Cluster Analysis on Longitudinal Data of Parkinson's Disease Subjects

Mohammad R. Salmanpour[1,2], Abdollah Saberi[3], Ghasem Hajianfar[4], Mojtaba Shamsaei[1], Hamid Soltanian-Zadeh[5], Arman Rahmim[2]

1 Amirkabir University of Technology, Tehran, Iran

2 University of British Columbia, Vancouver BC, Canada

3 Islamic Azad University, Tehran, Iran

4 Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Science, Tehran, Iran

5 Henry Ford Health System, Detroit, USA

Qurit LAB — Quantitative Radiomolecular Imaging & Therapy

BC CANCER — UBC

JULY 12–16 2020 VIRTUAL JOINT AAPM | COMP MEETING — EASTERN TIME [GMT-4]

➢ INTRODUCTION

- ✓ Parkinson's disease (PD) is a chronic, heterogenous and progressive neurodegenerative disease and the 2nd most common neurodegenerative disorder after Alzheimer's disease (2–3% of the population >65 years of age) [1,2,3].
- ✓ It is important to identify distinct progression pathways in PD, for improved understanding of disease and improved powering of clinical trials of disease modifying therapies.

➢ AIM

- ✓ To identify distinct disease progression pathways in PD making use of imaging features and advanced time-series (longitudinal) clustering algorithms.

➢ METHODS

- ✓ We studied 885 PD-subjects derived from longitudinal datasets (years 0,1,2,4; Parkinson's Progressive Marker Initiative), with 980 features, e.g. Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) measures, a range of task/exam performances, socioeconomic/family histories and SPECT image features.
- ✓ Segmentation of regions-of-interest (ROIs; caudate and putamen) on DaT SPECT images were performed via MRI images.
- ✓ Radiomic features (RFs) were extracted for each ROI using our standardized SERA software.
- ✓ We first performed unsupervised clustering to identify disease subtypes in a given year (3 clusters robustly identified in another poster, applicable to all years). We then created 2 longitudinal datasets with same patients followed longitudinally.
- ✓ 1st longitudinal dataset included 84 patients which had all features in each year.
- ✓ 2nd dataset consisted of 143 patients (based on year 4) with some missing data in some years that we filled using ensemble hybrid machine-learning (majority-voting) system consisting of 8 feature-selection algorithms (FSAs), 8 dimensionality-reduction algorithms, and 6 classifiers.

➢ METHODS

- ✓ We finally performed longitudinal-clustering of disease progression pathways, using K-Means Longitudinal Clustering (KMLC), an extension of standard K-mean clustering. Ray-Turi clustering evaluation method was used for optimal selection of number of pathway clusters [4,5].

➢ RESULT

- ✓ Our analysis revealed significant heterogeneity in disease projection.
- ✓ We identified 7 distinct progression trajectories/clusters from 1st dataset, confirmed by analysis of 2nd dataset.
- ✓ The pathways included those with consistent disease escalation (5 pathways, 66% of patients), slow progression (red pathway, 24%), and slow improvement (yellow pathways, 10 %).

Fig 1. Fusion of MRI segmentations on SPECT images for radiomics analysis

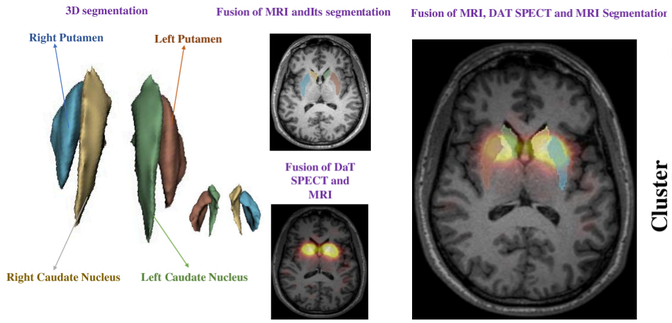


Fig 2. Spider plot of three kinds of features, non-motor (N), motor (M) and imaging (I), for our 3 identified PD sub-clusters.

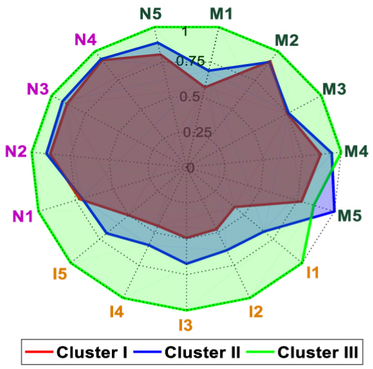


Fig 3. Identified Major Disease Progression Pathways using K Means Longitudinal Clustering (thick lines provides averaged cluster center values for the identified 7 pathways)

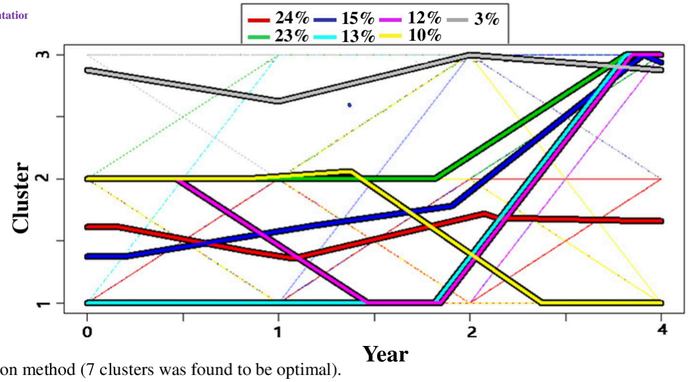
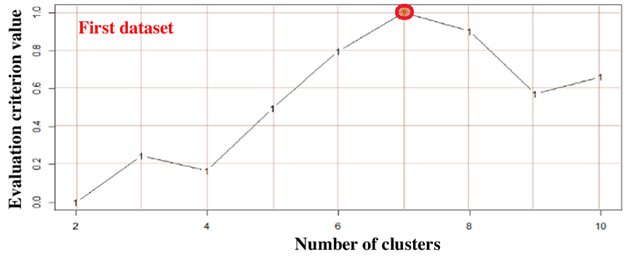
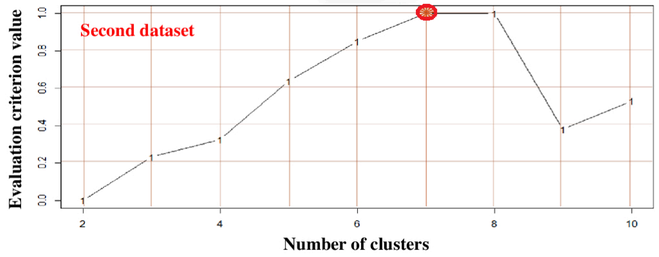


Fig.4. Plots of cluster number evaluation using Ray-Turi clustering evaluation method (7 clusters was found to be optimal).





➢ INNOVATION & IMPACT

- ✓ In this study, measurements are not restricted to individual features or cross-section data only, but includes numerous features studied longitudinally.
- ✓ This works moves beyond cross-sectional PD subtyping to longitudinal disease pathway progression clustering.

➢ CONCLUSIONS

- ✓ Advanced longitudinal missing-data filling and unsupervised-clustering demonstrated 7 distinct longitudinal clusters, depicting significant heterogeneity in PD disease progression.
- ✓ As shown, the majority of the patients (66%) would progress to the most severe subtype (cluster III), while 24% had slow progression and 10% experienced some improvement.
- ✓ As shown, some patients had an improvement in during 5-year following up.

➢ ACKNOWLEDGEMENTS

✓ The project was supported by the Michael J. Fox Foundation, including use of data available from the PPMI—a public-private partnership—funded by The Michael J. Fox Foundation for Parkinson's Research and funding partners (listed at www.ppmi-info.org/fundingpartners). This work was also supported by the Natural Sciences and Engineering Research Council of Canada.

➢ REFERENCES

1. M. Hely, J. Morris and et al, "The sydny multicentre stuly of Parkinson's Disease : progression and mortality at 10 years," J Neurol Neurosurg Psychiarty, vol. 67, pp. 300-3007, 1999.
2. M. Salmanpour, M. Shamsaei and e. al, "Machine learning methods for optimal prediction of motor outcome in Parkinson's disease," Physica Medica, vol. 69, pp. 233-240, 2020.
3. M. Salmanpour, M. Shamsaei and et al, "Optimized machine learning methods for prediction of cognitive outcome in Parkinson's disease," Computers in Biology and Medicine, vol. 111, pp. 1-8, 2019.
4. C. Genolini and B. Falissard, "Kml: A package to cluster longitudinal data," Computer Methods and Programs in Biomedicine, vol. 107, no. 3, pp. e112-e121, 2011.
5. C. Genolini, R. Ecochard and et al, "kmlShape: An Efficient Method to Cluster Longitudinal Data (Time-Series) According to Their Shapes," PLoS One, vol. 11, no. 6, pp. 1-24, 2016.

➢ CONTACT INFORMATION

Arman.rahmim@ubc.ca, msalma@phas.ubc.ca, msalman@bccrc.ca, msalmanpoor66@gmail.com