2-1-2022

Longitudinal clustering analysis and prediction of Parkinson's disease progression using radiomics and hybrid machine learning

Mohammad R. Salmanpour

Mojtaba Shamsaei

Ghasem Hajianfar

Hamid Soltanian-Zadeh
Henry Ford Health, Hsoltan1@hfhs.org

Arman Rahmim

Recommended Citation

Longitudinal clustering analysis and prediction of Parkinson's disease progression using radiomics and hybrid machine learning

Mohammad R. Salmanpour[1,2], Mojtaba Shamsaei[1], Ghasem Hajianfar[3], Hamid Soltanian-Zadeh[4,5], Arman Rahmim[2,6]

[1]Department of Energy Engineering and Physics, Amirkabir University of Technology, Tehran, Iran; [2]Department of Physics & Astronomy, University of British Columbia, Vancouver BC, Canada; [3]Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Science, Tehran, Iran; [4]CIPCE, School of Electrical & Computer Engineering, University of Tehran, Tehran, Iran; [5]Departments of Radiology and Research Administration, Henry Ford Health System, Detroit, USA; [6]Department of Radiology, University of British Columbia, Vancouver BC, Canada

Contributions: (I) Conception and design: MR Salmanpour, A Rahmim; (II) Administrative support: MR Salmanpour, A Rahmim; (III) Provision of study materials or patients: MR Salmanpour, G Hajianfar; (IV) Collection and assembly of data: MR Salmanpour, G Hajianfar; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Arman Rahmim, PhD. Department of Radiology, University of British Columbia, Vancouver BC, Canada. Email: arman.rahmim@ubc.ca.

Background: We employed machine learning approaches to (I) determine distinct progression trajectories in Parkinson's disease (PD) (unsupervised clustering task), and (II) predict progression trajectories (supervised prediction task), from early (years 0 and 1) data, making use of clinical and imaging features.

Methods: We studied PD-subjects derived from longitudinal datasets (years 0, 1, 2 & 4; Parkinson's Progressive Marker Initiative). We extracted and analyzed 981 features, including motor, non-motor, and radiomics features extracted for each region-of-interest (ROIs: left/right caudate and putamen) using our standardized standardized environment for radiomics analysis (SERA) radiomics software. Segmentation of ROIs on dopamine transposer - single photon emission computed tomography (DAT SPECT) images were performed via magnetic resonance images (MRI). After performing cross-sectional clustering on 885 subjects (original dataset) to identify disease subtypes, we identified optimal longitudinal trajectories using hybrid machine learning systems (HMLS), including principal component analysis (PCA) + K-Means algorithms (KMA) followed by Bayesian information criterion (BIC), Calinski-Harabatz criterion (CHC), and elbow criterion (EC). Subsequently, prediction of the identified trajectories from early year data was performed using multiple HMLSs including 16 Dimension Reduction Algorithms (DRA) and 10 classification algorithms.

Results: We identified 3 distinct progression trajectories. Hotelling's t squared test (HTST) showed that the identified trajectories were distinct. The trajectories included those with (I, II) disease escalation (2 trajectories, 27% and 38% of patients) and (III) stable disease (1 trajectory, 35% of patients). For trajectory prediction from early year data, HMLSs including the stochastic neighbor embedding algorithm (SNEA, as a DRA) as well as locally linear embedding algorithm (LLEA, as a DRA), linked with the new probabilistic neural network classifier (NPNNC, as a classifier), resulted in accuracies of 78.4% and 79.2% respectively, while other HMLSs such as SNEA + Lib_SVM (library for support vector machines) and t_SNE (t-distributed stochastic neighbor embedding) + NPNNC resulted in 76.5% and 76.1% respectively.

Conclusions: This study moves beyond cross-sectional PD subtyping to clustering of longitudinal disease trajectories. We conclude that combining medical information with SPECT-based radiomics features, and optimal utilization of HMLSs, can identify distinct disease trajectories in PD patients, and enable effective prediction of disease trajectories from early year data.

Introduction

Parkinson's disease (PD) as a progressive (1-3) and heterogeneous disease (4-7) is the second-most common neurodegenerative disorder after Alzheimer's disease (8,9). PD is characterized by motor (10,11) and non-motor symptoms (12-15). Even though there is currently no proven permanent therapies for PD, but symptomatic treatments with levodopa (16) and dopaminergic agonists (17) are considered in order to temporally control symptoms (18,19). A multicenter study (20) showed different progression rate in PD for 10 years follow-up. Identification of PD progression could enhance comprehension of PD mechanisms as well as improve design of clinical trials (21). While studies focused on identification of PD subtypes (6,22-29), identification of longitudinal progression in the course of several years can led to better understanding of underlying mechanisms (30-32).

Studies for prediction of individual outcomes in PD patients have recently been emerging (33-43). Discovery of new biomarkers of PD may enable improved treatment planning (24). Although motor dysfunctionalities were first proven as primary indicators of disease progression at screening (44), non-motor symptoms were shown to be significant considerations (45). Our past efforts (27-29,42,46) focusing on identifications of PD subtypes as well as prediction of motor and subtype outcome in PD showed that radiomics features, beyond conventional imaging measures, enable improvement in both tasks. Moreover, using HMLSs enabled significant improvements in both clustering and prediction tasks (38,42,43). In some studies, no significant correlations between conventional imaging features and clinical features (46) and no improvement in prediction when employing their mixture (42) were observed.

A recent study (30) focused on specifying progression rate of PD using some relevant features. The authors first clustered PD patients by global composite outcome score, and then compared the progression of this score among different subtypes in 4.5-year follow-up. Another effort (31)

focused on specifying the heterogeneity of PD (subtypes) and then investigated the progression of some motor and non-motor features among sub-clusters in 6-year follow-up. Patient subtyping can be defined as a clustering problem (47), where the subjects within a cluster are similar to each other. Another longitudinal study (48) investigated the annual progression rate in activity and participation measures in 5-year follow-up. A study (49) also focused on clinical progression of postural instability gait disorders compared to tremor dominant using longitudinal clinical data in 4-year follow-up. This study employed the linear mixed-effects models to specify differences in progression rate between the two groups. Furthermore, a study explored the main clinical variables of PD progression in a selected population of subjects, identified by a long-term preserved response to dopaminergic treatment in 30-year follow up. They assessed the clinical and neuropsychological progression of 19 patients, treated with subthalamic nucleus deep brain stimulation.

Han *et al*. (50) applied unsupervised medical anomaly detection generative adversarial network (deep learning technique) on multi-sequence structural MRI to detect brain anomalies (Alzheimer disease) at a very early stage. In another study, Nakao *et al*. (51) employed an unsupervised anomaly detection method based on the variational autoencoder- generative adversarial network (VAE-GAN, as a deep learning technique) to detect various lesions using a large chest radiograph dataset. Rundo *et al*. (52) employed a fully automatic method for necrosis extraction, after the whole gross tumor volume (GTV) segmentation, using the Fuzzy C-Means (FCM) algorithm to detect the necrotic regions within the planned GTV for neuro-radiosurgery therapy. In another study (53), particle swarm optimization algorithm was employed to improve FCM algorithm via selecting the initial cluster centers optimally.

To the best of our knowledge, there has been very limited work focused on longitudinal clustering to stratify PD progression trajectories. Since patients in the same cross-sectional subtype can have similar properties but with different progression rates, clustering longitudinal
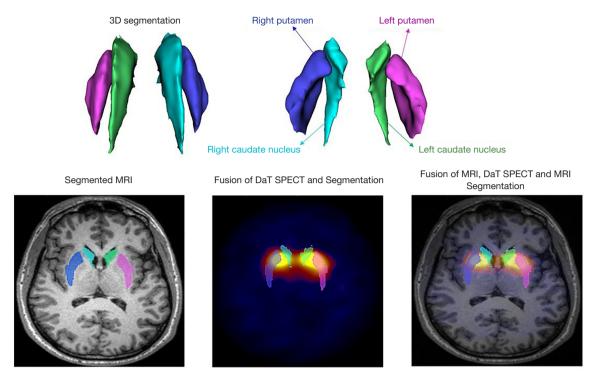
Figure 1 Example 3D MRI segmentation (top) and SPECT-MRI fusion (bottom) from our image processing pipeline.

trajectories constructed using PD subtypes can improve interpretation of disease progression instead of following specific features among PD subtypes over time (32). Our present effort includes derivation of distinct trajectories of PD progression during 4-year follow-up (unsupervised task) via longitudinal clustering, as well as prediction of these trajectories (supervised task) from early year data using HMLSs. Our methods and materials are described next, followed by results, discussion, and conclusion.

We present the following article in accordance with the MDAR checklist (available at https://dx.doi.org/10.21037/qims-21-425).

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). We extracted patient data from the PPMI database (www.ppmi-info.org/data) including clinical information and DAT SPECT and MRI, for years 0 (baseline), 1, 2 and 4. As detailed next, image processing pipeline was implemented to perform MRI segmentation, then DAT SPECT/MRI fusion, and feature extraction from each segmented ROI were done.

We then employed HMLSs, consisting of DRAs combined with clustering/classification algorithms, in order to identify the most optimal progression trajectories in PD and to predict these trajectories using data in years 0 and 1.

Image segmentation, fusion, and feature extraction

As shown in Figure S1 and elaborated in Appendix 1, we employed multiple steps towards extraction of radiomics feature from ROIs in dorsal striatum (DS) (left and right caudate and putamen). The FreeSurfer package was first utilized to directly segment T1 MRI (29), and then register DAT SPECT images to MRI. Example images are shown in *Figure 1*. We provided the ROIs to our standardized SERA software package (54) to extract radiomics features.

Patient features

We first select 885 patients (original dataset) who had 981 features: the features included motor, non-motor features and radiomics features extracted for each ROI using our standardized SERA software (these are further elaborated in Appendix 1 and the features were publicly shared as

linked in "Data and Code Availability"). Since some information of some patients were missing in some years, for longitudinal clustering task, we arrived at dataset of 143 patients (years 0, 1, 2 and 4). For the prediction task based on year 0 and/or 1, we created 3 datasets: (I) CSD0 with 143 subjects: the use of cross-sectional dataset in year 0 as input and the related trajectories as outcome; (II) TD01 with 286 subjects: the use of datasets in year 0 and 1 as timeless approach as input and the related trajectories as outcome (i.e., effectively doubling the number of our cases; hypothesized to improve task due to added statistics); (III) CSD01 with 143 subjects: the use of cross-sectional datasets in year 0, 1 (putting cross-sectional datasets longitudinally next to each other) as input and related trajectories as outcome. We constructed the timeless dataset (TD01) by appending cross sectional datasets within a single set of data. This approach enables us to gather data with larger (double) number of patients.

Machine learning methods

We utilized HMLSs, utilizing 3 groups of algorithms including: (I) feature extraction algorithms, (II) clustering algorithms, and (III) classification algorithms. These are elaborated next.

Feature extraction algorithms (FEAs)

We employed FEAs to reduce high-dimensional data into fewer dimensions (i.e., as DRAs), in order to tackle overfitting issues in both unsupervised and supervised tasks. Unsupervised FEAs use no label for attribute extraction, and rely on patterns emerged between input features (55-57). In this study, 16 FEAs (only unsupervised) were employed (as elaborated in Appendix 1): (I) principle component analysis (PCA) (58); (II) Kernel PCA (59); (III) t-SNE (60); (IV) factor analysis (FA) (61); (V) Sammon Mapping algorithm (SMA) (62,63); (IV) Isomap algorithm (IsoA) (64); (VII) LandMark Isomap algorithm (LMIsoA) (65); (VIII) laplacian eigenmaps algorithm (LEA) (66,67); (IX) LLEA (68); (X) multidimensional scaling algorithm (MDSA) (69); (XI) diffusion map algorithm (DMA) (70,71); (XII) stochastic proximity embedding algorithm (SPEA) (72); (XIII) gaussian process latent variable model (GPLVM) (73,74); (XIV) SNEA (75); (XV) symmetric stochastic neighbor embedding algorithm (Sym_SNEA) (76); and (XVI) autoencoders algorithms (AA) (77). These methods were implemented in R 2020 respectively.

Clustering algorithms

For the unsupervised longitudinal clustering tasks, we employ clustering algorithm to group multidimensional data based on similarity measures (78-80). In our recent study involving cross-sectional and timeless data (27,28) (but not performing longitudinal clustering, which the present work pursues), the KMA (as elaborated in Appendix 1) outperformed various clustering algorithms. In the present work, we also study KMA as applied to longitudinal clustering. KMA was implemented in MATLAB R 2020 b platform.
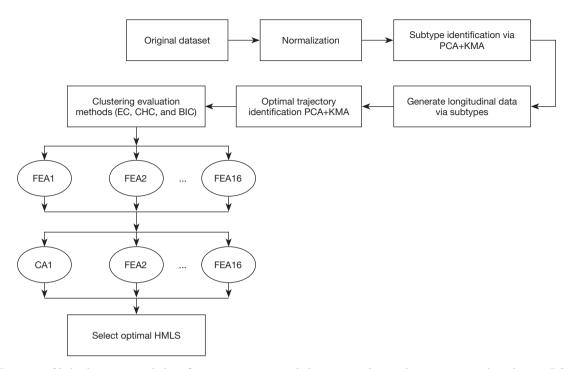
Clustering evaluation methods (CEM): To evaluate clustering trajectories, the so-called EC (81), CHC (82) and BIC (83) were utilized (Appendix 1), selecting optimal number of clusters from a range of cluster numbers spanning 2 to 9.

Classification algorithms (CAs)

For the supervised prediction task, we utilize a range of optimal CAs selected between various families of learner algorithms. These are all listed in Appendix 1). Specifically, we selected 9 CAs: (I) decision tree classification (DTC) (84-86); (II) Lib_SVM (87-89); (III) K nearest neighborhood classifier (KNNC) (90,91); (IV) ensemble leaner classifier (ELC) (92,93); (V) linear discriminant analysis classifier (LDAC) (94,95); (VI) NPNNC (96,97); (VII) error-correcting output codes model classifier (ECOCMC) (98,99); (VIII) multilayer perceptron_back propagation classifier (MLP_BPC) (100,101); (IX) random forest classifier (RFC) (102,103); and (X) recurrent neural network classifier (RNNC) (20,104). We performed 5-fold cross validation for all CAs and automatically adjusted intrinsic hyperparameters through automated machine learning hyperparameter tuning (as elaborated in Appendix 1) (38). We implemented all CAs in MATLAB R 2020b.

Analysis procedure

As shown in *Figure 2*, our work includes two main stages: (I) optimal identification of progression trajectories in PD (unsupervised task); and (II) usage of HMLSs to improve prediction performance (supervised task). For the first stage (unsupervised clustering), we first utilized a normalized original dataset (every feature was normalized based on minimum and maximum values for that feature) in order to cluster PD subjects in each year ("original subtypes"). We reproduced 3 sub-clusters as elaborated in our prior work (29),

Figure 2 Diagram of hybrid systems, including first stage unsupervised clustering and second stage supervised prediction. PCA, principle component analysis; KMA, KMeans algorithm; EC, elbow criterion; CHC, Calinski-Harabatz criterion; BIC, Bayesian information criterion; FEA, feature extraction algorithms; CA, classification algorithms.

namely (I) mild, (II) intermediate, and (III) severe. Our prior work (29) showed that these clusters were consistent year to year (i.e., while patients of course may move from cluster to cluster in different years, the 3 clusters identified in each year are consistent with one another. Subsequently, a HMLS including PCA and KMA was applied to the longitudinal data as clustered in each year to a particular subtype. We then generated 4 components, and we used the first 2 components for 2-dimension visualization while we used all components to identify progression of trajectories. To optimize the number of longitudinal trajectories (clusters), we applied BIC and CHC to our results (for a range of 2–9 longitudinal clusters/trajectories) as generated by HMLS. Our optimized number of trajectories were further confirmed by EC as applied on clustering results provided by PCA + KMA. We also applied a statistical test, HTST, described in Appendix 1 (105) to measure similarities between sub-clusters. Subsequently, in stage two of our efforts (supervised task), prediction of the identified trajectories based on early years (data in year 0 and 1) was performed using multiple HMLSs, including 16 FEAs coupled to 10 CAs, as enlisted previously.

Data and code availability

All code (included prediction algorithms and feature extraction algorithms etc.) and all datasets are publicly shared at: https://github.com/MohammadRSalmanpour/Longitudinal-task

Results

First stage analysis for optimal identification of trajectories in PD

In a previous study of ours (29), involving analysis of cross-sectional data (not longitudinal clustering), HMLS including PCA + KMA enabled us to find 3 distinct subtypes consisting of: (I) mild, (II) intermediate and (III) severe. *Figure 3* shows these distinct subtypes identified. We utilize these subtypes in our longitudinal data, to first cluster data from each year, followed by analysis of progression and prediction.

In this effort, we studied progression trajectories via following PD subtypes during year 0, 1, 2, and 4. In fact, the features define the subtypes and these subtypes change over
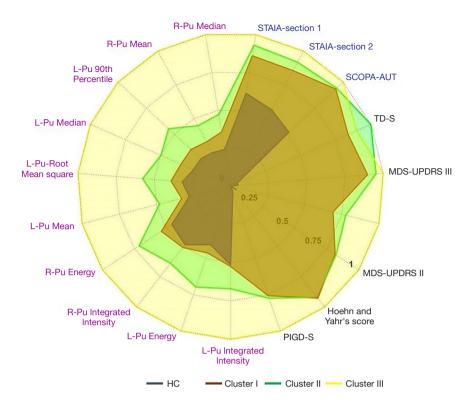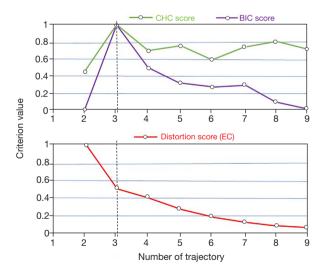
Figure 3 Spider plot of 3 types of features, motor symptoms (black text), imaging features (purple text) and non-motor symptoms (blue text) for 3 identified PD sub-clusters: (I) mild (light brown area), (II) intermediate (green area), and (III) severe (yellow area); and healthy control groups (HC, dark brown area).
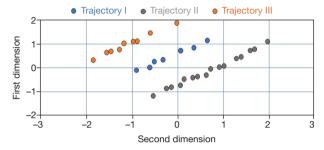

Figure 4 Plots of cluster number evaluation. (Top) Criterion values from CHC and BIC methods, and (bottom) as provided by EC method. The highest criterion scores resulting from CHC and BIC both indicate 3 optimal longitudinal trajectories, as also confirmed by EC method. CHC, Calinski-Harabatz criterion; BIC, Bayesian information criterion; EC, elbow criterion.

time as features change. In fact, we used all imaging and non-imaging features as well as biologically defining PD subtypes in previous paper (29). Our important discovery was, by looking at timeless data as well as cross-sectional data, that disease in any year could be clustered to 1 of 3 distinct clusters; i.e., these clusters were significantly the same from year to year; in other words, while a patient can surely move between different clusters in different years, the 3 specific clusters themselves from year to year are very consistent. Thus, subtypes are means of representing the collective features, and changes of these subtypes over time display changes of these features over time. Next, employing 3 CEMs linked with PCA + KMA enabled us to find optimal number of trajectories. As shown in *Figure 4* (top), the highest scores provided by CHC and BIC both belong to 3 optimal longitudinal trajectories, and results (bottom) provided by EC method confirms the finding.
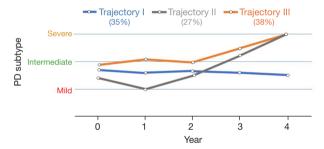
As shown in *Table 1*, P values calculated by HTST show significant difference between sub-clusters. The 3 P values were adjusted via Bonferroni correction test. In short, this table demonstrates that our sub-clusters are quite distinct.

912

Salmanpour et al. Longitudinal clustering and prediction of PD progression

Table 1 P values provided by HTST between different sub-clusters

Subgroup (SG)	SG1	SG2	SG3
SG1	1	<0.001	<0.001
SG2	<0.001	1	<0.001
SG3	<0.001	<0.001	1

Figure 5 PD progression subtypes provided via PCA + KMA. Y and X axes indicate the first and second dimensions provided by PCA.

Figure 6 Identified major progression trajectories, including stable disease (trajectory I: 35% of patients), and disease escalation (trajectory II: 27% and trajectory III: 38% of patients).

After specifying optimal number of trajectories, we applied the dataset with PD subtypes to an HMLS including PCA + KMA. *Figure 5* shows 3 distinct trajectory sub-clusters. The Y and X axes denote the first 2 principal components obtained by PCA, and each sub-cluster is shown using different colors. There are clear separations between the sub-clusters. In fact, we employed this plot as an independent visualization test in our effort to derive PD progression subtypes.

Figure 6 shows three distinct trajectories. Each point on the trajectories was calculated via averaging patient subtype values in specific year. PD patients in trajectory I

(age range in year 69.4±9.5, female: 19, male: 31) showed slower progression in 4-year follow-up compared to other trajectories. In trajectory II (age range in year 67±10, female: 10, male: 28), patients show an improvement of the disease in the first 2 years and then enhanced progression in during years 2–4. Trajectory III (age range in year 68.2±11.5, female: 19, male: 36) illustrated significantly enhanced progression of PD in 4-year follow-up.

Second stage analysis including prediction of trajectories using data in year 0 and 1

After generating new dataset using data in year 0 and 1 as input and utilizing the identified trajectories as outcome (as elaborated in section II, part B), we applied the dataset to multiple HMLSs. The HMLSs consisted of FEAs and CAs (as elaborated in section II, part D). The performances are shown in *Figure 7*.

As shown in *Figure 7*, the hybrid HMLSs resulted in different performances for the TD01 dataset. The performances were averaged among 5 folds cross-validations tests and reported. Furthermore, their standard deviations were also calculated and reported. The best results were observed when SNEA as well as LLEA, were followed with NPNNC, resulting in accuracies of 78.4%±6.9% and 79.2%±3.6% respectively. Some other HMLSs such as SNEA linked with Lib_SVM and t_SNE linked with NPNNC resulted in 76.5%±8.9% and 76.1%±6.1% respectively. Other HMLSs such as Kernel PCA followed by RNNC, t-SNE followed by Lib-SVM, and SMA followed by NPNNC also resulted in accuracies over 70%. Meanwhile, most HMLSs including PCA + RNNC, FA + Lib-SVM, Kernel PCA + ECOCMC, DMA + KNNC, SNEA + RFA, AA + MLP-BP, etc. resulted in accuracies over 60%. Meanwhile, no improvement was observed when we used CSD0 nor CSD01 as input and the related trajectories as outcome. We achieved the maximum performance 53.6%±6.3% and 66.4%±7.6% obtained by CSD0 and CSD01 respectively (P<0.05; *t*-test). Moreover, we reached a maximum performance around 51%±3.3% (P<0.05; *t*-test) when we used sole classifiers. Employing classifiers followed by DRAs (HMLSs) has often proven to increase performance (38,43,106). In our previous study (42,106), we aimed to predict cross-sectional PD subtypes in year 4 by employing 3 kinds of dataset (such as CSD0, CSD01 and TD01) linked with different HMLSs. We also showed that the usage of TD01 only enabled us to achieve high accuracies while the usage of the other datasets such as

Figure 7 Performance heat map plot for application of the 16 FEAs linked with 10 CAs to TD01 dataset. The cells display average of accuracies (± standard deviation) derived from 5-fold cross validation. X-axis lists the CAs and the Y-axis shows FSAs.

CSD0 or CSD01 resulted poorer performances.

Discussion

Identification of PD subtypes can be defined as a clustering task so that patients within the same cluster would have more similar properties than the patients within other clusters (47), but these patients may have different progression speeds in the future (21). Longitudinal clustering enables us to identify progression of the disease in several-year follow-up (32). Clustering longitudinal trajectories constructing from the subtypes can improve the interpretation of the disease progression instead of following some specific features during several years (32) because these subtypes depict an outline of all features (29). A recent study (20) showed that different patients had different progression rates in 10-year follow-up despite the employment of various targeted treatment strategies so that 9 of 126 patients were diagnosed to confine to bed or a wheelchair unless aided, whereas 13 patients showed no significant functional restrictions. In short, PD is defined as a heterogeneous disease (4,5,7) and longitudinal clustering can help provide an easier way for the interpretation of the disease progression.

Even though prediction of PD progression was recently expressed as an important and challenging problem (107),

a number of benefits are obtained by accurate prediction of the outcome in PD (108), including the improvement of the treatment methods, better comprehension of the disease progression, and the improvement in clinical interpretation for symptomatic therapy. This is particularly relevant as PD progression is heterogeneous; such variability encounters us with the challenge of predicting progression in PD patients (108). Since, large input dimensionality and small sample sizes are often common problems in research studies, dimensionality reduction algorithms have been used to deal with these problems (109). In this work, we aimed to identify optimal patterns (trajectories) of PD progression as well as predict these trajectories using data in years 0 and 1. In the first part, 3 optimal trajectories were identified by three CEMs. In the second part, we were able to accurately predict the trajectories in 4-year follow-up by HMLS that included LLEA and SNEA linked with NPNNC.

In the first effort, we analyzed the original dataset using HMLS including PCA + KMA to identify optimal PD subtypes, as described in our previous work (29). Subsequently, we utilized longitudinal datasets, where patients for each given year were associated with a subtype; the resulting longitudinal tracks were then provided to HMLSs including EC, CHC and BIC linked with KMA + PCA, for optimal longitudinal clustering. As depicted in *Figure 4*, the analysis of data resulted in 3 optimal

914

Salmanpour et al. Longitudinal clustering and prediction of PD progression

trajectories. *Table 1* showed that there is no similarity between the cluster, and they are quite distinct. *Figure 5* shows these 3 distinct clusters; we employed this plot as an independent visualization test of our study to specify optimal PD trajectories. The plots show clear distinctions among the clusters. *Figure 6* obviously depicts progression of 3 distinct trajectories in 4-year follow-up. Trajectory I showed a relatively smooth line during the 4-year follow-up. Thus, the patients in this trajectory experienced a slower progression compared to other patients. Trajectory II first showed a decrease during the first 2 years and then showed an increase to year 4. The patients in this trajectory first experienced an improvement and then a progression to the worse level (severe). The patients in trajectory III illustrated progression of PD to the worse level (severe) at the 4-year follow-up.

A challenge, and dichotomy, in data-driven discoveries is that subsequent studies may be needed to ascertain biological and clinical plausibility and value of these findings. This is also an issue in related fields. For instance, in the field of radiomics, there is significant evidence on the value of using advanced, and sometimes computationally-intense imaging features as biomarkers of disease; however, that is only one step, and subsequent or parallel analyses including biological plausibility and meaning are needed (110). Thus, there is a need for follow-up studies to further interpret and study data-driven identification of PD progression pathways.

Fereshtehnejad *et al.* (30) utilized the global composite outcome score from clinical data such as some motor and non-motor features to identify progression rate in PD. They first clustered PD patients using these scores and then compared progression of the scores between different subtypes in 4.5-year follow-up. Zhang *et al.* (31) attempted to specify the heterogeneity of PD (subtypes) using clinical data and conventional imaging features, as well as their progression rates using Long-Short Term Memory (a deep learning algorithm) applied to longitudinal clinical dataset. As proven in our previous paper (29), subtypes identified from the dataset without radiomics features are not robust to the variations in features and samples, and these subtypes may depend on the size of the features and samples. Thus, we do not compare the previous studied to this study. Meanwhile, unlike this study, to the best of our knowledge, previously published studies did not cluster the trajectory of the PD progression (30,31,38,39,43,46,111).

In the second part of this study, we employed HMLSs including multiple DRAs and CAs to predict these

trajectories using TD01. As shown in *Figure 7*, HMLSs such as SNEA + NPNNC and LLEA + NPNNC generated accuracies of 78.4% and 79.2%, respectively. Meanwhile, Some HMLSs such as SNEA + Lib_SVM and t_SNE+ NPNNC also had appropriate results, although they were less consistent than the best HMLSs. In a recent study of ours to improve prediction (46) and diagnosis tasks (111), radiomics features extracted from DAT SPECT images, going beyond conventional imaging features, were observed to obtain significant performances. We later observed that utilizing HMLSs, predictive algorithms linked with feature subset algorithms enables very good predictions of motor (43) and cognitive (38) outcomes. In addition, employing deep learning methods enables us to significantly improve prediction of motor outcome (39). The usage of the other datasets such as CSD0 and CSD01 did not improve prediction of the trajectories. Furthermore, in our past effort (42), we employed HMLSs including feature selection algorithms linked with CAs to predict cross-sectional PD subtypes in year 4. In that study, we saw limited predictive performance when employed CSD0 and CSD01 while we reached accuracies over 90% using TD01. As a result, constructing datasets as timeless, resulting in an increase in the sample size, can enhance prediction performance.

In this study, there is a restricting factory in the size of patients for outcome prediction. To tackle the restriction, we employed a timeless approach and it led to improved prediction performance. In this work, we utilized FEAs for dimensionality reduction to avoid over-fitting, although employing feature selection algorithms such as GA and least absolute shrinkage and selection operator (LASSO) were also possible.

Our study has several strengths. Importantly, we combined different categories of databases including motor, non-imaging, conventional imaging features, and radiomics features to generate a very comprehensive set of features. In addition, we focused on added value of radiomic features which was provided by standardized manner as well as based on guidelines from the image biomarker standardization initiative, aiming towards identifying robust progression trajectories in PD, and then predicting these trajectories using data in years 0 and 1.

Conclusions

We aimed to identify robust PD progression trajectories in 4-year follow-up, incorporating clinical and imaging

data. We also aimed to predict these trajectories using data in years 0 and 1. Based on three CEMs linked with PCA + KMA, we were able to identify three progression trajectories in 4-year follow-up. The identified trajectories were distinct. Trajectory I showed no enhanced progression for the period, while trajectories II and III showed enhanced progression. We also investigated a range of HMLSs including CAs linked with FEAs in order to predict these trajectories from early year data. SNEA + NPNNC and LLEA + NPNNC provided highest accuracies among all HMLSs. Overall, we conclude that combining clinical information with SPECT-based radiomics features, coupled with optimal utilization of HMLSs, can identify distinct trajectories and predict outcome in PD. In the future, we aim to predict non-motor outcome in year 4 by applying machine learning methods on datasets including clinical as well as radiomics features. Moreover, we aim to identify and predict different categories of response to treatment in PD patients, utilizing clustering of on/off-drug data as well as predicting these categories. Furthermore, we aim to predict start-date of taking drug and the administered dose. We also aim to predict rate of increase in the dose in subsequent years.

Footnote

Reporting Checklist: The authors have completed the MDAR checklist. Available at https://dx.doi.org/10.21037/qims-21-425

Conflicts of Interest: The authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/qims-21-425). AR reports that he received grant from Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2019-06467. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Simuni T, Sethi K. Nonmotor manifestations of Parkinson's disease. Ann Neurol 2008;64 Suppl 2:S65-80.
2. Wolter E. Non-motor extranigral signs and symptoms in Parkinson's disease. Parkinsonism Relat Disord 2009;3:6-12.
3. Bayulkem K, Lopez G. Non-motor fluctuations in Parkinson's disease: clinical spectrum and classification. J Neurol Sci 2010;289:89-92.
4. Foltynie T, Brayne C, Barker RA. The heterogeneity of idiopathic Parkinson's disease. J Neurol 2002;249:138-45.
5. Berg D, Postuma RB, Bloem B, Chan P, Dubois B, Gasser T, Goetz CG, Halliday GM, Hardy J, Lang AE, Litvan I, Marek K, Obeso J, Oertel W, Olanow CW, Poewe W, Stern M, Deuschl G. Time to redefine PD? Introductory statement of the MDS Task Force on the definition of Parkinson's disease. Mov Disord 2014;29:454-62.
6. Lewis SJ, Foltynie T, Blackwell AD, Robbins TW, Owen AM, Barker RA. Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. J Neurol Neurosurg Psychiatry 2005;76:343-8.
7. Eggers C, Pedrosa DJ, Kahraman D, Maier F, Lewis CJ, Fink GR, Schmidt M, Timmermann L. Parkinson subtypes progress differently in clinical course and imaging pattern. PLoS One 2012;7:e46813.
8. Dai D, Wang Y, Wang L, Li J, Ma Q, Tao J, Zhou X, Zhou H, Jiang Y, Pan G, Xu L, Ru P, Lin D, Pan J, Xu L, Ye M, Duan S. Polymorphisms of DRD2 and DRD3 genes and Parkinson's disease: A meta-analysis. Biomed Rep 2014;2:275-81.

916

Salmanpour et al. Longitudinal clustering and prediction of PD progression

9. Lebouvier T, Chaumette T, Paillusson S, Duyckaerts C, Bruley des Varannes S, Neunlist M, Derkinderen P. The second brain and Parkinson's disease. Eur J Neurosci 2009;30:735-41.

10. McNamara P, Stavitsky K, Harris E, Szent-Imrey O, Durso R. Mood, side of motor symptom onset and pain complaints in Parkinson's disease. Int J Geriatr Psychiatry 2010;25:519-24.

11. Bell PT, Gilat M, O'Callaghan C, Copland DA, Frank MJ, Lewis SJ, Shine JM. Dopaminergic basis for impairments in functional connectivity across subdivisions of the striatum in Parkinson's disease. Hum Brain Mapp 2015;36:1278-91.

12. Bohnen NI, Albin RL. The cholinergic system and Parkinson disease. Behav Brain Res 2011;221:564-73.

13. Maril S, Hassin-Baer S, Cohen OS, Tomer R. Effects of asymmetric dopamine depletion on sensitivity to rewarding and aversive stimuli in Parkinson's disease. Neuropsychologia 2013;51:818-24.

14. Ventura MI, Baynes K, Sigvardt KA, Unruh AM, Acklin SS, Kirsch HE, Disbrow EA. Hemispheric asymmetries and prosodic emotion recognition deficits in Parkinson's disease. Neuropsychologia 2012;50:1936-45.

15. Cubo E, Martín PM, Martin-Gonzalez JA, Rodríguez-Blázquez C, Kulisevsky J; ELEP Group Members. Motor laterality asymmetry and nonmotor symptoms in Parkinson's disease. Mov Disord 2010;25:70-5.

16. Antonini A, Isaias IU, Canesi M, Zibetti M, Mancini F, Manfredi L, Dal Fante M, Lopiano L, Pezzoli G. Duodenal levodopa infusion for advanced Parkinson's disease: 12-month treatment outcome. Mov Disord 2007;22:1145-9.

17. Tippmann-Peikert M, Park JG, Boeve BF, Shepard JW, Silber MH. Pathologic gambling in patients with restless legs syndrome treated with dopaminergic agonists. Neurology 2007;68:301-3.

18. Lang AE, Lozano AM. Parkinson's disease. Second of two parts. N Engl J Med 1998;339:1130-43.

19. Savitt JM, Dawson VL, Dawson TM. Diagnosis and treatment of Parkinson disease: molecules to medicine. J Clin Invest 2006;116:1744-54.

20. Hely MA, Morris JG, Traficante R, Reid WG, O'Sullivan DJ, Williamson PM. The sydney multicentre study of Parkinson's disease: progression and mortality at 10 years. J Neurol Neurosurg Psychiatry 1999;67:300-7.

21. Fereshtehnejad SM, Zeighami Y, Dagher A, Postuma RB. Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression. Brain 2017;140:1959-76.

22. Graham JM, Sagar HJ. A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: identification of three distinct subtypes. Mov Disord 1999;14:10-20.

23. Erro R, Vitale C, Amboni M, Picillo M, Moccia M, Longo K, Santangelo G, De Rosa A, Allocca R, Giordano F, Orefice G, De Michele G, Santoro L, Pellecchia MT, Barone P. The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients. PLoS One 2013;8:e70244.

24. Post B, Speelman JD, de Haan RJ; CARPA-study group. Clinical heterogeneity in newly diagnosed Parkinson's disease. J Neurol 2008;255:716-22.

25. Gasparoli E, Delibori D, Polesello G, Santelli L, Ermani M, Battistin L, Bracco F. Clinical predictors in Parkinson's disease. Neurol Sci 2002;23 Suppl 2:S77-8.

26. Poletti M, Frosini D, Pagni C, Lucetti C, Del Dotto P, Tognoni G, Ceravolo R, Bonuccelli U. The association between motor subtypes and alexithymia in de novo Parkinson's disease. J Neurol 2011;258:1042-5.

27. Salmanpour M, Shamsaei M, Saberi A, Hajianfar G, Ashrafinia S, Davoodi-Bojd E, Soltanian-Zadeh H, Rahmim A. Hybrid Machine Learning Methods for Robust Identification of Parkinson's Disease Subtypes. J Nucl Med 2020;61:1429.

28. Salmanpour M, Shamsaei M, Saberi A, Hajianfar G, Soltanian-Zadeh H, Rahmim A. Radiomic features combined with hybrid machine learning robustly identify Parkinson's disease subtypes. Annual AAPM Meeting, 2020.

29. Salmanpour MR, Shamsaei M, Saberi A, Hajianfar G, Soltanian-Zadeh H, Rahmim A. Robust identification of Parkinson's disease subtypes using radiomics and hybrid machine learning. Comput Biol Med 2021;129:104142.

30. Fereshtehnejad SM, Romenets SR, Anang JB, Latreille V, Gagnon JF, Postuma RB. New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes. JAMA Neurol 2015;72:863-73.

31. Zhang X, Chou J, Liang J, Xiao C, Zhao Y, Sarva H, Henchcliffe C, Wang F. Data-Driven Subtyping of Parkinson's Disease Using Longitudinal Clinical Records: A Cohort Study. Sci Rep 2019;9:797.

32. Salmanpour M, Saberi A, Hajianfar G, Shamsaei M, Soltanian-Zadeh H, Rahmim A. Cluster analysis on longitudinal data of Parkinson's disease subjects. Proc Annual AAPM Meeting, 2020.

33. Nieuwboer A, De Weerdt W, Dom R, Bogaerts K.

Prediction of outcome of physiotherapy in advanced Parkinson's disease. Clin Rehabil 2002;16:886-93.

34. Grill S, Weuve J, Weisskopf MG. Predicting outcomes in Parkinson's disease: comparison of simple motor performance measures and The Unified Parkinson's Disease Rating Scale-III. J Parkinsons Dis 2011;1:287-98.

35. Arnaldi D, De Carli F, Famà F, Brugnolo A, Girtler N, Picco A, Pardini M, Accardo J, Proietti L, Massa F, Bauckneht M, Morbelli S, Sambuceti G, Nobili F. Prediction of cognitive worsening in de novo Parkinson's disease: Clinical use of biomarkers. Mov Disord 2017;32:1738-47.

36. Fyfe I. Prediction of cognitive decline in PD. Nat Rev Neurol 2018;14:316-7.

37. Gao C, Sun H, Wang T, Tang M, Bohnen NI, Müller MLTM, Herman T, Giladi N, Kalinin A, Spino C, Dauer W, Hausdorff JM, Dinov ID. Model-based and Model-free Machine Learning Techniques for Diagnostic Prediction and Classification of Clinical Outcomes in Parkinson's Disease. Sci Rep 2018;8:7129.

38. Salmanpour MR, Shamsaei M, Saberi A, Setayeshi S, Klyuzhin IS, Sossi V, Rahmim A. Optimized machine learning methods for prediction of cognitive outcome in Parkinson's disease. Comput Biol Med 2019;111:103347.

39. Leung K, Salmanpour MR, Saberi A, Klyuzhin I, Sossi V, Rahmim A. Using deep-learning to predict outcome of patients with Parkinson's disease. Sydney, Australia: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2018.

40. Emrani S, McGuirk A, Xiao W. Prognosis and Diagnosis of Parkinson's Disease Using Multi-Task Learning. New York, NY, USA: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17). Association for Computing Machinery, 1457-66.

41. Salmanpour MR, Shamsaei M, Saberi A, Setayeshi S, Taherinezhad E, Klyuzhin IS, Tang J, Sossi V, Rahmim A. Machine Learning Methods for Optimal Prediction of Outcome in Parkinson's Disease. Available online: https://rahmimlab.files.wordpress.com/2019/03/salmanpour_mic18_pd_prediction_machine_learning.pdf

42. Salmanpour M, Saberi A, Shamsaei M, Rahmim A. Optimal Feature Selection and Machine Learning for Prediction of Outcome in Parkinson's Disease. J Nucl Med 2020; 61:524.

43. Salmanpour MR, Shamsaei M, Saberi A, Klyuzhin IS, Tang J, Sossi V, Rahmim A. Machine learning methods for optimal prediction of motor outcome in Parkinson's

disease. Phys Med 2020;69:233-40.

44. Parashos SA, Luo S, Biglan KM, Bodis-Wollner I, He B, Liang GS, Ross GW, Tilley BC, Shulman LM; NET-PD Investigators. Measuring disease progression in early Parkinson disease: the National Institutes of Health Exploratory Trials in Parkinson Disease (NET-PD) experience. JAMA Neurol 2014;71:710-6.

45. Chaudhuri KR, Naidu Y. Early Parkinson's disease and non-motor issues. J Neurol 2008;255 Suppl 5:33-8.

46. Rahmim A, Huang P, Shenkov N, Fotouhi S, Davoodi-Bojd E, Lu L, Mari Z, Soltanian-Zadeh H, Sossi V. Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images. Neuroimage Clin 2017;16:539-44.

47. Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol 2011;95:629-35.

48. Miller SA, Mayol M, Moore ES, Heron A, Nicholos V, Ragano B. Rate of Progression in Activity and Participation Outcomes in Exercisers with Parkinson's Disease: A Five-Year Prospective Longitudinal Study. Parkinsons Dis 2019;2019:5679187.

49. Aleksovski D, Miljkovic D, Bravi D, Antonini A. Disease progression in Parkinson subtypes: the PPMI dataset. Neurol Sci 2018;39:1971-6.

50. Han C, Rundo L, Murao K, Noguchi T, Shimahara Y, Milacski ZÁ, Koshino S, Sala E, Nakayama H, Satoh S. MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. BMC Bioinformatics 2021;22:31.

51. Nakao T, Hanaoka S, Nomura Y, Murata M, Takenaga T, Miki S, Watadani T, Yoshikawa T, Hayashi N, Abe O. Unsupervised Deep Anomaly Detection in Chest Radiographs. J Digit Imaging 2021;34:418-27.

52. Rundo L, Militello C, Tangherloni A, Russo G, Vitabile S, Gilardi MC, Mauri G. NeXt for neuro-radiosurgery: A fully automatic approach for necrosis extraction in brain tumor MRI using an unsupervised machine learning technique. Int J Imaging Syst Technol 2017;28:21-37.

53. Mekhmoukh A, Mokrani K. Improved Fuzzy C-Means based Particle Swarm Optimization (PSO) initialization and outlier rejection with level set methods for MR brain image segmentation. Comput Methods Programs Biomed 2015;122:266-81.

54. Ashrafinia S. Quantitative Nuclear Medicine Imaging using Advanced Image Reconstruction and Radiomics. USA: Johns Hopkins University, 2019.

55. Mwangi B, Tian TS, Soares JC. A review of feature

918

Salmanpour et al. Longitudinal clustering and prediction of PD progression

reduction techniques in neuroimaging. Neuroinformatics 2014;12:229-44.

56. Reif M, Shafait F. Efficient feature size reduction via predictive forward selection. Pattern Recognition 2014;47:1664-73.

57. Patel MJ, Khalaf A, Aizenstein HJ. Studying depression using imaging and machine learning methods. Neuroimage Clin 2015;10:115-23.

58. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometr Intell Lab Syst 1987;2:37-52.

59. Wenming Z, Cairong Z, Li Z. An Improved Algorithm for Kernel Principal Component Analysis. Neural Process Lett 2005;22:49-56.

60. Maaten L, Hinton G. Visualizing Data Using t-SNE. J Mach Learn Res 2008;9:2579-605.

61. Tinsley H, Tinsley D. Uses of factor analysis in counseling psychology research. J Couns Psychol 1987;34:414-24.

62. Sammon JW. A Nonlinear Mapping for Data Structure Analysis. IEEE Transactions on Computers 1969;5:401-9.

63. Sun J, Crowe M, Fyfe C. Extending Sammon mapping with Bregman divergences. Information Sciences 2012;187:72-92.

64. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science 2000;290:2319-23.

65. Shi H, Yin B. Robust L-Isomap with a Novel Landmark Selection Method. Available online: https://www.hindawi.com/journals/mpe/2017/3930957/

66. Belkin M, Niyogi P. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. Available online: https://papers.nips.cc/paper/2001/file/f106b7f99d2cb30c3db1c3cc0fde9ccb-Paper.pdf

67. Lewandowski M, Rincon JMd, Makris D, Nebel J. Temporal extension of laplacian eigenmaps for unsupervised dimensionality reduction of time series. Istanbul, Turkey: 20th International Conference on Pattern Recognition, 2010.

68. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science 2000;290:2323-6.

69. Mead A. Review of the Development of Multidimensional Scaling Methods. J R Stat Soc Ser C Appl Stat 1992;41:27-39.

70. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, Zucker SW. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. Proc Natl Acad Sci U S A 2005;102:7426-31.

71. Coifman R, Lafon S. Diffusion maps. Appl Comput

Harmon Anal 2006;21:5-30.

72. Agrafiotis DK. Stochastic proximity embedding. J Comput Chem 2003;24:1215-21.

73. Li P, Chen S. A review on Gaussian Process Latent Variable Models. CAAI Trans Intell Technol 2016;1:366-76.

74. Lawrence N. Learning for Larger Datasets with the Gaussian Process Latent Variable Model. J Mach Learn Res 2007;2:243-50.

75. Hinton G, Roweis S. Stochastic Neighbor Embedding. Adv Neural Inf Process Syst 2003;15:857-64.

76. Nam K, Je H, Choi S. Fast stochastic neighbor embedding: a trust-region algorithm. Budapest: IEEE International Joint Conference on Neural Networks, 2004.

77. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313:504-7.

78. Jain A, Duin R, Mao J. Statistical Pattern Recognition: A Review. Available online: https://ieeexplore.ieee.org/document/824819

79. Jain M, Murty M, Flynn P. Data Clustering: A Review. ACM Comput Surv 1999;31:264-323.

80. Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa LDF, Rodrigues FA. Clustering algorithms: A comparative approach. PLoS One 2019;14:e0210236.

81. Bholowalia P, Kumar A. Kmeans: A clustering technique based on elbow method and k-means in WSN. Int J Comput Appl 2014;105:17-24.

82. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat Theory Methods 1974;3:1-27.

83. Schwarz G. Estimating the dimension of a model. Ann Stat 1987;6:461-4.

84. Goodman RM, Smyth P. Decision tree design using information theory. Available online: https://www.sciencedirect.com/science/article/abs/pii/S1042814305800202

85. Chourasia S. Survey paper on improved methods of ID3 decision tree. Available online: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.402.7503

86. Denison, D, Holmes C, Mallick B, Smith A. Bayesian Methods for Nonlinear Classification and Regression. New York: John Wiley and Sons, 2002.

87. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol 2011;2:1-27.

88. Cortes C, Vapnik V. Support-Vector Networks. Mach Learn 1995;20:273-97.

89. Crammer K, Singer Y. On the algorithmic implementation of multiclass kernel-based vector machines. J Mach Learn

2001;2:265-92.

90. Altman N. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 1992;46:175-85.

91. Suguna N, Thanushkodi K. An Improved k-Nearest Neighbor Classification Using Genetic Algorithm. Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.404.1475&rep=rep1&type=pdf

92. Talbot J, Lee B, Kapoor A, Tan D. EnsembleMatrix: interactive visualization to support machine learning with multiple classifiers. Boston, MA: CHI '09: CHI Conference on Human Factors in Computing Systems, 2009.

93. Shan J, Zhang H, Liu W, Liu Q. Online Active Learning Ensemble Framework for Drifted Data Streams. IEEE Trans Neural Netw Learn Syst 2019;30:486-98.

94. Lu Q, Qiao X. Sparse Fisher's linear discriminant analysis for partially labeled data. Stat Anal Data Min 2018;11:17-31.

95. McLachlan G. Discriminant Analysis and Statistical Pattern Recognition. New Jersey: John Wiley & Sons, 1992.

96. Kusy M, Zajdel R. Probabilistic neural network training procedure based on Q(0)-learning algorithm in medical data classification. Appl Intell 2014;41:837-54.

97. Specht DF. Probabilistic neural networks and the polynomial Adaline as complementary techniques for classification. IEEE Trans Neural Netw 1990;1:111-21.

98. Joutsijoki H, Haponen M, Rasku J, Aalto-Setälä K, Juhola M. Error-Correcting Output Codes in Classification of Human Induced Pluripotent Stem Cell Colony Images. Biomed Res Int 2016;2016:3025057.

99. Hamming R. Error Detecting and Error Correcting Codes. Bell System Technical Journal 1950;29:147-60. Available online: https://course.ece.cmu.edu/~ece447/s13/lib/exe/fetch.php?media=bstj29-2-147.pdf

100. Alsmadi M, Omar K, Noah S. Back Propagation Algorithm: The Best Algorithm. Available online: http://paper.ijcsns.org/07_book/200904/20090451.pdf

101. Rumelhart D, Geoffrey E. Leaner Representations By back-Propagating errors. Nature 1986;323:533-6.

102. Breiman L. Random Forests. Mach Learn 2001;45:5-32.

103. Jehad A, Khan R, Ahmad N. Random Forests and Decision Trees. Available online: https://www.uetpeshawar.edu.pk/TRP-G/Dr.Nasir-Ahmad-TRP/Journals/2012/Random%20Forests%20and%20Decision%20Trees.pdf

104. Nijole M, Aleksan. Investigation of financial market prediction by recurrent neural network. Available online: http://journal.kolegija.lt/iitsbe/2011/Maknickiene-investigation-IITSBE-2011-2(11)-3-8.pdf

105. Hu Z, Tong T, Genton MG. Diagonal likelihood ratio test for equality of mean vectors in high-dimensional data. Biometrics 2019;75:256-67.

106. Salmanpour MR, Shamsaei M, Rahmim A. Feature selection and machine learning methods for optimal identification and prediction of subtypes in Parkinson's disease. Comput Methods Programs Biomed 2021;206:106131.

107. Tiwari K, Jamal S, Grover S, Goyal S, Singh A, Grover A. Cheminformatics Based Machine Learning Approaches for Assessing Glycolytic Pathway Antagonists of Mycobacterium tuberculosis. Comb Chem High Throughput Screen 2016;19:667-75.

108. Marras C, Rochon P, Lang AE. Predicting motor decline and disability in Parkinson disease: a systematic review. Arch Neurol 2002;59:1724-8.

109. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007;23:2507-17.

110. Tomaszewski MR, Gillies RJ. The Biological Meaning of Radiomic Features. Radiology 2021;298:505-16.

111. Cheng Z, Zhang J, He N, Li Y, Wen Y, Xu H, Tang R, Jin Z, Haacke EM, Yan F, Qian D. Radiomic Features of the Nigrosome-1 Region of the Substantia Nigra: Using Quantitative Susceptibility Mapping to Assist the Diagnosis of Idiopathic Parkinson's Disease. Front Aging Neurosci 2019;11:167.

Patient data

Segmentation of dorsal striatum (DS) on DaT SPECT images via T1 weighted MRI

Imaging was performed 4±0.5 h following injection of DAT SPECT (^{123}I-ioflupane; 111–185 MBq). Thyroid update was blocked via pre-treatment of subjects with saturated iodine solution (10 drops in water) or perchlorate (1,000 mg) prior to injection. Data acquisition consisted of 128×128 raw SPECT projection data acquired every 3 degrees, 120 projections, 20% symmetric photopeak windows centered on 159 and 122 keV, and a total scan duration of ~30–45 min. A HERMES system (Hermes Medical Solutions, Stockholm, Sweden) was used to perform iterative OSEM reconstruction on the input raw SPECT projection data, for all studies to ensure consistency. Subsequently, PMOD (PMOD Technologies, Zurich, Switzerland) was used for attenuation correction. Ellipses where drawn on the images and Chang 0 attenuation correction were applied invoking a site-specific mu as empirically derived from phantom data (as acquired in site initiation for the trial). Following this, standard 3D Gaussian post-smoothing (6.0 mm FWHM) was applied (43,46).

As shown in *Figure S1*, pre-processing process is included 3 parts. We firstly resized pixels of images as 1×1×1. After fixing orientation of images, we modified intensity bias of images. The original aspect ratio of the MR images was maintained. After pre-processing section, we applied these images to Free surfer for segmenting region of interests (ROI) such as left and right caudate as well as putamen. Co-registration of SPECT on MRI were performed through two steps. In first step, we manually co-reiterated SPECTS on their MRI through Mango software. We tried various combinations of transform options, cost function options and search cost options on each image. We experimentally considered best solution for each image. There were various options for transform options includes (I) 2D rigid body, (II) translation only, (III) rigid body, (IV) global scale, (V) full scale, and (VI) full affine. There were also different options for cost function options and search cost options included: (I) correlation ratio, (II) mutual information, (III) norm mutual information, (IV) normalized correlation, and least square. Finally, we employed the rigid co-registration algorithm accompanied with normalized mutual information on the co-registered DAT SPECT images as well as Gaussian smoothing kernel with a width of 7 mm for all images. In the end, we overlaid the structures segmented in MRI on the co-registered SPECT images. After co-registration stage, we applied each mentioned ROI to the SERA to extract radiomics features. SERA has been extensively standardized in reference to the Image Biomarker Standardization Initiative (ISBI) (112), and studied in multi-center radiomics standardization publications by the IBSI (113) and the quantitative imaging network (QIN) (114). There is a total of 487 standardized radiomics features in SERA, including: 79 first-order features (morphology, statistical, histogram and intensity-histogram features), 272 higher-order 2D features, and 136 3D features. We included all 79 first-order features and 136 3D features (54,113,115).

List of features used

We considered the following 981 features (as shared publicly): multiple Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) measures, a range of task/exam performances, socioeconomic/family histories, genetic features, and SPECT image features. Segmentation of regions-of-interest (ROIs; left and right
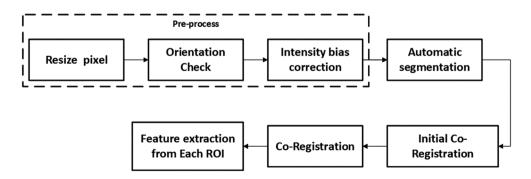

Figure S1 Diagram of image processing and radiomics feature extraction based on MRI images.

both 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. For consistency, we only included patients who were off medication (e.g., Levodopa/dopamine agonist) for >6 hours prior to testing/imaging (47). we separately collected information for patients based on each year. Subsequently, timeless datasets were constructed by appending cross sectional datasets within a single set of data. This approach aims to gather data with larger number of subjects and features.

Machine learning algorithms

Feature extraction algorithms

Principal component analysis (PCA)
PCA is a known tool for linear dimensionality reduction and feature extraction. Using an orthogonal transformation enables us to convert a dataset with correlated variables into new dataset with linearly uncorrelated variables called principal components. The first principle component has the highest variance in compered to other principle components. Moreover, number of principal components is less than or equal to the number of original variables (58).

Kernel PCA
Kernel PCA is an extended nonlinear form of the PCA using techniques of kernel methods. it is more useful to extract the complicated spatial structure of high-dimensional features in compared to simple PCA. Thus, the Kernel PCA is increasingly using in machine learning application (59).

t-distributed stochastic neighbor embedding (t-SNE)
t-SNE is a machine learning algorithm which performs non-linear dimensionality reduction to embed high-dimensional data for visualization in a low-dimensional space of two or three dimensions. Specifically, it models each high dimensional object by a two or three-dimensional point in such a way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability (60).

Factor analysis (FA)
FA, as an analytic technique enables us to reduce a large number of correlated variables to a smaller number of dimensions. The goal of the FA is to achieve parsimony by minimizing explanatory concepts to explain the maximum amount of common variance in a correlation matrix (61).

Sammon mapping algorithm (SMA)
SMA is a non-linear algorithm which maps a high dimensional dataset to a low dimensional dataset. It also preserves the structure of inter-point distances in the high dimensional dataset in the lower dimensional space.it works based on minimizing error function (called Sommon error or Stress error), as shown in equation [1] (62).

$$E = \frac{1}{\sum_{i<j} d_{ij}^*} \sum_{i<j} \frac{\left(d_{ij}^* - d_{ij}\right)^2}{d_{ij}^*} \quad [1]$$

Where d_{ij}^* is the distance between ith and jth datapoint in the original space and d_{ij} is the distance between their projections.

The stress function was improved using left Bregman divergence and right Bregman divergence (63).

Isomap algorithm (IsoA)
IsoA, as a nonlinear dimensionality reduction algorithm, is one of low-dimensional embedding methods. It was employed for compute a quasi-isometric in order to embed low-dimensional points of a set of high-dimensional data points (64).

LandMark Isomap algorithm (LMIsoA)
LMIsoA is a fast IsoA which is faster than the Isomap. It works based on landmark MDS (Multi-Dimension Scaling) so that it selects a group of points termed as Landmarks and implements classical MDS on them. After computing the shortest path from each data point to the landmark points, the geodesic distance matrix is applied on classical MDS to find the low-dimensional embedding of the landmark points (65).

Laplacian eigenmaps algorithm (LEA)
LEA, as a non-linear dimensionality reduction, aims to build a graph from neighborhood connections of the dataset. The discrete approximation of the low-dimensional manifold in the high-dimensional space are considered as connections between nodes which constructing by each data point. Minimizing the cost function based on the graph enables us to guarantee that close points are mapped close to each other in the low-dimensional space (66,67).

Locally linear embedding algorithm (LLEA)
LLEA, as an unsupervised learning algorithm, computes

low dimensional, neighborhood preserving embeddings of high dimensional data. It works based on exploiting the local symmetries of linear reconstructions in order to find nonlinear structure in high dimensional data (68).

Multidimensional scaling algorithm (MDSA)

MDSA, known as Principal Coordinates Analysis, provides a visual representation of the pattern of proximities (i.e., correlation matrix or distance matrix) among a set of objects. It aims to map these dissimilarities as distances between points in a low dimensional space so that these distances correspond as closely as possible to the dissimilarities (69).

Diffusion map algorithm (DMA)

DMA, as a non-linear technique, reduces high dimensional space to low dimensional space by re-organising data according to parameters of its underlying geometry. It computes a family of embeddings of dataset into Euclidean space using coordinates from the eigenvectors and eigenvalues of a diffusion operator on the data. These distances among datapoints in the embedded spaces are equal to diffusion distances between probability distributions centered at those points (70,71).

Stochastic proximity embedding algorithm (SPEA)

SPEA, as a novel self-organizing algorithm, produces meaningful underlying dimensions from proximity data. It reduces high dimensional dataset to low-dimensional Euclidean embeddings so that the similarities between a set of related observations preserve. This algorithm initially selects a random configuration and then adjusting their coordinates according to iteratively refining it by repeatedly selecting pairs of objects at random (72).

Gaussian process latent variable model (GPLVM)

GPLVM, as a dimensionality reduction method, is a flexible Bayesian non-parametric modeling method that learns a low-dimensional representation of high-dimensional data by a Gaussian process. In this case, the kernel and learning hyperparameters of Gaussian process regression are selected to describe the mapping of high dimensional dataset to low dimensional dataset (73,74).

Stochastic neighbor embedding algorithm (SNEA)

SNE, as a dimensionality reduction method, is a probabilistic method of embedding objects. It maps a high-dimensional vectors or pair wise dissimilarities into a lower dimensional space so that neighbor identities are preserved.

Dissimilarities, providing via applying a Gaussian model to each object in the high-dimensional space, are used to define a probability distribution which has potential neighbors of the object (75).

Symmetric stochastic neighbor embedding algorithm (Sym_SNEA)

SNEA has slow convergence; a fast SNE algorithm was proposed that is approximately 4-6 times faster. This algorithm works based on a trust-region method to discover a reliable direction as well as efficient step size with the help of a quadratic model of the objective function (76).

Autoencoders algorithms (AA)

AA, a three-layered neural network, constructs the "building block" of deep learning. It can be converted high-dimensional data to low-dimensional codes by training a multilayer neural network. Gradient descent can be employed for updating weight matrix in the AA (77).

K-Means algorithm (KMA) for unsupervised clustering

The KMA, as a type of unsupervised learning method, tries to partition the dataset into K distinct non-overlapping sub-groups. It works based on minimizing the sum of the squared distance between the subjects and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) similar subjects in a cluster (116,117).

We considered 3 first features of the first subject as initial centroid values to cluster patients (original sub-cluster), while we used 3 first columns of the first patient as initial centroids to identify progression trajectories. We also discovered that different combination of initial centroids resulted in same sub-clusters and merely the order of sub-clusters was changed. In future work, we plan to explore other initialization methods. Moreover, we considered 1,000 epochs (maximum iteration) for this algorithm.

Clustering evaluation methods

Calinski-Harabatz criterion

The Calinski-Harabasz index of a clustering is the ratio of the between-cluster variance (which is essentially the variance of all the cluster centroids from the dataset's grand centroid) to the total within-cluster variance. The within-cluster variance will decrease as k increases; the rate of decrease should slow down past the optimal k. The between-cluster variance will increase as k, but the rate of increase should slow down past the optimal k. So, in theory,

the ratio of between-cluster variance to within-cluster variance should be maximized at the optimal k. The ratio is formally defined as:

$$C(k) = \frac{Trace(B)}{Trace(W)} \frac{n-k}{k-1} \qquad [2]$$

where B is the between-cluster covariance matrix [so high values of Trace(B) denote well-separated clusters], W is the within-cluster covariance matrix [so low values of Trace(W) correspond to compact clusters], n is the number of the data points and k is the number of the clusters (82).

Bayesian information criterion (BIC)

In statistics, the BIC is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the Akaike information criterion (AIC). Both BIC and AIC attempt to resolve this problem by introducing a penalty term for the number of parameters in the model; the penalty term is larger in BIC than in AIC. The BIC is formally defined as (83):

$$BIC = k \ln(n) - 2\ln(\hat{L}) \qquad [3]$$

where

$\hat{L}$ = the maximized value of the likelihood function of the model

x = the observed data;

n = the number of data points in x, the number of observations, or equivalently, the sample size;

k = the number of parameters estimated by the model.

Elbow criterion

In cluster analysis, the elbow method is a heuristic used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use (81).

Classifiers

Decision tree classification (DTC)

DTC technique is one of the most popular techniques in the emerging field of data mining. There are various methods for constructing the DTC. Induced decision tree (ID3) is the basic algorithm for constructing the DTC (84). There are many algorithms based on classification that is sample based, neural networks, Bayesian networks, support vector machine, and decision tree. The DTC classifies samples by sorting them down the tree from the root to some leaf node, which provides the classification of samples. Each node in the tree specifies a test of some attribute of the sample and each branch descending from that node corresponds to one of the possible values for this attribute (104). In this specific work, the maximum depth was not set so the algorithm would continue until all leaves were pure. To measure the quality of a split, a "Gini" function was used (GINI function describes the impurity of each node; each child node was purer than its parent node so that the GINI function was minimized).

Library for support vector machines (Lib_SVM)

LIBSVM is a library for support vector machines (SVM) (87). SVM, as a supervised learner, was initially designed for binary classification, defined by a separating hyperplane. Optimal hyperplane, which categorizes new examples, is regulated by labeled training data. In two-dimensional space, this hyperplane is a line dividing a plane in two parts where in each class lay in either side (88). To extend SVM to the multi-class scenario, several classification models were proposed such as the one by Crammer and Singer (89). It was replaced the misclassification error of an example with the piecewise linear bound.

K nearest neighborhood classifier (KNNC)

k-NN, as a supervised and non-parametric algorithm, employs for classification and regression tasks. On the other words, it works based on instance-based learning, where the function is only approximated locally. In both cases, the input consists of the k closest training examples in the feature space as well as output is a class membership. Thereby, objects are classified by a plurality vote of their neighbors (90,91).

Ensemble leaner classifier (ELC)

ELC, as a supervised learner, works according to voting process of multiple classifiers. All classifiers, combined to solve a common problem, participate in prediction process. It mostly results in better predictive performance than use of a sole classifier (92,93).

Linear discriminant analysis classifier (LDAC)

LDAC, as a generalization of Fisher's linear discriminant, is employs in machine learning area to discover a linear combination of features that separates two or more classes of objects. Thus, the combination may be used as a linear classifier or as a dimensionality reduction before the

classification (94,95).

New probabilistic neural network classifier (NPNNC)
PNNC, as a supervised feed-forward neural network with a complex structure consists of an input layer, a pattern layer, a summation layer and an output layer. A single training parameter (probability density functions) is considered to activate the neurons in the pattern layer (96,97).

Error-correcting output codes model classifier (ECOCMC)
ECOCMC, as a general classification framework and a supervised algorithm, needs multiple classifiers, similarly to the ELC. It works based on two stages: encoding and decoding. Encoding phase consists of designing coding matrix. The columns and rows of the matrix show binary classifiers and codewords for classes respectively. Designing of a coding matrix can be made using binary coding and ternary coding. In decoding phase, we need to find which one of the classes' codewords is the closest one the test example's codeword (98,99).

Multi-layer perceptron-back propagation (MLP-BP)
A multilayer perceptron is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output so it is a modified MLP that uses three or more layers of neurons (nodes) with nonlinear activation functions, and is more powerful than the perceptron in that it can distinguish data that are not linearly separable, or separable by a hyper plane (46,100). In this specific work, we used a three-layer neural network and the number of neurons in each layer was adjusted via automated machine learning hyperparameter tuning automatically.

Random forest algorithm (RFA)
Random Forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them (102,20). Depth of structure was adjusted via automated machine learning hyperparameter tuning automatically. Number of trees and number of splits were set to 1000 and 5, respectively.

Recurrent neural network (RNN)
Recurrent neural network is a deep learning algorithm. The RNN as fundamentally different neural network from feed-forward architectures was investigated for modelling of nonlinear behavior (41,104). In this work, we used a model with many inputs to one output. In this specific work, we used a three-layer neural network and the number of neurons in each layer was adjusted via automated machine learning hyperparameter tuning automatically.

Automated machine learning hyperparameter tuning

In this work, automated machine learning hyperparameter tuning was employed to automatically adjust intrinsic parameters such as the number of neurons, and the number of layers in the classification algorithms. We applied this approach to various algorithms such as LOLIMOT, RBF, RNN, MLP-BP, RFA to automatically tune the parameters. Automated tuning, which was implemented with our in-house code, executes an error minimization search scheme to optimize the hyperparameters starting with the random initialization. Employing this approach enables us to pursue a systematic trial-and-error search scheme for tuning the parameters (38).

Hoteling's t squared test

This statistical test is used to evaluate the equality of the mean vectors of two populations (with n1 and n2 samples). Each of two groups has p features. Assume that population 1 is distributed as $N_p(\mu_1, \Sigma_1)$ and population 2 is distributed as $N_p(\mu_2, \Sigma_2)$, where $N_p(\mu,\Sigma)$ is the p-variable multivariate normal distribution with mean vector μ and covariance matrix Σ. The null hypothesis that $\mu_1 = \mu_2$ can be tested using the test statistic:

$$T2 = \frac{n_1 n_2}{n_1 + n_2} (Y_1 - Y_2)' S_{pooled}^{-1} (Y_1 - Y_2) \qquad [4]$$

where Y_1 and Y_2 are the two sample mean vectors, n_1 and n_2 are the two sample sizes, and S_{pooled}^{-1} is the inverse of the pooled covariance matrix which is calculated using:

$$S_{Pooled} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} \qquad [5]$$

Here, S_1 and S_2 are the estimated covariance matrices calculated from the two samples. If we make the additional assumption that $\Sigma_1 = \Sigma_2$, T2 follows Hotelling's T-squared distribution when the null hypothesis is true. That is, $T2 = T_{P,n_1+n_2-2}^2$. Reject the null hypothesis if $T2 \geq T_{P,n_1+n_2-2}^2$. Note that rejecting the null hypothesis concludes that at

least one pair of the p sets of group response means are unequal (105).

References

112. Image biomarker standardisation initiative. Available online: https://arxiv.org/abs/1612.07003
113. Hall JF, Crocker TF, Clarke DJ, Forster A. Supporting carers of stroke survivors to reduce carer burden: development of the Preparing is Caring intervention using Intervention Mapping. BMC Public Health 2019;19:1408.
114. McNitt-Gray M, Napel S, Jaggi A, Mattonen SA, Hadjiiski L, Muzi M, et al. Standardization in Quantitative Imaging: A Multicenter Comparison of Radiomic Features from Different Software Packages on Digital Reference Objects and Patient Data Sets. Tomography 2020;6:118-28.
115. Ashrafinia S, Dalaie P, Yan R, Huang P, Pomper M, Schindler T, Rahmim A. Application of Texture and Radiomics Analysis to Clinical Myocardial Perfusion SPECT Imaging. J Nucl Med 2018;59(Supplement 1):94.
116. Kanungo T, Member S, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans Pattern Anal Mach Intell 2002;24:881-92.
117. Francis BK, Babu SS. Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. J Med Syst 2019;43:162.