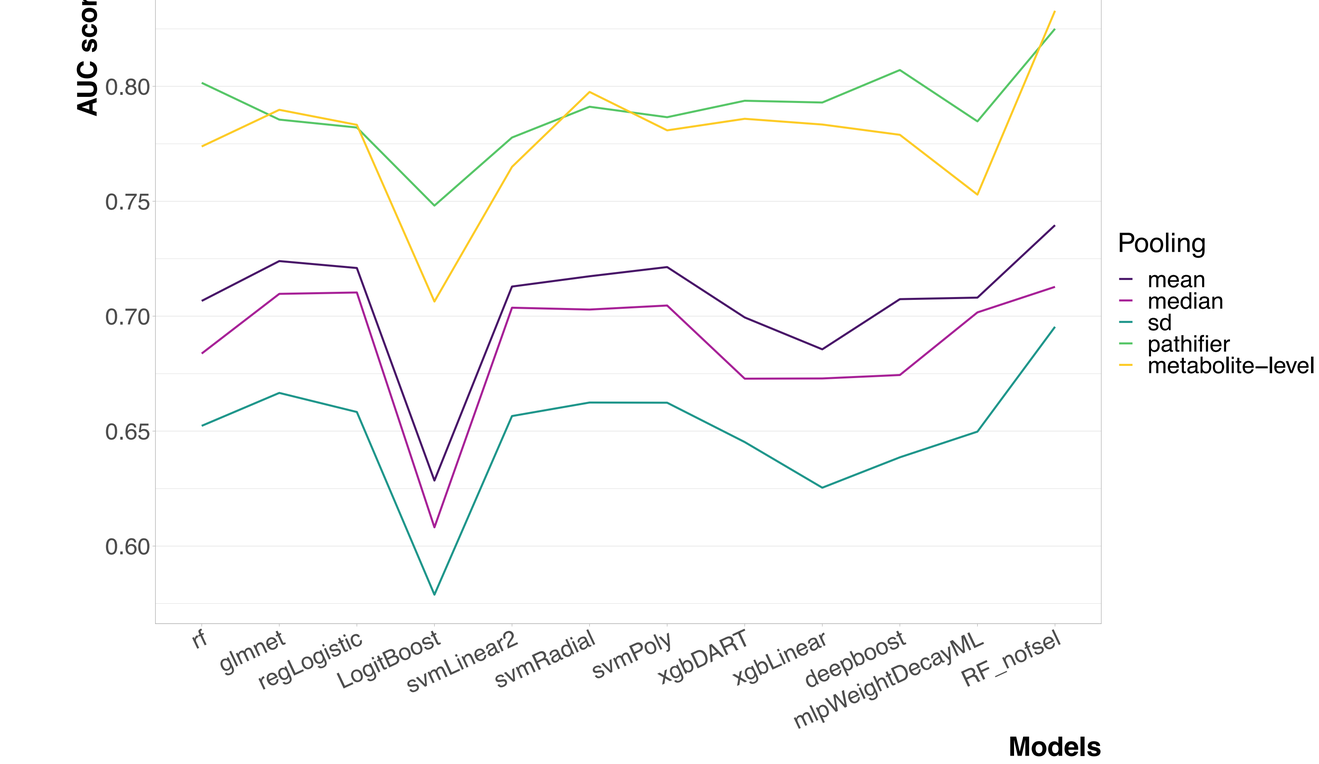
Machine learning analysis of blood-based omics data is a promising non-invasive approach to finding molecular fingerprints associated with PD that may enable an early and accurate diagnosis.[2] However, genes and metabolites don't act isolated. Higher order functional representations of omics data such as pathways allow better and more meaningful interpretation.[3]

Methods

Here, we applied ML classification methods to transcriptomics and metabolomics data from PD case/control studies (PPMI[4] & NCER-PD cohort[5] respectively).

- Higher order functional representations were generated via aggregation statistics (mean, median, sd) and deregulation scores based on principal curves[6] "pathifier scores" (A)
- External two-level cross-validation was used, including nested feature selection (B)
- Models' performance and most relevant predictive features were compared with individual feature level predictors (C)





Conclusions

- Significant AUC scores for cross-validation & external testing.
- Pooled representations of omics data can perform as well as single-level omics predictors to classify PD versus controls samples.
- Plausible biological pathways associated with PD diagnosis.

Limitations:

- Unknown confounders
- Large variability among PD patients makes identifying common trends difficult
- Data represents late stages of the disease

Future work:

- Modelling of other PD prognostic outcomes (e.g. motor dysfunction scores)
- Graph representation of the data via protein-protein interactions, metabolic networks

References

1. Tolosa, E., Garrido, A., Scholz, S. W. & Poewe, W. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol* **20**, 385–397 (2021).

2. Redenšek, S., Dolžan, V. & Kunej, T. From Genomics to Omics Landscapes of Parkinson's Disease: Revealing the Molecular Mechanisms. *OMICS* **22**, 1–16 (2018).

3. Zheng, F., Wei, L., Zhao, L. & Ni, F. Pathway Network Analysis of Complex Diseases Based on Multiple Biological Networks. *Biomed Res Int* **2018**, (2018).

4. Marek, K. et al. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol 95, 629 (2011).

5. Hipp, G. et al. The Luxembourg Parkinson's Study: A Comprehensive Approach for Stratification and Early Diagnosis. Front Aging Neurosci 10, (2018).

6. Drier, Y., Sheffer, M. & Domany, E. Pathway-based personalized analysis of cancer. Proc Natl Acad Sci U S A 110, 6388–6393 (2013).