**LINK TO DATA:** [**https://www.dropbox.com/sh/u2p9oc80iguskwa/AADVV85oo6mQKuk8TMukHKPra?dl=0**](https://www.dropbox.com/sh/u2p9oc80iguskwa/AADVV85oo6mQKuk8TMukHKPra?dl=0)

**Identifying the Clusters within Nonmotor Manifestations in Early Parkinson’s Disease by Using Unsupervised Cluster Analysis (4 pages)**

**Summary:**

The paper focuses on much ignored area related to non motor manifestations in PD and proposes cluster analysis to identify non motor manifestations in early PD.

Classically, patients are divided by their motor phenotype into tremor-dominant (TD) or postural instability-gait disturbance (PIGD) subgroups, or classified by onset age as juvenile, young, or late onset PD. . Several cluster analysis studies have shown that the age of onset rate of disease progression [5–10], and motor phenotype are the major dimensions of PD subgroup classification. The classical and data-driven approaches have focused primarily on the motor symptoms of PD. Some of the nonmotor features, such as depression or cognitive decline were associated with previous classifications based mainly on motor phenotypes.

**Data used:**

Clinical data on current age, age at onset of PD, gender, Hoehn and Yahr (H-Y) stage, dopaminergic drugs in a levodopa-equivalent daily dose (LEDD, mg/day), presence of constipation and responses to the screening questionnaire of the nonmotor features.

Constipation was defined as having fewer than three bowel movements per week.

Patients had to be followed for at least 1 year to be included.

None of the de novo PD patients, a subset of the all PD samples, had previously taken antiparkinsonian medication

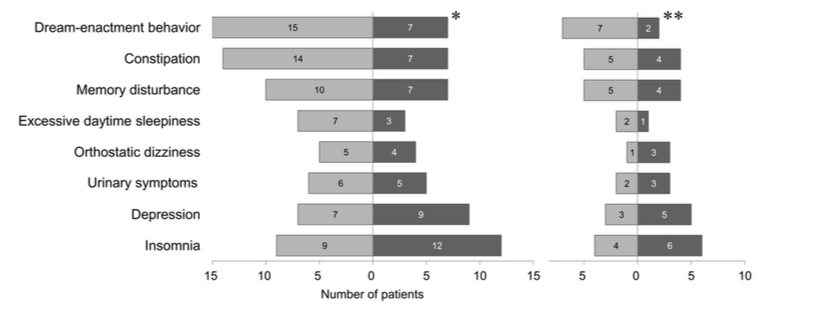
**Method:**

The paper is first of the kind that did analysis on non motor manifestations in PD using a combination of cluster analysis and Korean version of stiffin stick (KVSS). The cluster analysis was performed using unsupervised hierarchical cluster analysis (HCA). The analysis was done on early PD on patients that were with disease onset within 3 years.

Unsupervised HCA was performed and the clustering results are shown using a dendrogram.

The paper used **Yule’s Q** as a measure of similarity for asymmetric binary variables. The dissimilarity between clusters were done using between group average method. Most of the analysis was done in R

**Non motors clustered**



**Conclusion**: The existence of different clusters of nonmotor manifestations in early PD by using unsupervised hierarchical clustering is possible. However the dataset used was small and paper does acknowledge that some of the non motor conditions may be due to other things (non PD related) and as well as other medications. But the study does provide the existence of different nonmotor symptom clusters in early stage PD patients and support the idea of a premotor PD subtype based on nonmotor manifestations. Also Conventional PD classification based on motor phenotype (tremor-dominant or PIDG) cannot be applied to these premotor PD; therefore, classifications that are wholly based on nonmotor symptoms are needed. The finding of this study can have implications for developing the premotor PD classification based on nonmotor features.

**A Data-Driven Approach to the Study of Heterogeneity in Idiopathic Parkinson’s Disease: Identification of Three Distinct Subtypes (11 pages)**

Idiopathic Parkinson’s disease (IPD) has been sub-classified on the basis of predominant motor symptomatology, age at disease onset, depressive affect, and cognitive performance. However, subgroups are usually arbitrarily defined and this paper provides a data driven approach.

**METHODS**

Patients

One hundred seventy-six outpatients with IPD (83 women, 93 men; mean age 63.2 years, standard deviation 10.2 years; mean disease duration 7.5 years, standard deviation 6.4 years)

**MEASURES**

**Motor Function, Mood & Affect, Global Cognitive Function, Demographics**

**ANALYSIS**

All continuous variables were normalized and then subjected to a non-hierarchical cluster analysis (k-means method). Post-hoc comparisons of scores were conducted using either one-way analyses of covariance or independent samples *t* tests with strict application of the **Bonferroni correction**. The **chi-square technique**, or **Fisher’s Exact** when expected values were small, with Yate’s correction where relevant, was used for the comparison of categorical data. Relationships between variables were investigated with **Pearson’s product-moment** correlation corrected for multiple comparisons with the Larzelere and Mulaik test.

The non-hierarchical cluster analysis converged with a 0.01 criterion in nine iterations resulting in five distinct clusters

**Conclusion**

Cluster analysis confirmed the clinical heterogeneity of IPD by identifying three naturally occurring sub- groups at a mean disease duration of 5.6 years (groups 1, 2, and 3) and two at a mean disease duration of 13.4 years (groups 4 and 5). Group 1 was characterized by good motor control without cognitive impairment. Group 2 was characterized by good motor control with cognitive deficits of the ‘‘executive’’ type. Group 3 was characterized by an older age of disease onset, poor motor control with motor complications, and mild global cognitive impairment. Group 4 was characterized by poor motor control without cognitive impairment. Group 5 was characterized by poor motor control with moderately severe global cognitive impairment.

**The Identification of Parkinson’s Disease Subtypes Using Cluster Analysis: A Systematic Review**

The paper reviews the cluster analysis used by different papers and tries to find the similarity and differences in each of the study. The cluster profiles ‘‘old age-at-onset and rapid disease progression’’ and ‘‘young age-at-onset and slow disease progression’’ emerged from the majority of studies. The paper provides explanation that cluster analysis will bring out un identified features whereas most of the other analysis are biased on clinical trials. However in the review it does mention that most of the cluster analysis are not done similarly thereby stating that the results are varying.

How is cluster analysis performed?

Use K Means or HCA

Find Local Optimum

Determine K: Calinski and Harabasz index (pseudo F-statistic)

Cluster Validation: Perform cross validation

Also do discriminant analysis or F Scores

**Variable selection:**

Motor symptoms (UPDRS)

Onset symptom

Cognitive impairment

Depression

Apathy

Hallucinations

Motor complications

Time to MF/dysk, years

Time to falls, years

Disease progression

Disease severity

Disease duration, years

Age at onset, years

Age, years

Medication

ADL

**Data Preprocessing**

Using standardized Z Score

**Recommendation from review**

In spite of the methodological differences, a profile characterized by **higher age-at-onset and faster rate of disease progression** and a profile characterized by **lower age-at-onset and slower rate of disease progression**, emerged from most studies.

1. select a sample of PD patients with a preferably similar disease duration;
2. critically select a set of conceptually similar clinical variables that adequately represent the clinical spectrum of PD and are relevant in discriminating phenotypic profiles;
3. take the limitations of K-means into account or apply another CA technique that does not have limitations;
4. critically evaluate the cluster results: Are they clinically meaningful and interpretable? Which variables discriminate best between clusters? How do the clusters differ with respect to variables that were not included in the CA?;
5. validate the results in independent samples. Studies that apply a similar design in different cohorts and take into account the aforementioned recommendations will likely increase our knowledge on subtypes in PD.

**An RNN Architecture with Dynamic Temporal Matching for**

**Personalized Predictions of Parkinson’s Disease**

**Abstract:**

Goal is to identify more homogeneous sub-populations. Challenge: find similarity among patients because of longitudinal/temporal records. Solution: A RNN which learns patient similarity based on multi-modal and longitudinal data by matching patients based on temporal patterns.

**Intro:**

Features: age of onset, motor impairment, motor severity, cognitive status, sleep disorders, cardiac autonomic dysfunction.

Heterogeneity from AO and progression rate, caused by different underlying biological mechanisms, such as different dopamingeric dysfunction levels. Goal is to enable better prognosis and targeted therapy for homogenous sub-populations.

Methods like local spline regression (LSR) embed vectors of patient events into an intrinsic space and measure Euclidean distance, but this ignores temporality.

Authors tried Word2Vec to train two-layer neural-network, which reflects high contextual similarity. But this too does not explicitly model dynamic temporal data.

Other experiments only use EHR, which doesn’t incorporate data from multiple modalities.

This paper’s contributions: (1) RNN for multi-modal, heterogeneous, longitudinal data, (2) Dynamic Time Warping (DTW) with 2D-RNN to identify similarity between two temporal sequences, (3) extensive testing that improves performance.

**Background:**

RNN feed-forward neural net that computes a fixed sequence of learned non-linear transformations to convert an input pattern into an output pattern. Uses Gated Recurrent Unit (GRU) to overcome the vanishing gradient problem.

Dynamic Time Warping (DTW) measures the similarity between two temporal sequences which may vary in speed; it uses a dynamic programming approach to minimize a distance measure so that two time series are aligned through a warping path. Implemented using a 2D-RNN.

**Related Work:**

Patient Similarity -- mostly ignores temporal relations in data.

Personalized Prediction in Healthcare -- nothing new.

**The RNN Architecture with Dynamic Temporal Matching:**

1. Obtain the distance between two patient data sequences
2. Apply 2D-RNN to compute global distance between two patient data tensors
3. Apply a linear scoring function to obtain final distance

*Similarity between Patient Data Sequence:*

Defines distance metric (Euclidean distance), but this doesn’t incorporate temporality.

*Dynamic Temporal Matching:*

Defines DTW distance dtw(i, j)

*2D-GRU for Dynamic Temporal Matching*

Demonstrates formula and image of GRU.

*Linear Scoring Function:*

Gets distance from 2D-GRU

*Optimization:*

Defines loss function.

**Personalized Prediction:**

1. Retrieve top N similar patients to form sub-population.
2. Train predictive models based on similar patients to make personalized prediction for queried patient.
3. Predict ordered scores that reflect cognitive stages of patients with PD.

**Experiments:**

*Datasets:*

PPMI: extract all features and select features that are observed in at least 400 patients’ records.

Extract 683 patients: 466 cases and 217 control.

Cleaning:

use latest occurrence carry forward strategy. If all are empty for given patient, use mean. Categorical -- one-hot form.

Remove patients with less than 3 sequences.

Loss function is pairwise, creates triplets of (P1, P2+, P2-)

*Features and Targets:*

319 raw features in 7 categories: motor symptoms/complications, cognitive functioning, autonomic symptoms, psychotic symptoms, nighttime sleep problems and excessive daytime sleepiness, depressive symptoms, hospital anxiety and depression scale.

Targets: Hoehn and Yahr (NHY) scale from last 3 months as prediction target, which ranges from 0 to 5. Score 1.0 - PD limited to one side of body. 2.0 - problems affecting both sides. After that, the higher the more severe the condition.

*Evaluation Method:*

Patient Similarity: Precision@K performance measure. See this section for formula.

Personalized Prediction: RMSE for NHY score prediction.

Divide 617 patients into training, validation, and testing ets with 8:1:1 ratio.

*Experiment Results:*

SGD batch size: 10.

Params initialized randomly using uniform distribution.

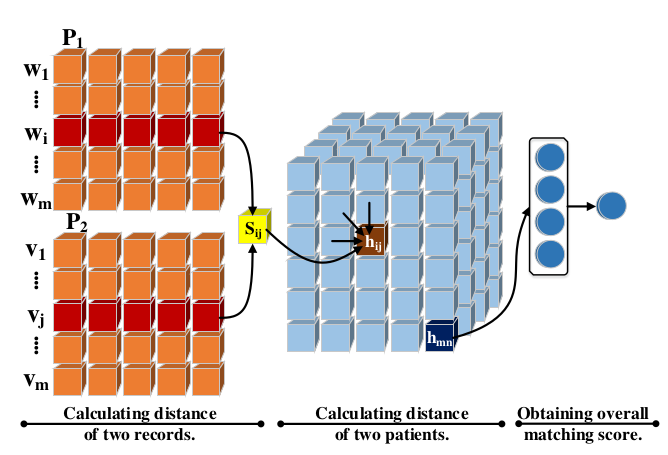
RNN d = 5 after testing 2, 5, 10.

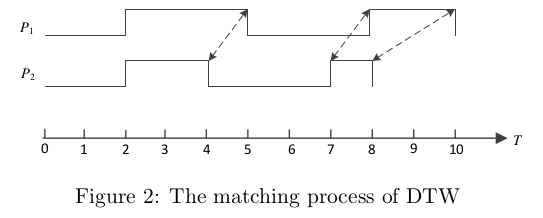
**Discussion:**

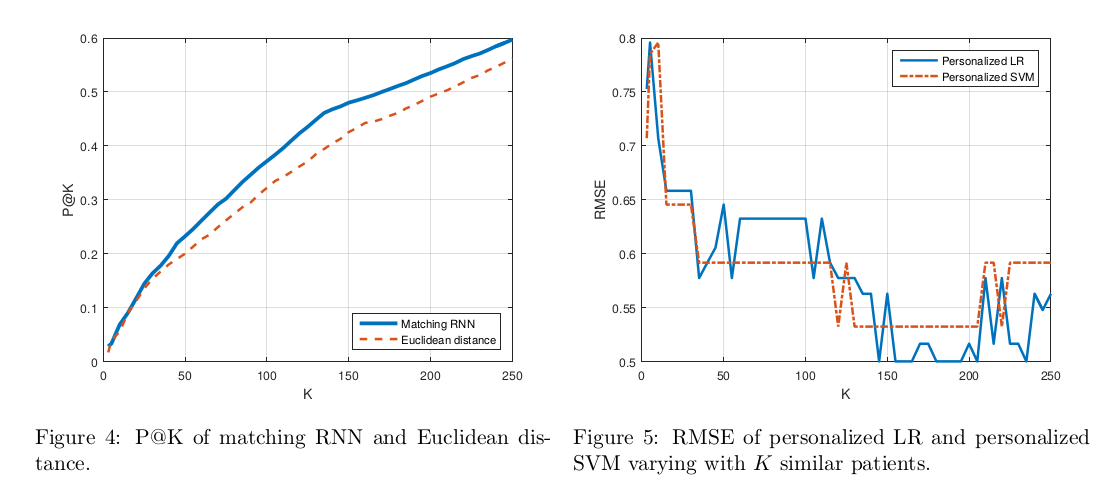
K Increases, RSME decreases.

Personalized prediction outperforms LSTM and KNN.

HNY scores are unbalanced. Most patients in 0 or 2, and few or none in 3, 4, 5. Thus stages 1 and 3 next to stage 0 and 2 prone to misclassification.







Performance:

