An Integrative Network Approach for Longitudinal Stratification in Parkinson's Disease

Barry Ryan, Riccardo Marioni & T. Ian Simpson

barry.ryan@ed.ac.uk

Webpage

GitHub

In this research, we perform longitudinal and cross sectional experiments to identify which data modalities or combination of modalities are informative at different time points of Parkinson's Disease. We use clinical, genomic, and proteomic data from the Parkinson's Progression Marker Initiative. We adopt a multi omic approach based on network fusion and graph neural network classification algorithms.

Introduction

Parkinson's Disease (PD) is a neurodegenrative disorder characterised by the loss of dopamine-producing neurons in the brain.

There is a known genetic component of PD and genomic datasets have helped to uncover some aspects of the disease[1].

Understanding the longitudinal variability of PD is essential as it has been theorised that there are different triggers and underlying disease mechanisms at differest disease stages[2].

Parkinson's Progression Markers Iniative (PPMI)[3]

The PPMI was curated to improve this understanding and consists of over 900 PD, 800 Prodromal and 200 Healthy Controls.

Genomic modalities were generated at 4 time points from whole-blood samples at enrollment and yearly intervals thereafter.

PD individuals were recruited less than 2 years post diagnosis and prior to taking any medication.

In this research, we integrated clinical, genomic, and proteomic modalities from the PPMI dataset with the goal of stratifying individuals.

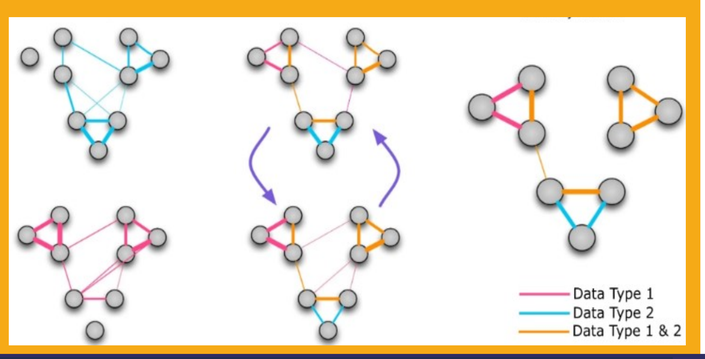
Multi Omic Graph Diagnosis[4]

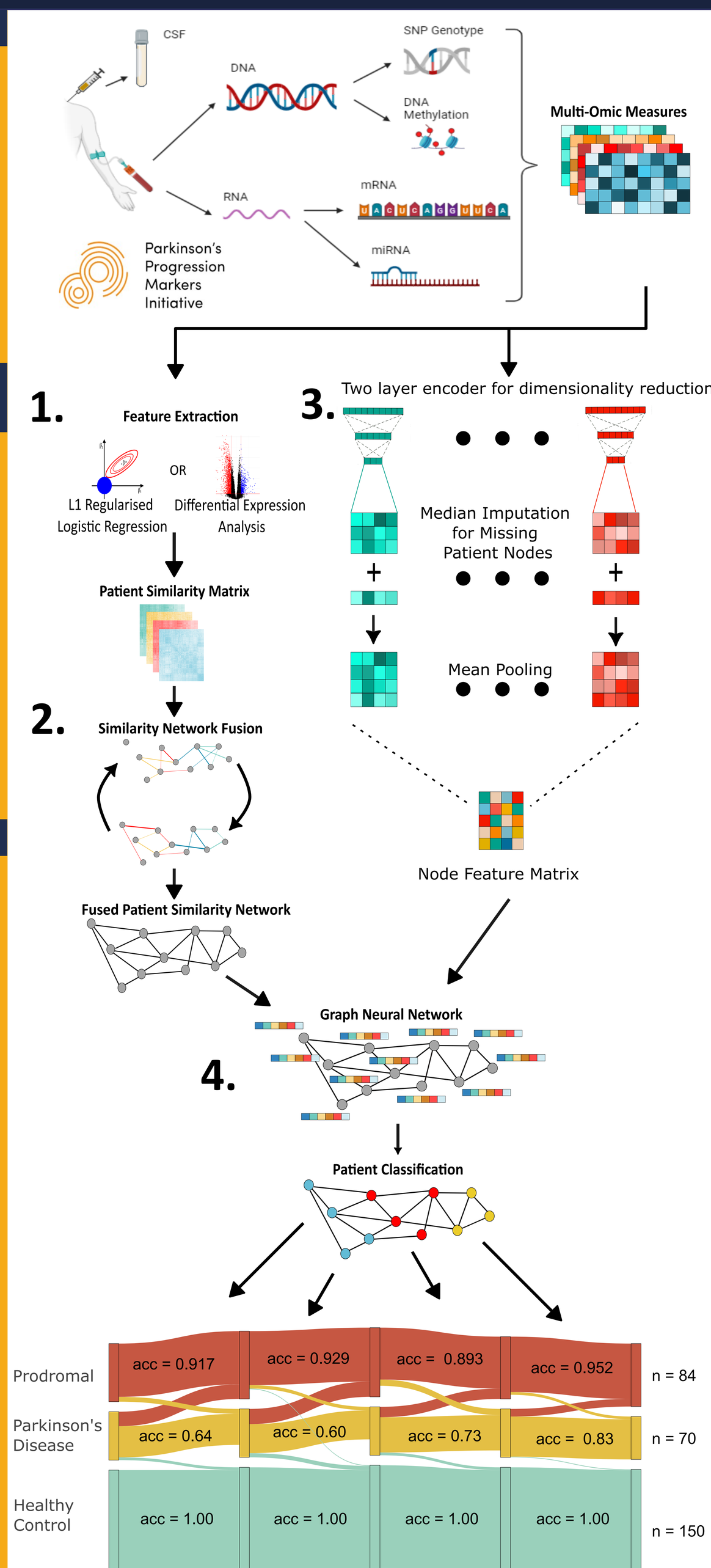
We use a flexible integrative approach based on a network taxonomy to incorporate many aspects of the PPMI dataset, notably the longitudinal component.

Our method, named Multi Omic Graph Diagnosis (MOGDx), is based on 4 key principles/ components.

1. Patient Similarity Network (PSN)
Each modality is represented as a PSN. Similarity is measured by calculating distance between participants' extracted features in each modality. The PSN is generated using the K Nearest Neighbours algorithm.

2. Similarity Network Fusion (SNF)[5]
SNF is used to combine individual modalities PSN's. Missing samples in a modality are retained. Weak relationships are dampened, strong relationships are enforced and new relationships are found through SNF.





Multi Omic Graph Diagnosis[4]

3. Multi Modal Encoder (MME)
Each modality is encoded for dimensionality reduction using an individual. The reduced encoded layer from each modality is decoded to a shared latent space using mean pooling.

4. Graph Convolutional Network (GCN)
GCN's have shown powerful classification on several benchmark network datasets[6]. The GCN implemented consists of two layers with relu activation and batch normalisation

Experiments

Stratification
Individuals were stratified according to thier recruiment status of Healthy Control, Prodromal or PD.

Disease Subtyping
Experiments were performed by grouping individuals into their disease subtypes and jointly. There were two disease subtypes : Genetic - Pariticipants with a mutation in a gene associated with PD and Idiopathic - Participants with no known genetic PD association.

Cross Sectional and Longitudinal
Models were trained and tested cross sectionally at each time point. One model was trained on the genetic subgroup and tested longitudinally at each time point.

Flexible Integration
All combinations of modalities were tested cross sectionally to identify the optimal combination at each disease time point.

Results

An integrative approach is optimal when classifying individauls with PD over time.

Flexibility in integration of modalities facilitates a biological signal for PD to be learnt from whole-blood samples and protein markers in PPMI study participants.

There are possible similarities in the DNAm signatures of idiopathic participants and participants who have a mutation in one of three genes associated with PD.

An integrative model trained using SNP and DNAm at a late disease stage could form a viable early diagnostic tool for stratifying individuals with a mutation in of three genes associated with PD

Conclusion

Accurate longitudinal stratification of PD can be achieved when integrating multiple genomic measures generated from whole-blood samples.

Acknowledgements

This work was supported by the United Kingdom Research and Innovation [grant EP/S02431X/1], UKRI Centre for Doctoral Training in Biomedical AI at the University of Edinburgh, School of Informatics.

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database, RRID:SCR 006431. For up-to-date information on the study, visit www.ppmi-info.org

With thanks to my supervisors Ian Simpson and Riccardo Marioni for their advice and help throughout.

UK Research and Innovation

Biomedical AI
UKRI CENTRE FOR DOCTORAL TRAINING

THE UNIVERSITY of EDINBURGH

References

1. Klein C, Westenberger A. Genetics of Parkinson's disease. Cold Spring Harb Perspect Med. 2012 Jan;2(1):a008888. doi: 10.1101/cshperspect.a008888. PMID: 22315721; PMCID: PMC3253033.
2. Wüllner U, Borghammer P, Choe CU, Csoti I, Falkenburger B, Gasser T, Lingor P, Riederer P. The heterogeneity of Parkinson's disease. J Neural Transm (Vienna). 2023 Jun;130(6):827-838. doi: 10.1007/s00702-023-02635-4. Epub 2023 May 11. PMID: 37169935; PMCID: PMC10174621.
3. Simuni T, Siderowf A, Lasch S, et al. Longitudinal Change of Clinical and Biological Measures in Early Parkinson's Disease: Parkinson's Progression Markers Initiative Cohort. Mov Disord. 2018;33(5):771-782. doi:10.1002/mds.27361
4. Ryan, B., Marioni, R. E., & Simpson, T. I. (2023). Multi-Omic Graph Diagnosis (MOGDx): A data integration tool to perform classification tasks for heterogeneous diseases (p. 2023.07.09.23292410). medRxiv. https://doi.org/10.1101/2023.07.09.23292410
5. Wang, B., Mezlini, A., Demir, F. et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods 11, 333–337 (2014). https://doi.org/10.1038/nmeth.2810
6. Jie Zhou et al. "Graph neural networks: A review of methods and applications". en. In: AI Open 1 (Jan. 2020), pp. 57–81. ISSN: 2666-6510. DOI: 10 . 1016 / j . aiopen . 2021 . 01 . 001. URL: https : / / www .sciencedirect.com/science/article/pii/S2666651021000012 (visited on 10/12/2022).