2-1-2021

Robust identification of Parkinson's disease subtypes using radiomics and hybrid machine learning

Mohammad R. Salmanpour

Mojtaba Shamsaei

Abdollah Saberi

Ghasem Hajianfar

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

See next page for additional authors

Recommended Citation

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

Authors

Mohammad R. Salmanpour, Mojtaba Shamsaei, Abdollah Saberi, Ghasem Hajianfar, Hamid Soltanian-Zadeh, and Arman Rahmim

Computers in Biology and Medicine 129 (2021) 104142

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

Robust identification of Parkinson's disease subtypes using radiomics and hybrid machine learning

Mohammad R. Salmanpour [a,b], Mojtaba Shamsaei [a], Abdollah Saberi [c], Ghasem Hajianfar [d], Hamid Soltanian-Zadeh [e,f], Arman Rahmim [b,g,*]

[a] Department of Energy Engineering and Physics, Amirkabir University of Technology, Tehran, Iran
[b] Department of Physics & Astronomy, University of British Columbia, Vancouver, BC, Canada
[c] Department of Computer Engineering, Islamic Azad University, Tehran, Iran
[d] Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Science, Tehran, Iran
[e] School of Electrical & Computer Engineering, University of Tehran, Tehran, Iran
[f] Departments of Radiology and Research Administration, Henry Ford Health System, Detroit, USA
[g] Department of Radiology, University of British Columbia, Vancouver, BC, Canada

ARTICLE INFO

Keywords:
Unsupervised clustering
Machine learning
Parkinson's disease (PD)
Radiomics features
DAT SPECT Imaging
Disease subtypes

ABSTRACT

Objectives: It is important to subdivide Parkinson's disease (PD) into subtypes, enabling potentially earlier disease recognition and tailored treatment strategies. We aimed to identify reproducible PD subtypes robust to variations in the number of patients and features.

Methods: We applied multiple feature-reduction and cluster-analysis methods to cross-sectional and timeless data, extracted from longitudinal datasets (years 0, 1, 2 & 4; Parkinson's Progressive Marker Initiative; 885 PD/163 healthy-control visits; 35 datasets with combinations of non-imaging, conventional-imaging, and radiomics features from DAT-SPECT images). Hybrid machine-learning systems were constructed invoking 16 feature-reduction algorithms, 8 clustering algorithms, and 16 classifiers (C-index clustering evaluation used on each trajectory). We subsequently performed: i) identification of optimal subtypes, ii) multiple independent tests to assess reproducibility, iii) further confirmation by a statistical approach, iv) test of reproducibility to the size of the samples.

Results: When using no radiomics features, the clusters were not robust to variations in features, whereas, utilizing radiomics information enabled consistent generation of clusters through ensemble analysis of trajectories. We arrived at 3 distinct subtypes, confirmed using the training and testing process of k-means, as well as Hotelling's T2 test. The 3 identified PD subtypes were 1) mild; 2) intermediate; and 3) severe, especially in terms of dopaminergic deficit (imaging), with some escalating motor and non-motor manifestations.

Conclusion: Appropriate hybrid systems and independent statistical tests enable robust identification of 3 distinct PD subtypes. This was assisted by utilizing radiomics features from SPECT images (segmented using MRI). The PD subtypes provided were robust to the number of the subjects, and features.

1. Introduction

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disease [1–3]. This disease more affects the population over 65 years of age [4,5]. PD is defined by widespread neuronal loss [6,7], resulting in a range of primary motor symptoms. Motor and non-motor symptoms significantly impact quality of life in PD patients [8–12]. A recent study shew that cognitive and psychiatric changes are directly linked with PD progression [13]. Although, there is currently no permanent therapy for PD, but temporary symptomatic therapies with levodopa [14] and dopaminergic agonists [15] can improve the quality of life through alleviation from early symptoms [16,17].

A multicenter study in 10 years following up [18] showed that PD is a heterogeneous disease so that 9 out of 126 patients needed a wheelchair unless aided, whereas 13 patients did not show any significant functional restriction. Studies of disease progression can be broadly divided into two categories involving: i) individual symptoms/measures, or ii) collection of symptoms/measures. In first approach, some studies

* Corresponding author. Department of Physics & Astronomy, University of British Columbia, Vancouver, BC, Canada.
E-mail address: arman.rahmim@ubc.ca (A. Rahmim).

https://doi.org/10.1016/j.compbiomed.2020.104142
Received 8 September 2020; Received in revised form 20 November 2020; Accepted 21 November 2020
Available online 25 November 2020

focused on the prediction of outcomes in PD subjects [19–30]. Novel prognostic biomarkers of PD enable better implementation for disease-modifying trials [31]. For example, Prashos et al. [32], showed that motor symptoms are more reliable primary symptoms of disease progression. In another study by Noyce et al. [33], it was demonstrated that higher UPDRS III scores cause a lower quality of life in patients with PD.

Imaging phenotypes, beyond pure usage of clinical measures, have the potential to add further value to the assessment of PD [21,21, 34–37]. As an example, identification of patients that are symptomatic without evidence of dopamine deficit (SWEDD) as performed using dopamine transporter (DAT) SPECT imaging has already made a difference in the recruitment of patients in clinical trials [38,39]. Furthermore, beyond the usage of conventional imaging features, the field of radiomics has the potential to provide further analysis of imaging data [21,37,40,41]. In an effort of ours [21], we focused on prediction of motor outcome, showing that radiomics features, when added to clinical features, can significantly improve prediction of outcome. We later showed that by using hybrid systems (predictive algorithms accompanied by feature subset selector algorithms), one can obtain excellent predictions of motor [26] and cognitive [27] outcomes. Furthermore, our recent ongoing efforts linked with deep learning-based prediction of outcome showed significant improvements through the discovery of patterns in images [28].

PD represents heterogeneity in various (not merely individual) phenotypes for PD subjects [42–45]. As such, definition of PD subtypes based on collective data could help provide a better understanding of underlying disease mechanisms, predict disease course, and design clinical trials [46]. Some studies based on clinical features have attempted to derive subtypes of PD as a clustering task, as listed in supplemental Table S1 [31,44,47–50]. Parkinson's disease (PD) has been classically introduced as a progressive degenerative motor disease associated with the degeneration of striatal dopamine neurons [51–55]. Hoehn and Yahr (YH) [56] categorized disease severity into 5 stages. Since then, stage 0 has been added, and stages 1.5 and 2.5 have been proposed and are widely used [57]. Patients with H&Y stage 2 have lower levels of dopamine binding than stage 1 [58], while patients with H&Y stages 3 and 4 have significantly impaired language function, working memory and visuospatial function compared to those in stage 2 [59]. In a study by Szewczyk-Krolikowski et al. [60], phenotypic heterogeneity was documented across age and gender in both motor and non-motor symptoms. As summarized in Supplemental Table S1, studies making use of clinical information in the past have categorized patients into 2 [49], 3 [31,46,61], 4 [44,48,50,62–65], 5 [47,66], 6 [67] or 7 [56,57] sub-clusters. In a recent study [68], previously published data-driven PD subtype classification systems were re-assessed, demonstrating limited reproducibility and suggesting a need for the establishment of standards for validation and use of clustering systems [68].

However, clinical measurements based on visual examinations in different clinical centers are prone to errors for defining globally reproducible PD subtypes. In fact, in general, there have been significant challenges with the reproducibility of PD subtypes, and a desire for specific standards for the clustering of PD subjects [68]. There exists an emerging usage of automated processing invoking machine learning algorithms for improved task performance without the need for explicit programming [69]. Approaches based on machine learning aim to enable improved and automatic classification, prediction, or clustering by capturing statistically robust patterns present in the analyzed data.

In the present work, we specifically aim to robustly identify subtypes of PD (an unsupervised clustering task). This is an important key to better understand underlying disease mechanisms, predict disease course, and design clinical trials. Furthermore, we incorporate and study the role of imaging in our analysis, as functional imaging enables spatial localization of molecular changes as well as accurate and consistent quantification of their distribution [70,71]. Many machine learning algorithms are not able to work with many input features, and thus it is necessary to reduce high feature dimensions into few dimensions to be used as inputs, as pursued in this work. Overall, in the present work, we select a range of hybrid algorithms amongst various families of learner algorithms, and also, various combinations of datasets based on timeless and cross-sectional approaches are considered, to select robust sub-clusters of PD which minimally depend on sample size and feature size.

2. Materials and methods

In what follows we outline our various data selection, image processing, machine learning and analysis methods. Data and code are made publicly available (details at the end of this manuscript). In short, we generated multiple datasets (timeless and cross-sectional), performed segmentation and feature extraction, and utilized hybrid ML systems. We also employed two methods to compare clusters as generated from different datasets. Finally, by dividing datasets on small groups, we evaluated our finding based on variations into sample size. Fig. 1 shows different steps of implementation in this study, as elaborated next.

2.1. Image processing and feature extraction from regions of interest (ROIs)

We performed segmentation via two methods, as elaborated next.

2.1.1. Segmentation of dorsal striatum (DS) on DAT SPECT images via T1 weighted MRI

As shown in Supplemental Figure S1 and elaborated in supplemental section II (part A, i), following few preprocessing steps, we employed the Free Surfer package to automatically segment T1 weighted-MRI images and to register SPECT images to MRI data. Fig. 2 shows images related to the segmentation registration steps. Subsequently, we utilized our standardized SERA software package [72] (publicly available at: https://qurit.ca/software/sera/) to extract radiomics features from the ROIs (left and right caudate and putamen). SERA has been extensively standardized in reference to the Image Biomarker Standardization Initiative (ISBI) [73] and studied in multi-center radiomics standardization publications by the IBSI [74] and the Quantitative Imaging Network (QIN) [75]. There are 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 [72,74,76].

2.1.2. Direct segmentation of dorsal striatum (DS) on DAT SPECT images

As an alternative approach, we performed direct segmentation based on the SPECT image itself [77], for greater ease, given the lack of need for MRI images and computational speed. The multiple steps (x1-x10) are elaborated in Supplemental Figure S2 (in supplemental section II, part A.ii). The results of each step are shown in Fig. 3.

2.2. Patient data

As shown in Supplemental Figure S3, 35 datasets extracted from the PPMI database (www.ppmi-info.org/data) were analyzed. For consistency, we only considered patients being off medication (e.g. Levodopa/dopamine agonist) for >6 h before testing/imaging [78]. Two types of data-gathering frameworks were used, namely the use of timeless and cross-sectional datasets. In cross-sectional datasets, we separately collected information for patients based on each year. Subsequently, timeless datasets were constructed by appending cross-sectional datasets within a single set (885/1139 PD subject visits with/without imaging). This approach aims to gather data with a larger number of subjects and features. As such, we were able to select datasets with

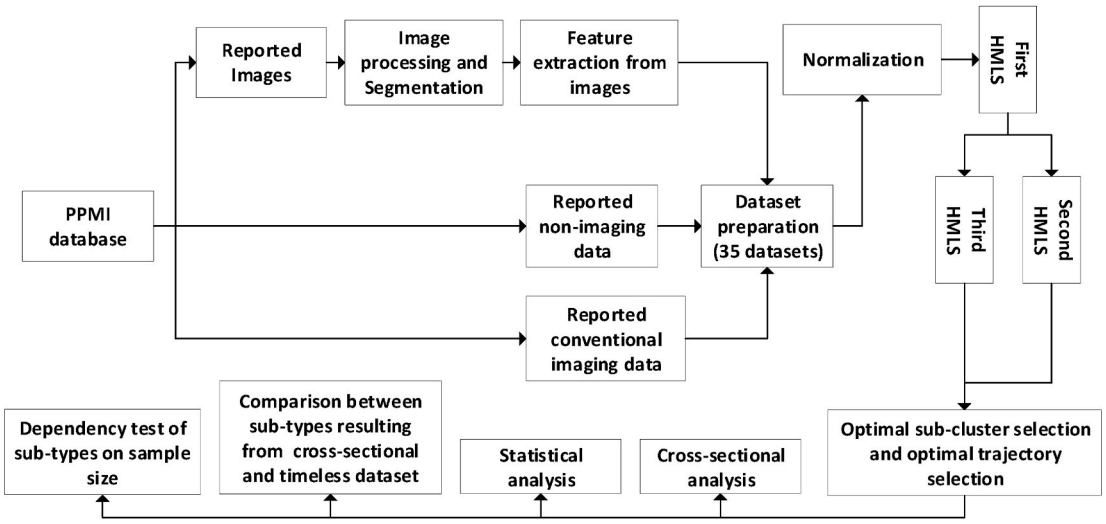

Fig. 1. The workflow in our study. The 3 types of HMLS are discussed and depicted in subsection C and Supplemental Figure S4.

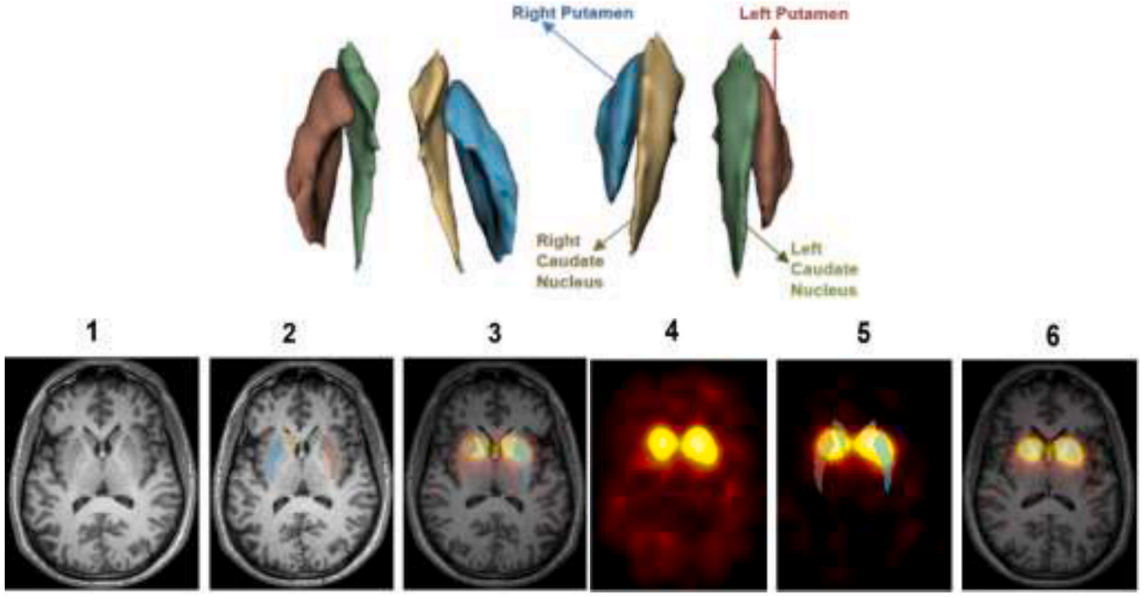

Fig. 2. Segmentation and fusion process based on MRI segmentation: (*top*) 3D segmentation of DS; (*bottom*) (1) T1-weighted MRI image, (2) Segmented MRI, (3) Fusion of MRI, DAT SPECT, and MRI Segmentation, (4) DAT SPECT, (5) Fusion of DAT SPECT and Segmentation, and (6) Fusion of DAT SPECT and MRI.

different sample sizes and feature sizes (see Supplemental Table S2), to better study different combinations of features within clustering tasks. Additionally, we considered 163 healthy-control (HC) subject visits. Variables included non-imaging clinical features (NCF) and conventional imaging features (CIF), as well as radiomics features extracted from SPECT (RFS) images as segmented via MRI (RFS-M) and SPECT itself (RFS-S), as elaborated in supplemental section II, part B. Traditionally, quantitative analysis, if performed at all, has been restricted to assessment of mean regional uptake in different ROIs [36]; by contrast, in the present work, we move beyond inclusion of conventional features, to study additional inclusion of radiomics features. To determine the effect of different categories, we employed combinations of the above-mentioned categories to study PD clustering:

1. NCF only
2. RFS-M only
3. NCF + CIF
4. NCF + CIF + RFS-M
5. NCF + CIF + RFS-S
6. NCF + RFS-S
7. NCF + RFS-M

2.3. Machine learning methods

We constructed 3 types of hybrid machine learning systems (HMLSs) (Supplemental Figure S4):

(1) First type employed 3 groups of algorithms: i) 16 Dimensionality Reduction Algorithms (DRA); ii) 8 Clustering Algorithms (CA); iii) 1 Individual Clustering Evaluation Method (ICEM) for the assessment of cluster number for each sole trajectory.

(2) Second and (3) Third HMLSs included the above-mentioned algorithms in addition to Collective Clustering Evaluation Methods CCEM1 and CCEM2, respectively, for *overall* optimization of cluster number considering various trajectories. CCEM1 focused on correlation analysis between clusters formed from different trajectories, while CCEM2 additionally utilized 16 Classifiers (C) to assess classification

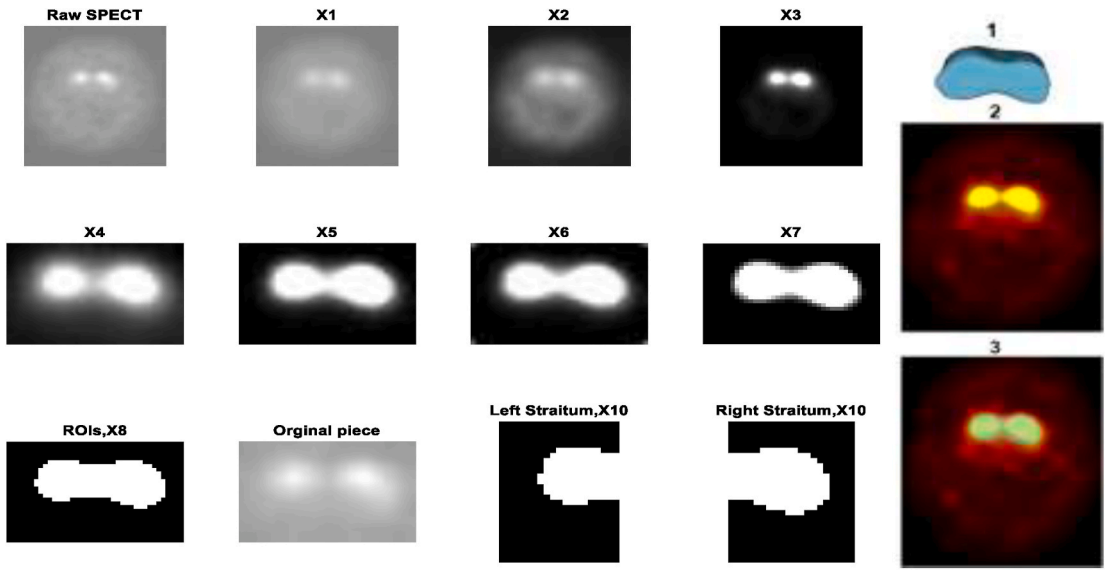

Fig. 3. Segmentation process based on SPECT: (*Left*) Steps in segmentation (described in supplemental section II, Part A,ii); and (*right)* showing (1) 3D segmentation of DS, (2) DAT SPECT, and (3) superposition of DAT SPECT and Segmentation.

accuracy, to arrive at optimal overall cluster number.

These algorithms are elaborated in Supplement Section D and briefly mentioned next.

2.3.1. Dimensionality reduction algorithms (DRA)

Feature reduction methods are employed to reduce the number of features in high-dimensional data to a limited number of most relevant features, thus reducing overfitting. Supervised DRAs requires outcome for feature subset selection. Unsupervised DRAs need no label for feature reduction, thus relying on information based on patterns found among input features [79–81]. In this study, 16 DRAs (only unsupervised) were employed (as elaborated in Supplemental section II, part D, i): 1) PCA (Principle Component Analysis) [82], 2) Kernel PCA [83], 3) t-SNE (t-distributed Stochastic Neighbor Embedding) [84], 4) Factor analysis (FA) [85], 5) SMA (Sammon Mapping algorithm) [86,87], 6) IsoA (Isomap Algorithm) [88], 7) LMIsoA (LandMark Isomap Algorithm) [89], 8) LEA (Laplacian Eigenmaps Algorithm) [90,91], 9) LLEA (locally linear embedding algorithm) [92], 10) MDSA (multidimensional scaling Algorithm) [93], 11) DMA (Diffusion map Algorithm) [94,95], 12) SPEA (Stochastic Proximity Embedding Algorithm) [96], 13) GPLVM (Gaussian Process Latent Variable Model) [97,98], 14) SNEA (Stochastic Neighbor Embedding Algorithm) [99], 15), Sym_SNEA (symmetric Stochastic Neighbor Embedding Algorithm) [100], and 16) AA (Aut oencoders Algorithms) [101].

2.3.2. Clustering algorithms (CA)

Clustering algorithms, as unsupervised machine learning methods, employ to identify natural groupings or clusters within multidimensional data based on some similarity measure [102–104]. These are all listed in the Supplement (as elaborated in Supplemental section II, part D, ii). Our study includes multiple CAs including 1) SOMA (Self--Organizing Map) [105], 2) APA (Affinity Propagation) [106], 3) HC-WM (Hierarchical clustering-Ward's Method) [107], 4) HC-CLA (Hierarchical Clustering -Complete Linkage Algorithms) [108], 5) HC-WLA (Hierarchical Clustering -Weighted Linkage Algorithm) [109], 6) KMA (K-Means Algorithm) [110,111], 7) KMeA (K-Mediods Algorithm) [112, 113], and 8) CwGMMA (Clustering with Gaussian Mixture Model Algorithms) [114].

2.3.3. Classifiers (C)

Classifier algorithms are all listed in the Supplement (as elaborated in Supplemental section II, part D, iii). Specifically, we selected 16 classifiers: 1) DTC (Decision Tree Classification) [115–117], 2) Lib_SVM (Library for Support Vector Machines) [118–120], 3) NBC (Naïve Bayes Classifier) [121,122], 4) KNNC (K Nearest Neighborhood Classifier) [123,124], 5) ELC (Ensemble Leaner Classifier) [125,126], 6) LDAC (Linear Discriminant Analysis Classifier) [127,128], 7) New PNNC (New probabilistic neural network Classifier) [129,130], 8) ECOCMC (Error-Correcting Output Codes Model Classifier) [131,132], 9 MLP_BP (Multilayer Perceptron_ Back Propagation) [133,134], 10)RFC (Random Forest Classifier) [135,136], 11) RNNC (Recurrent Neural Network Classifier) [137,138], 12), RBFA (Radial Basis Function Algorithm) [139], 13) LOLIMOT (Local Linear Model Trees Algorithm) [140,141], and 14, 15, 16) GLMCs (Gaussian Maximum likelihood Classifiers) types I, II and III (as described in supplemental file) [142–144].

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 [26].

2.3.4. Individual Clustering Evaluation Method (ICEM)

To evaluate clustering solutions, the so-called C-index was utilized (Supplemental section II, part D, iv). A small value for the C-index indicates good clustering [145,146]. We utilized this C-index criterion to evaluate a range of cluster numbers spanning 2 to 10.

2.3.5. Collective Clustering Evaluation Methods (CCEM 1 and 2)

Because of the variety of optimal cluster solutions in different trajectories of the first HMLS (as shown in Supplemental Figure S4), we considered 2 collective cluster evaluations for final overall cluster number selection. They utilized 1) Average of Correlation Factors (AOCF), and 2) Average of Classifier Performances (AOCP; classification accuracy), constituting final steps in 2nd and 3rd HMLSs, respectively (Supplemental Figure S4). For a given number of clusters, AOCF assessed how well results from different clustering methods correlate with one another, while AOCP assesses accuracies in ultimate

classification. Overall, given various ML trajectories, AOCF quantifies and optimizes the overall reproducibility of clusters, while AOCP evaluates overall reliability (for ultimate classification).

2.4. Analysis procedure

As shown in Fig. 1, we first generated 35 datasets that included different combinations as listed in section B. All the datasets were normalized based on the minimum and maximum of each feature. In the first HMLSs, we employed different hybrid trajectories and determined the optimal number of clusters for each trajectory (using ICEM). The optimal numbers of clusters were close to one another (often around 3 clusters), but not always consistent. We then performed ensemble analysis of trajectories, utilizing CCEM1 as well as CCEM2 (the latter accompanied by a range of classifiers), to arrive at a collective selection of optimal cluster number (see Supplemental Figure S4).

For the rest of our analysis, we utilized principal component analysis (PCA) for feature reduction and k-means algorithm (KMA) for clustering, as they each arrived at most consistent number of clusters (explained later). To assess robustness of our clusters, we pursued 4 approaches: (i) in cross-sectional analysis, for sub-groups obtained via the hybrid system, we assessed similarity between subtypes (see Supplemental Figure S5). To do this we utilized a training and testing process: following training of KMA using a specific year, the subjects of another year were processed using the trained model, and the results were compared with training based on the new set itself (testing). This enables assessment of the robustness of clusters between datasets. (ii) We employed High Dimensional Hoteling's T Squared Test [147] as a statistical approach (Supplemental section II, part E) for re-confirming findings. (iii) Furthermore, we made comparisons between clusters from cross-sectional vs. timeless datasets (see Supplemental Figure S6), and (iv) finally, by dividing our large timeless datasets into smaller groups, and then apply the clustering algorithm, to each group, we evaluated the dependency of our subtype identification on variations in sample sizes. All algorithms were implemented in the Matlab R 2020 b platform.

3. Results

3.1. First stage analysis including optimal cluster selection

First, we applied our 35 datasets to the first category of HMLS. Initially selected numbers of disease subtypes selected via ICEM were not consistent across different trajectories and datasets. Subsequently, we applied all datasets to the second and the third HMLSs enabling ensemble, collective selection of an optimal number of clusters. Tables 1 and 2 show the resulting AOCF and AOCP methods applied to timeless data respectively. It is seen that for different datasets and methods, the optimal number of PD sub-clusters is consistently found to be 3. Moreover, we also applied these methods to cross-sectional datasets as shown in Supplemental Tables S4 and S5, showing again that the most common optimal number of clusters is 3. For the rest of our work (below), the PCA + KMA trajectory was considered for generation and analysis of disease subtypes, because PCA and KMA were each seen (via ICEM) to result in the most consistent number of identified clusters when couple with other methods.

3.2. Second stage analysis including cross-linking different sub-groups across years

To discover similarity between sub-groups, we performed different tests as mentioned in section II.D. As shown in Supplemental Figure S5, after arranging two comparing datasets based on common features, we re-clustered our datasets with common features. Datasets without radiomics features were not able to reach consistent results in comparison to when we used all features (see Supplemental Tables S6 and S7). Meanwhile, datasets including radiomics features were able to reach consistent subtypes (Supplemental Table S8). This comparison shows a lack of the reproducibility of results based on the usage of clinical features alone, as also seen elsewhere [148]. Subsequently, we focused on the additional usage of radiomics features. First, we applied a dataset in a specific year to KMA. After training the algorithm via a specific year, we compared subjects of another year as clustered using the trained model with direct training based on the new year (testing process), to assess robustness of the identified subtype. Among various datasets, datasets including RFS-M reached more robust performance (Table 3) compared to datasets with RFS-S (see Supplemental Tables S9 and S10). Furthermore, we saw that datasets including RFS-M (e.g. without and with clinical features) resulted in exactly similar performances. As such, we concluded that radiomics featured, derived properly using MRI segmentation, are very important in the clustering of PD.

As an example, cell (2,1) in the top left Table 3 shows how many patients in G12" (year 1, sub-cluster 2) were associated with G01 (year 0, cluster 1) when we applied test data in year 1 to the network trained via specific data in year 0. Fig. 4 shows a consistent relationship between the various sub-clusters as identified in different years (i.e. that they are effectively similar).

Analysis of both cross-sectional and timeless data arrived at 3 optimal numbers of clusters. As shown in Table 3, the cluster groups obtained in datasets including RFS-M were more robust compared to cluster groups obtained by other datasets (see Supplemental Tables S9 and S10), when performing training and testing processes. Moreover, High Dimensional Hoteling T Squared Test re-confirmed our findings (p-value$<<$0.001).

3.3. Third stage analysis including comparison of timeless sub-groups and cross-sectional sub-groups

Additional tests aimed to further assess the reproducibility of our results, and sensitivity to the number of samples. After clustering

Table 1
Clustering scores are based on the AOCF method for timeless data. The maximum score in each dataset (column) is shown in bold. The last column is created by summing the previous columns.

Cluster number	NCF + CIF	NCF + CIF+ RFS-S	NCF + CIF + RFS-M	NCF	NCF + RFS-S	NCF + RFS-M	RFS-M	Overall score
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3.00	**0.26**	**0.27**	**0.43**	**0.27**	**0.36**	**0.41**	**0.49**	**2.50**
4.00	0.20	0.19	0.32	0.20	0.27	0.31	0.37	1.85
5.00	0.16	0.16	0.29	0.19	0.18	0.31	0.30	1.59
6.00	0.12	0.00	0.30	0.12	0.20	0.28	0.25	1.26
7.00	0.10	0.13	0.20	0.13	0.23	0.18	0.21	1.18
8.00	0.11	0.00	0.15	0.13	0.00	0.00	0.24	0.63
9.00	0.08	0.00	0.16	0.00	0.08	0.11	0.00	0.42
10.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.14

Table 2
Clustering scores are based on the AOCP method for timeless data. The maximum score in each dataset (column) is shown in bold. The last column is created by summing the previous columns.

Cluster number	NCF + CIF	NCF + CIF+ RFS-S	NCF + CIF + RFS-M	NCF	NCF + RFS-S	NCF + RFS-M	RFS-M	Overall score
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	0.90
3.00	**0.92**	**0.91**	**0.89**	**0.92**	**0.88**	**0.88**	**0.90**	**6.28**
4.00	0.92	0.91	0.88	0.91	0.00	0.87	0.00	4.48
5.00	0.91	0.91	0.88	0.91	0.88	0.87	0.89	6.24
6.00	0.00	0.00	0.88	0.00	0.88	0.00	0.89	2.64
7.00	0.00	0.00	0.88	0.91	0.00	0.87	0.00	2.66
8.00	0.91	0.00	0.00	0.91	0.00	0.00	0.00	1.82
9.00	0.91	0.91	0.00	0.00	0.00	0.87	0.00	2.68
10.00	0.91	0.90	0.88	0.00	0.88	0.87	0.00	4.43

Table 3
Association tables for cross-sectional linkage of datasets with RFS-M
Left Table: year-0 dataset used for training, and year-1 dataset used for testing; Right Table: year-1 dataset used for training, and year-0 dataset used for testing. GXY shows subtype-Y resulting from a training dataset in year-X, and GZY" shows the associated patients (in derived subtype-Y from another year-Z) used for testing.

	G01	G02	G03		G11	G12	G13
G11″	**0.77**	0.08	0.01	**G01″**	**0.96**	0.03	0.02
G12″	0.09	**0.92**	0	**G02″**	0.14	**0.86**	0
G13″	0.14	0	**0.99**	**G03″**	0.23	0	**0.77**

Left Table: year-1 dataset used for training, and year-2 dataset used for testing; Right Table: year-2 dataset used for training, and year-1 dataset used for testing

	G11	G12	G13		G21	G22	G23
G21″	**0.69**	0	0	**G11″**	**1**	0	0
G22″	0	**1**	0.24	**G12″**	0	**0.43**	0.57
G23″	0.31	0	**0.76**	**G13″**	0.18	0	**0.82**

Left Table: year-2 dataset used for training, and year-4 dataset used for testing; Right Table: year-4 dataset used for training, and year-2 dataset used for testing

	G21	G22	G23		G41	G42	G43
G41″	**1**	0.17	0	**G21″**	**0.79**	0.21	0
G42″	0	**0.77**	0	**G22″**	0	**1**	0
G43″	0	0.06	**1**	**G23″**	0	0.13	**0.87**

patients in the timeless dataset, we then split the clustered timeless dataset into 4 parts based on the original year of each study (i.e. years 0, 1, 2, and 4). As we had also independently clustered cross-sectional data, this approach allows further investigation of the robustness of our framework, by comparing the two datasets. We denote patients in a cluster Y as determined from the timeless dataset, who were originally from year X, as G_TXY. We also denote patients clustered into a cluster Y directly from the cross-sectional dataset of year X as G_CXY. Thus, we compare subtype in each split category obtained from timeless versus cross-sectional datasets, as depicted in Table 4.

As an additional investigation, in reverse, we concatenated patients with similar subtype in cross-sectional datasets of different years into a clustered timeless set, referred to as G_CY for a given sub-type Y. We then compared this set to G_TY, as directly obtained for different subtype Y from the timeless data. The comparisons are shown in Table 5.

As shown in Tables 4 and 5, results reached from RFS-M datasets are relatively consistent. These results are in fact more robust compared to when using NCF + CIF + RFS-S datasets (See Supplemental Tables S11 and S12). At the same time, result of NCF + RFS-S shown in Supplemental Tables S13 and S14 also depicted good performance, but usage of this dataset did not enable one to reach as a high cross-linking performance compared to datasets with RMS-M. As a result, RFS-M performs well in the clustering task, while other datasets result in less robust findings. Furthermore, we showed that conventional imaging features were not effective on their own towards robust clustering (unlike radiomics using RFS-M).

3.4. The identified subtypes

We had 3 categories of features, namely motor, non-motor, and imaging information. We considered 163 HC subject visits for additional comparison with the identified PD subtypes (images for 40 HC subjects were analyzed). After normalizing the features based on the minimum and maximum of each feature, we then assessed PD sub-clusters according to whether their features had significant differences with HC subjects (p-value < 0.01; Bonferroni corrected). For plotting in Fig. 5, we selected features that were significantly correlated with the identified subtype (Spearman correlation, p-value < 0.01).

Fig. 6 shows a radar plot of the above-mentioned features. What is different, is that, for improved visibility, we have reversed the orientation of some features (where now higher values mean the worst progression for all features; which we know by correlation with 5 well-understood features). And to further simplify, we show the radar after final normalization based on the maximum value of each feature.

As illustrated, HC subjects (age range: 68 ± 11.4, female: 59, male: 104, Hoehn, and Yahr's score range: 0.01 ± 0.11) have the lowest scores compared to PD sub-clusters. PD patients in Cluster I (age range: 69.5 ± 9.8, female: 88, male: 117, Hoehn and Yahr's score range: 1.71 ± 0.58) showed slower progression in all domains, including motor and non-motor symptoms, and imaging, compared to other PD sub-clusters, while they have higher scores compared to scores in HC subjects. Patients in Cluster II (Age range: 70 ± 9.7, female: 167, male: 125, Hoehn and Yahr's score range: 1.70 ± 0.55) have significantly enhanced tremor among PD sub-clusters, although other symptoms were mostly larger than those in Cluster I. Cluster III (age range: 71.9 ± 10.9, female: 58, male: 44, Hoehn and Yahr's score range: 1.85 ± 0.54) illustrated significantly enhanced values for various features (except for tremor) especially the imaging features. The 3 identified subtypes can be referred to as 1) mild, 2) intermediate, and 3) severe, especially in terms of dopaminergic deficit (imaging), while also having some escalating motor and non-motor manifestations.

3.5. Final reproducibility test

This section shows how many samples are needed to reproduce our findings. As clustering of subjects in datasets containing RFS-M was similar, we only considered two timeless datasets: i) RFS-M (with 885 subject visits), and ii) NCF (with 1139 subject visits) for reproducibility analysis. We divided our datasets into various subsets (divisions) from 2 to 30 parts. After applying each data-subset to KMA, we compared the subtype obtained by split parts, with the original subtype obtained by the timeless dataset. These correlation coefficients can be introduced as a standard for the reproducibility test. Figs. 7 and 8 show how much clustering results changed when sample sizes were reduced. As shown in Fig. 8, sub-clusters resulting from the NCF dataset were more dependent

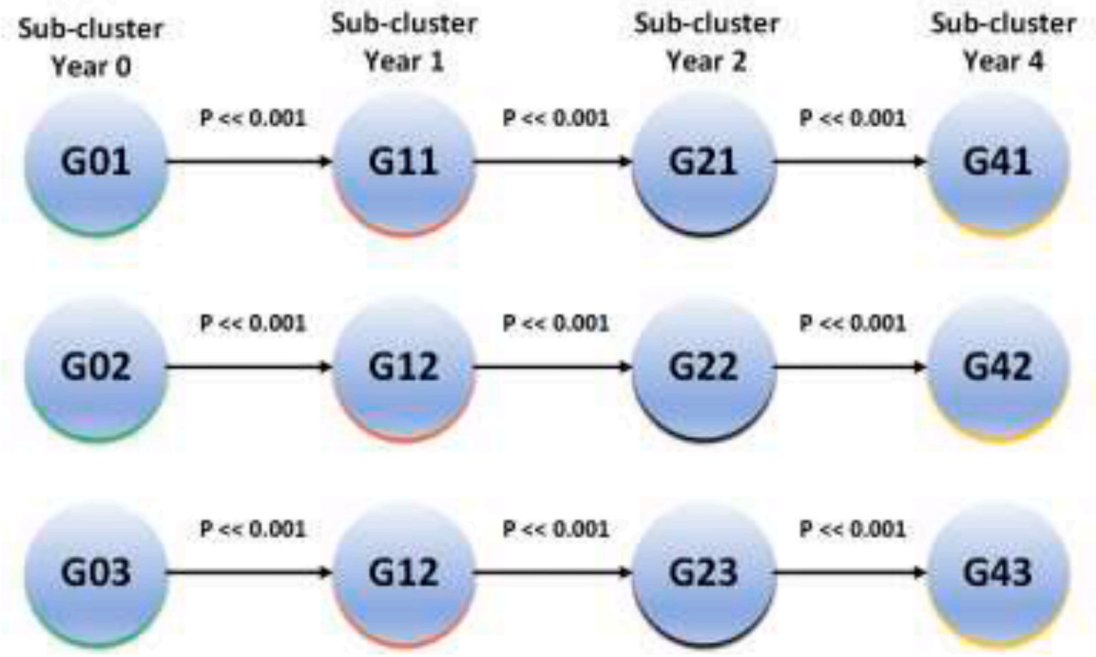

Fig. 4. Diagram of the relationship between different sub-cluster groups when utilizing RFS-M for analysis. P-values are calculated by High Dimensional Hotelling's T2 Test.

Table 4
Comparison of the split timeless subtype with subtype directly derived from cross-sectional datasets including RFS-M. G_TXY represents the sub-cluster Y obtained by the timeless dataset as well as split based on year X, and G_CXY represents the sub-cluster Y obtained by the cross-sectional dataset in year X.

	G_T01	G_T02	G_T03		G_T11	G_T12	G_T13
G_C01	**0.59**	0.41	0.01	**G_C11**	**0.73**	0.27	0
G_C02	0	**1**	0	**G_C12**	0	**1**	0
G_C03	0.31	0	**0.69**	**G_C13**	0.37	0	**0.63**
	G_T21	**G_T22**	**G_T23**		**G_T41**	**G_T42**	**G_T43**
G_C21	**0.96**	0.04	0	**G_C41**	**1**	0	0
G_C22	0	**1**	0	**G_C42**	0	**1**	0
G_C23	0.16	0	**0.84**	**G_C43**	0.4	0	**0.6**

Table 5
Comparison of computed subtype obtained by cross-sectional datasets included RFS-M with subtype obtained by timeless dataset including RFS-M. G_TY represents the sub-cluster Y obtained by the timeless dataset, and G_CY represents the formed sub-cluster Y obtained by the cross-sectional datasets.

	G_T1	G_T2	G_T3
G_C1	**0.76**	0.24	0
G_C2	0	**1**	0
G_C3	0.29	0	**0.71**

on sample sizes, while it was already seen before that datasets without radiomics features strongly depended on features (Supplemental Tables S6 and S7). The last sub-figure (averaged correlation) especially can be introduced as a standard to measure reproducibility.

3.6. Comparing our findings to previous studies

As shown in Fig. 8, the subtypes obtained by the clinical dataset without radiomics features depend more strongly on sample sizes compared to those obtained by datasets with radiomics features. In any case, differences in inclusion criteria from datasets, feature selection, and methodology between cluster analysis studies make it difficult to compare subtypes derived from different studied [148]. Though PD is a heterogeneous disease and longitudinal data is needed to assess PD more extensively, many studies [31,44,47,48,50,62–64] have used baseline datasets for clustering PD subjects. Previous efforts did not focus on the use of radiomics features, and often no imaging information, from which we conclude that the results may have more strongly dependent on sample size, and reproducibility remains challenging [148]. Recent studies and analyses [68,148] have noted prior efforts to have limited reproducibility, which can also be understood in the context of our findings. At the same time, we attempt to compare our findings to other works in Table 6, where we report the average values of features in sub-clusters. We considered all features reported in each study for our assessment, though we show then more relevant features in the table. These comparisons were performed based on the relative progression of disease between the sub-clusters.

3.7. Independent validation: t-SNE (T-distributed Stochastic Neighbor Embedding) plot

t-SNE [84] is a machine learning algorithm that performs non-linear dimensionality reduction to embed high-dimensional data for visualization in a low-dimensional space of two or three dimensions. Specifically, it transforms each high-dimensional dataset to 2 or 3 dimensions in such a way that similar datapoints are modeled by nearby datapoints and dissimilar datapoints are modeled by distant datapoints with high probability. We utilized this plot as an independent visualization test of our efforts to identify PD subtypes. Fig. 9 plots t_SNE results for cross-sectional years 0, 1, 2 and, 4, as well as timeless data, color-coding our derived ML-based subtypes and HC group on the plots. The plots show clear distinctions (particularly along the first dimension) among the sub-groups.

4. Discussion

Robust identification of PD subtypes has the potential to guide the design, and interpretation of clinical trials involving neuroprotective and symptomatic therapy [149]. To this end, we explored the application of hybrid machine learning systems (HMLSs) to two datasets,

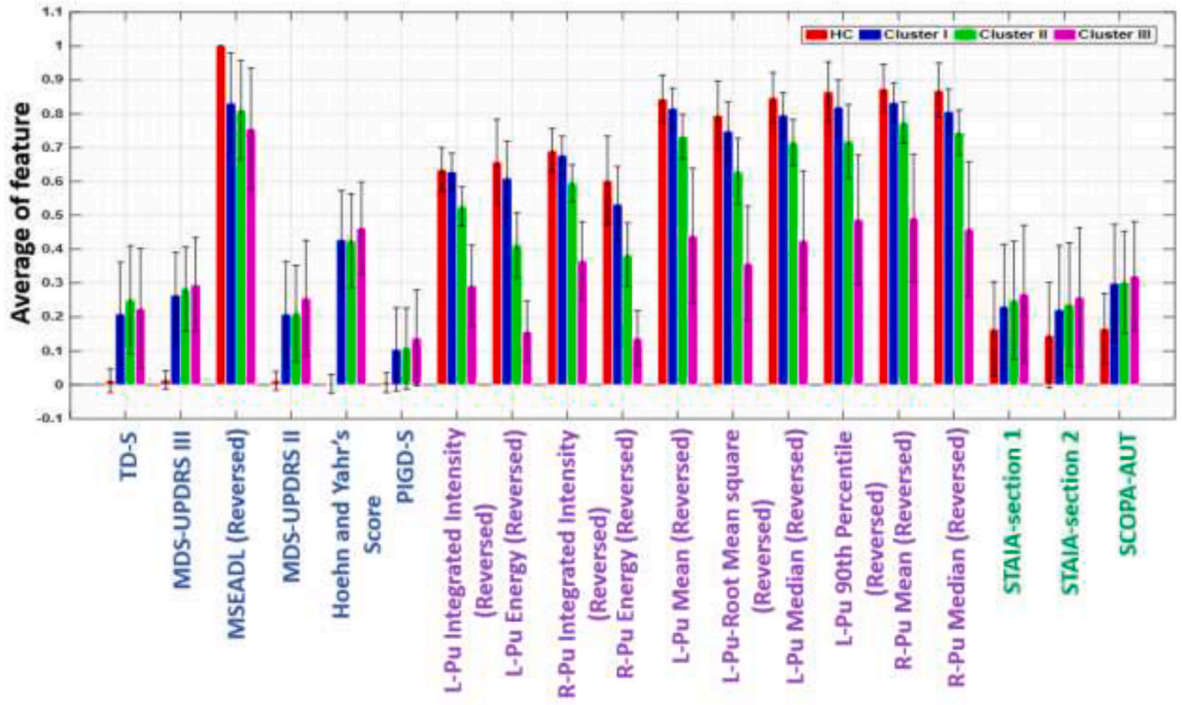

Fig. 5. Bar plot of three kinds of features, motor (dark blue text), imaging (purple text) and non-motor (green text) for our 3 identified PD sub-clusters and HC groups.

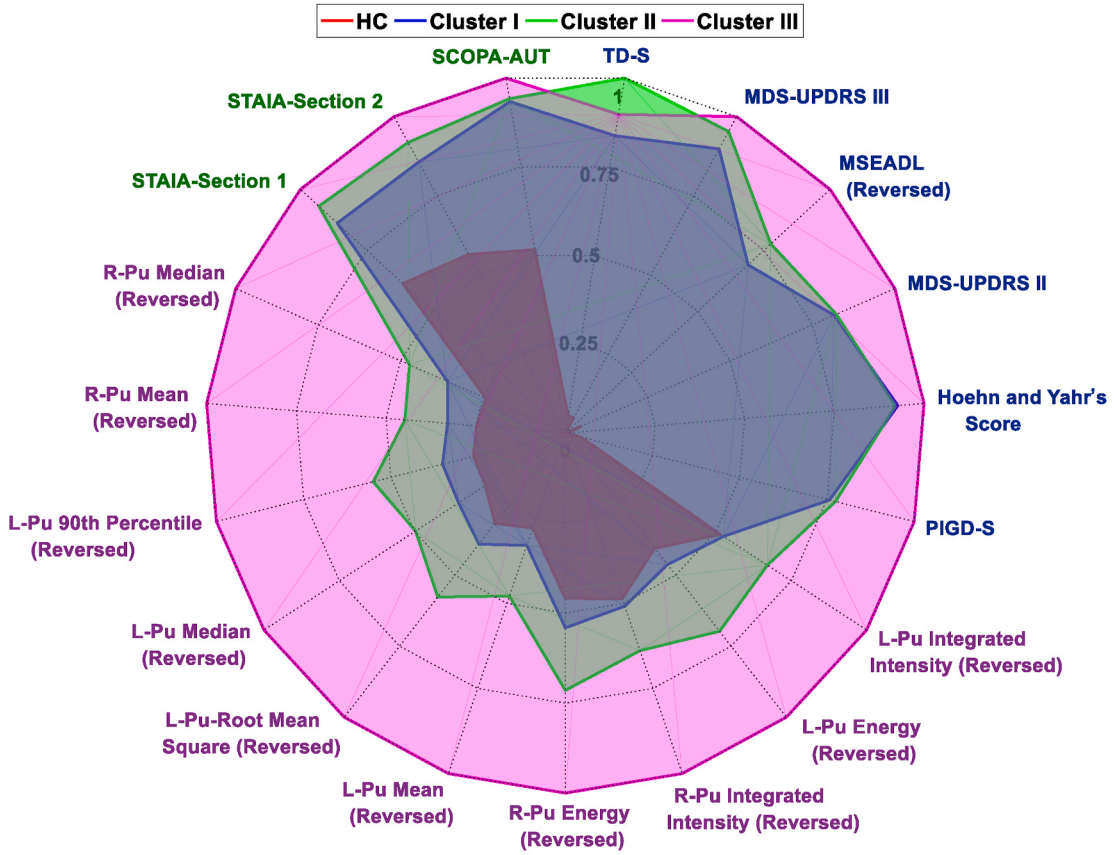

Fig. 6. Radar plot of three kinds of features, motor (dark blue text), imaging (purple text) and non-motor (green text) for our 3 identified PD sub-clusters and HC groups.

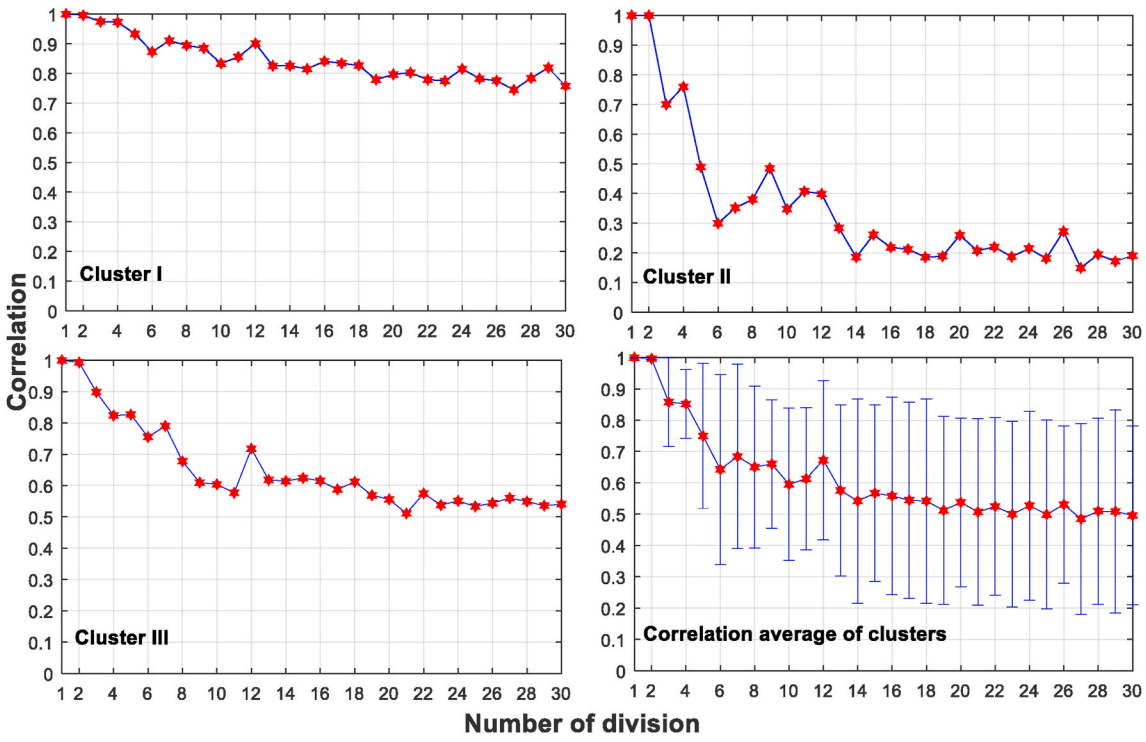

Fig. 7. Reproducibility test based on timeless RFS-M dataset. X axis represents the number of parts the dataset was divided into, and Y axis represents the correlation between specific subtype resulting from each data subset (division) and the original subtype from the entire dataset. The first three plots show the correlations for the three clusters and the fourth one shows the means and standard deviations of the correlations in the three clusters.

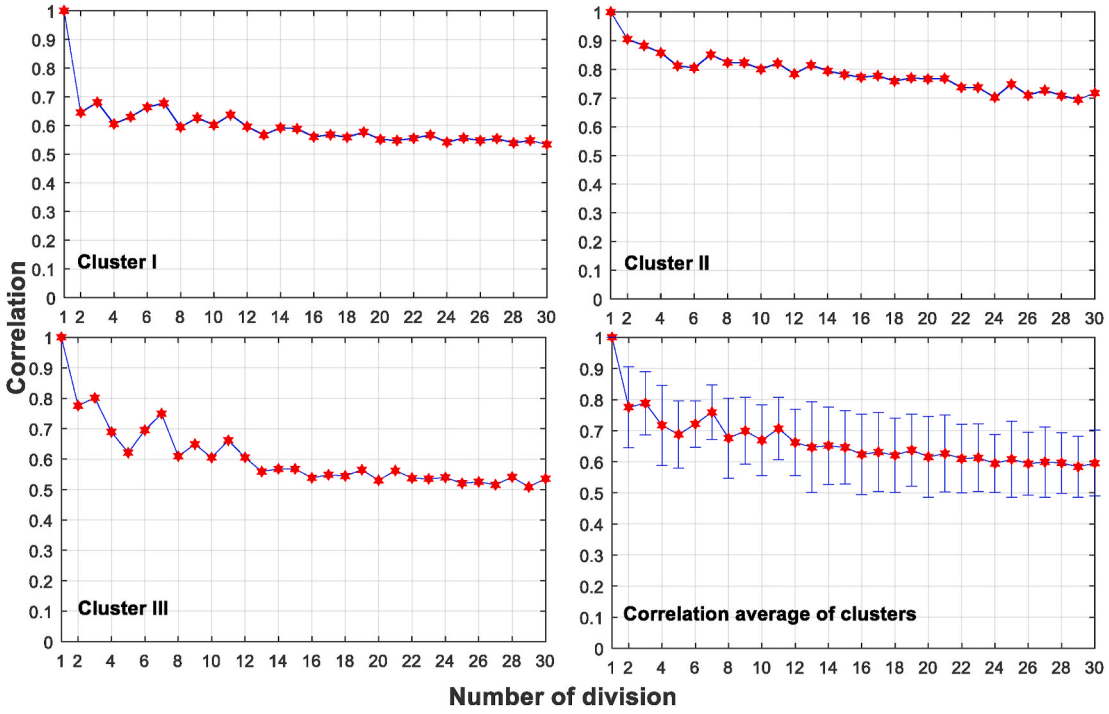

Fig. 8. Reproducibility test based on timeless NCF dataset. X axis represents the number of parts the dataset was divided into, and Y axis represents the correlation between specific subtype resulting from each data subset (division) and the original subtype from the entire dataset. The first three plots show the correlations for the three clusters and the fourth one shows the means and standard deviations of the correlations in the three clusters.

namely cross-sectional and timeless, both of which resulted in robust PD subtypes. Overall, we arrived at 3 distinct subtypes among PD patients using cluster analysis with a comprehensive range of variables including both non-imaging information and radiomics features.

In our work, using solely clustering cost (sum of subject distances in each cluster compared to their cluster centers) did not enable one to select a unique optimal cluster number (see Supplemental Table S3). Thus, we employed additional costs such as AOCF (reproducibility test)

Table 6
Comparison of our findings with other related works. Subtypes (e.g. I) in the different studies do not necessarily correspond to one another but are listed in relative order of severity.

Present work	**Sub-clusters**	**I. Mild**		**II. Moderate**		**III. Severe**	
	MS (U-III)	22.82 ± 10.77		24.22 ± 10.73		25.41 ± 11.96	
	NMS (ST-Sec-2)	31.75 ± 10.05		32.6 ± 9.64		33.72 ± 10.9	
	RMF (L-Pu-En)	(2.56 ± 0.45) e^{11}		(1.75 ± 0.4) e^{11}		(0.70 ± 0.37) e^{11}	
Fereshtenejad et al. [46]	**Sub-clusters**	**I. Mainly motor/slow progression**		**II. Intermediate**		**III. Diffuse/malignant subtype**	
	MS (U-III)	17.7 ± 7.5		24.1 ± 9		27.3 ± 9.4	
	NMS (U–I)	4.4 ± 3.1		6 ± 4		10.8 ± 4.9	
Lui et al. [65]	**Sub-clusters**	**III. Tremor dominant**	**IV. Young onset**	**I. Non-tremor dominant**		**II. Worse disease progression**	
	MS (UIII/DD)	4.82 ± 2.64	5.59 ± 3.39	16.84 ± 4.56		36.57 ± 7.89	
	NMS (MMSE)	27.02 ± 2.66	27.64 ± 2.07	26.79 ± 3.95		26.43 ± 2.76	
Reijnders et al. [63]	**Sub-clusters**	**II. Young onset**		**IV. Tremor-dominant**		**I. Worse disease progression**	**III. Non-tremor-dominant**
	MS (T)	0.6 ± 0.6		1.1 ± 0.8		0.9 ± 0.6	0.7 ± 0.8
	NMS (MMSE)	28.2 ± 1.8		27.7 ± 2.3		24 ± 4.7	17.4 ± 7.7
Lewis et al. [44]	**Sub-clusters**	**I. Younger disease onset**		**II. Tremor dominant**	**IV. Non-tremor dominant**	**III. Rapid disease progression**	
	MS (U-III)	21 ± 10		25 ± 12	28 ± 13	30 ± 11	
	NMS (VF)	46 ± 13		40 ± 12	33 ± 12	44 ± 9	
Gasparoli et al. [49]	**Sub-clusters**	**I. Slow progression**				**II. Rapid progression**	
	MS (UMS)	11.9 ± 3.9				25.1 ± 5.1	
	NMS (D)	9.5 ± 0				12.5 ± 0	
Rooben et al. [62]	**Sub-clusters**	**I. Mildly affected in all domains**	**II. Severe motor complications**	**III. Affected mainly on nondopaminergic domains without prominent motor complications**		**IV. Severely affected on all domains**	
	MS (T)	3.6 ± 1.9	3.1 ± 1.8	4.1 ± 2.2		1.5 ± 3.1	
	NMS (CI)	14.7 ± 5.5	15.8 ± 4.8	19.7 ± 5.4		25.7 ± 5.1	
Post et al. [31]	**Sub-clusters**	**I. Younger onset**		**II. Intermediate older onset**		**III. Oldest onset**	
	MS (UME/DD)	0.7		0.8		1.7	
	NMS (MMSE)	28.7		27.1		26.9	

MS: Motor symptoms, **MCS**: Motor Composition Score, **RMF**: Radiomics features, **U-III**: UPDRS III, **L-Pu_En:** Left Putamen Energy, **ST-Sec-2:** STAIA-section 2, **U–I**: UPDRS I, **UIII/DD**: UPDRS III/disease duration, **MMSE**: Mini-Mental State Examination, **T**: Tremor, **VF**: Verbal Fluency, **UMS**: UPDRS motor score, **D**: Depression, **CI**: Cognitive impairment, **UME/DD**: UPDRS-Motor examination per disease duration.

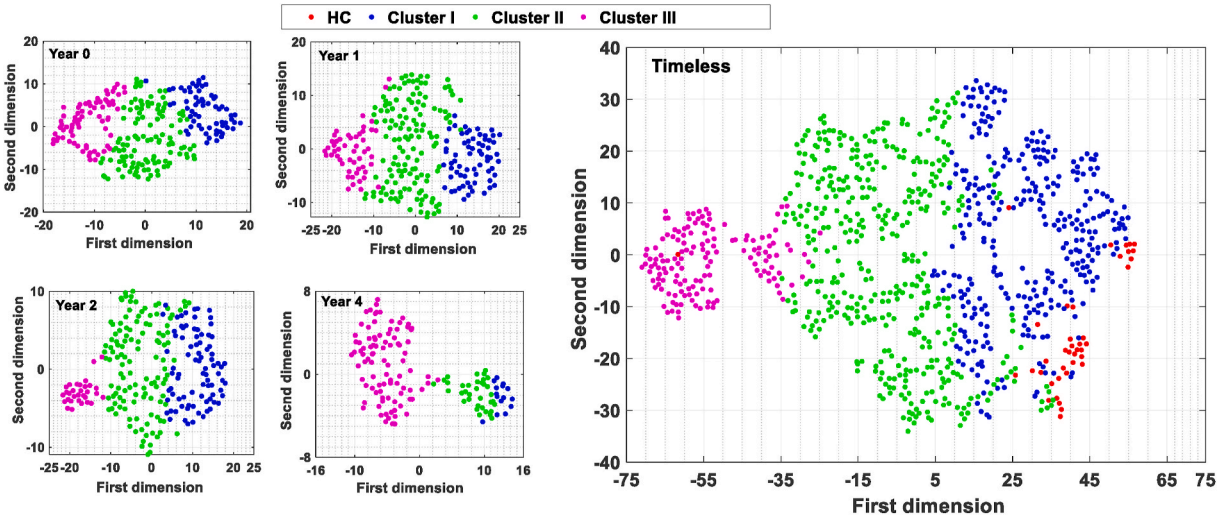

Fig. 9. t_SNE plot for NCF + CIF + RFS-M dataset. Clusters I, II and, III were identified in our ML-based effort, independently validated by this plot. The right figure also includes healthy-control (HC) patients for comparison.

and AOCP (reliability test) for selecting optimal cluster number in each dataset, overall arriving at 3 distinct PD subtypes. Moreover, we showed that the use of multiple HMLSs worked more effectively than the sole usage of clustering algorithms and classifiers (see section III, part A). As a result, the PCA + KMA trajectory was shown as an optimal trajectory among the various trajectories for the generation and analysis of disease subtypes. PCA and KMA were each seen (via ICEM) to result in the most consistent number of identified clusters (as shown in 3 last tables within Supplemental Table S3, resulted from the datasets with RFS-M). To further evaluate our identified subtypes, we performed similarity identification tests across cross-sectional datasets in each category. To identify the similarity between subtypes, first, we utilized the training and testing process of KMA. After training the algorithm by a specific year, we associated subjects of another year with the resulting clusters (testing process). After comparing the associated subjects with the original test sub-groups, we were able to identify similar clusters.

As shown in Supplemental Table S2, the datasets were generated according to the maximum sample size and maximum feature size. Before cross-linking cross-sectional phenotypes, we selected common features between 2 datasets comparing. After re-clustering cross-sectional datasets by common features, we were only enabled to reach similar clusters using datasets with radiomics features compared to when using all features (see Supplemental Table S8). According to Supplemental Table S6 and Supplemental Table S7, it seemed that the subtypes were sensitive to the reduction in clinical features when using no radiomics features.

In recent studies [68,148], it was recommended that researchers should employ methods resulting in reproducible subtypes. As seen in Supplemental Table S1, different studies employing a given dataset with limited features for clustering PD subjects resulted in different optimal numbers. The selections of specific features in clustering can result in a different number of optimal sub-clusters (Supplemental Table S3). One of the most important rules in clustering approaches is the dependency check of sub-groups on the size of features [148]. As seen in the results section, datasets without radiomics features are strongly dependent on the size of clinical features. Unlike this work, to the best of our knowledge, previously published works did not employ radiomics features or even imaging information for PD subtype identification (clustering) [31,44,46–50,61–64].

As shown in Table 3, associations in combinations with RFS-M were more robust compared to combinations with RFS-S (see Supplemental Table S9 and Supplemental Table S10). Thus, we were able to verify 3 distinct subtypes, across the cross-sectional datasets and timeless dataset. Moreover, Hotelling's T2 test re-confirmed our findings (see Fig. 4). As shown in Fig. 9, the distributions of cross-sectional datasets with RFS-M were similar to the timeless dataset. This plot can also be introduced as a powerful independency test among the sub-groups. These plots properly showed clear differences (especially on the first dimension) among the sub-groups.

In a recent study of ours [36], conventional mean uptake analysis showed no correlations with clinical measures in PD subjects in either the caudate and putamen. By contrast, significant correlations were observed for radiomics features, against motor and cognitive measures as well as disease duration. As a result, it can be concluded that RFS-M is potentially very important for the clustering of PD subjects. Moreover, radiomics features of DAT SPECT images, going beyond conventional imaging measures, were seen to provide significant improvements in prediction [21] and diagnosis tasks [37]. Overall, radiomics features have shown considerable potential as biomarkers of PD progression [34,35].

Subsequently, we assessed the dependency of our results on sample size. As shown in Tables 4 and 5, datasets with RMS-M overall resulted in better performance than other datasets shown in Supplemental Table S11-S12; the final reproducibility test confirmed our findings (see Fig. 7). Although the results of NCF + RFS-S shown in Supplemental Table S13- S14 also indicated good performance, usage of this dataset does not enable reaching appropriate cross-linking performance compared to datasets with RMS-M (see Supplemental Table S10). Overall, we believe, this is an important towards development of robust models, not considered in prior works [31,44,46–50,61,63,64].

Reproducing results of previous studies is difficult because the datasets used are different. Furthermore, clustering results did not employ imaging features and may depend on sample size or feature size as discussed already. In recent studies [68,148], analysis of previously published data-driven PD subtype classification systems suggested a lack of reproducibility and need for the establishment of standards for validation and the use of clustering systems. Clinical measurements based on visual examination in different clinical centers have several sources of error and can be a major challenge. A systematic analysis of images may provide help to this end. Thus, RFS-M, e.g. generated through our SERA radiomics software [72] standardized using guidelines from the IBSI (Image Biomarker Standardization Initiative) [73], can be a valuable step towards reproducible research. As shown, cluster results of combinations with RFS-M are more robust to variation in the size of the samples, and feature compared to other results, and the final reproducibility test confirmed our findings (see Fig. 7). Since motor symptoms [16,51–55] as well as non-motor symptoms [51,54,55,150–155] in PD subjects strongly correlate with neuronal dysfunction in the brain, radiomics features have the potential to capture and correlate with attributes of both symptomatic groups [71]. As seen, all combinations including RFS-M, even the use of RFS-M alone, enabled us to reach similar clusters. As a result, it can be seen that radiomic features uniquely determine PD subtypes identified in this work.

Fig. 6 showed the relative progression of disease between the sub-clusters. The subjects in the HC group depicted dramatically lower scores among the PD subtype in all domains, including motor symptoms, non-motor symptoms, and imaging information. Sub-cluster I had lower scores compared to other PD sub-clusters. Thereby, the patients in this sub-cluster appear in a less progressed stage of disease in all domains than those in other PD subtypes. The patients in sub-cluster II indicated worse tremor across all groups, although, other symptoms (except Hoehn and Yahr's score) were bigger than those in sub-cluster I, lower than those in sub-cluster III. Patients in sub-cluster III had the worst symptoms (except tremor) among PD sub-clusters. The patients in this sub-cluster had the worst neuronal loss in the substantia nigra with the loss of dopaminergic terminals in the basal ganglia [156,157], whilst, motor and non-motor symptoms had lower progression than those in other sub-clusters. In short, our 3 identified subtypes are 1) mild, 2) intermediate and 3) severe, especially in terms of dopaminergic deficit (imaging), but also with motor and non-motor manifestations.

Differences in inclusion criteria from datasets, in feature selection, and in methodology between cluster analysis studies have made it difficult to compare the subtypes [148]. We note that a recent study indicated that tremor is not an independent indicator of a benign disease course, thus we did not emphasize it for defining the name of sub-clusters [158]. As shown in Supplemental Table S1, Hoehn, and Yahr (YH) [56] introduced disease severity into 5 stages. Since then, stage 0 has been added, and stages 1.5 and 2.5 have been proposed and are widely used [56,57]. Overall, patients with higher H&Y scores have been reported to have a poorer quality of life [159]. Based on Table 6, we provide a relative comparison of sub-clusters obtained in different efforts. These comparisons were performed based on the relative progression of disease between the sub-clusters. Fereshtenejad et al. [46] found three sub-clusters. The relative progression of disease among their sub-clusters I, II and, III appears somewhat related to ours in sub-clusters I, II and, III. Post et al. [31] also defined 3 sub-clusters. By comparison, Lui et al. [65], Reijnders et al. [63], Lewis et al. [44] and, Rooben et al. [62] identified 4 sub-clusters, and we have attempted to roughly map their clusters to relevant columns, indicating the relative progression of the disease. Finally, Gasparoli et al. [49] found only two sub-clusters, which we also summarize. At the same time, as emphasized above, these various sub-clusters are likely not similar (e.g. as imaging was not

utilized, and different methods and patient data were utilized).

The limited size of a dataset is a limiting factor in PD subtype identification; as such, to maximize our numbers as well as to maximize our features, we had to select different sets. Another approach we used to tackle this limitation was the usage of timeless datasets. In our work, we used dimensionality reduction algorithms to reduce the size of features to avoid overfitting, although it was possible to utilize feature selection algorithms that we hope to explore in future work. Furthermore, our study is for a particular cohort of patients, and our findings should be externally validated on an independent database.

At the same time, there are many advantages to this study. Importantly, we utilized a very comprehensive database including a broad spectrum of non-imaging and imaging features. Furthermore, our study focused on the added value of imaging, including the generation of radiomics features, in a standardized manner, based on guidelines from the IBSI, aiming towards the identification of a robust PD subtype.

Furthermore, there is evidence that a cross-sectional only analysis cannot accurately reveal the progression of the disease [160,161], and the strength of our study is additional the usage of the timeless dataset formed from longitudinal data, enabling a more balanced definition of subtype. In the future, we aim to investigate optimal feature selection and machine learning methods for both tasks of i) clustering of PD subjects and ii) predicting PD outcome (e.g. disease subtype in year 4). Moreover, we aim to identify distinct disease progression pathways in Parkinson's disease (PD), utilizing clustering of time-series data (longitudinal clustering), and relating disease progression pathways to PD subtypes.

5. Conclusion

We aimed to perform robust identification of PD subtypes, incorporating clinical and imaging data. We utilized hybrid machine learning systems, with a comprehensive analysis of the trajectories, for optimal selection of clusters, and extensive analysis of robustness. In particular, the usage of radiomics features enabled the identification of clusters that were more robust to variations in features and samples. Based on cross-sectional as well as timeless datasets, including cross-linkage between them, three PD subtypes were identified: 1) mild, 2) intermediate and 3) severe, especially in terms of dopaminergic deficit (imaging), but also with motor and non-motor manifestations. Furthermore, clusters generated using SPECT-based segmentation remained less consistent when the number of subjects changed (reproducibility test), while the use of MRI to segment SPECT images enabled more robust identification of PD subtypes. Overall, we conclude that utilization of PCA + KMA trajectory and datasets including RFS-M can produce reproducible subtypes in PD patients.

Data and code availability

All codes (included clustering algorithms, classification algorithms, and dimensionality reduction Algorithms, etc.) and all datasets publicly exist at:

https://github.com/MohammadRSalmanpour/Unsupervised_study

Declaration of competing interest

The authors have no relevant conflicts of interest to disclose.

Acknowledgements

The project was supported by the Michael J. Fox Foundation, including the 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2020.104142.

References

[1] T. Simuni, K. Sethi, Nonmotor manifestations of Parkinson's disease, Ann. Neurol. 64 (2008) 65–80.
[2] E. Wolter, Non-motor extranigral signs and symptoms in Parkinson's disease, Park. Relat. Disord. 3 (2009) 6–12.
[3] K. Bayulkem, G. Lopez, Non-motor fluctuations in Parkinson's disease: clinical spectrum and classification, J. Neurol. Sci. 289 (2010) 89–92.
[4] D. Dai, Y. Wang, et al., Polymorphisms of DRD2 and DRD3 genes and Parkinson's disease: a meta-analysis, Biomedica 2 (2) (2014) 275–281.
[5] T. Lebouvier, T. Chaumette, et al., The second brain and Parkinson's disease, EGN 30 (5) (2009) 735–741.
[6] P. McNamara, E. Stavitsky, et al., Side of motor symptom onset and pain complaints in Parkinson's disease, Int. J. Geriatr. Psychiatr. 25 (2010) 519–524.
[7] P. Bell, M. Gilat, et al., Dopaminergic basis for impairments in functional connectivity across subdivisions of the striatum in Parkinson's disease, Mum Brain Mapp 36 (4) (2015) 1278–1291.
[8] N. Bohnen, R. Albin, The cholinergic system and Parkinson disease, Behav. Brain Res. 2 (221) (2011) 564–573.
[9] S. Maril, S. Hassin-Baer, et al., Effects of asymmetric dopamine depletion on sensitivity to rewarding and aversive stimuli in Parkinson's disease, Neuropsychologia 51 (5) (2013) 818–824.
[10] M. Ventura, K. Baynes, et al., Hemispheric asymmetries and prosodic emotion recognition deficits in Parkinson's disease, Neuropsychologia 50 (8) (2012) 1936–1945.
[11] E. Cubo, P. Martin, et al., Motor laterality asymmetry and nonmotor symptoms in Parkinson's disease, Mov. Disord. 25 (1) (2010) 70–75.
[12] J.G. GabrielHou, E.C. Lai, NON-MOTOR symptoms OF Parkinson's disease, Int. J. Gerontol. 2 (1) (2007) 53–64.
[13] E. Modestino, A. Reinhofer, et al., Hoehn and Yahr staging of Parkinson's disease in relation to neuropsychological measures, Front. Biosci. 23 (7) (2018) 1370–1379.
[14] A. Antonini, I. Isaias, et al., Duodenal levodopa infusion for advanced Parkinson's disease: 12-month treatment outcome, International Parkinson and Movement Disorder Society 22 (8) (2007) 1145–1149.
[15] M. Tippmann-Peikert, G. Park, et al., Pathologic gambling in patients with restless legs syndrome treated with dopaminergic agonists legs syndrome treated with dopaminergic agonists, Neurology 68 (2007) 301–303.
[16] M. Savitt, V. Dawson, et al., Diagnosis and treatment of Parkinson disease: molecules to medicine, JCI 7 (2006) 1744–1754.
[17] A. Lang, A. Lozano, Parkinson's disease. Second of two parts, N. Engl. J. Med. (1998) 1130–1143.
[18] M. Hely, J. Morris, et al., The sydny multicentre stuly of Parkinson's Disease : progression and mortality at 10 years, J Neurol Neurosurg Psychiarty 67 (1999) 300–3007.
[19] A. Nieuwboer, W.D. Weerdt, B. René Dom, Prediction of outcome of physiotherapy in advanced Parkinson's disease, SAGE Journals 16 (8) (2002) 886–893.
[20] S. Grill, J. Weuve, M. Weisskopf, Predicting outcomes in Parkinson's disease: comparison of simple motor performance measures and the Unified Parkinson's Disease Rating Scale-III, J. Parkinsons Dis. 3 (2011) 287–298.
[21] A. Rahmim, P. Huang, et al., Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images, Neuroimage: Clinical 16 (2017) 539–544.
[22] D. Arnaldi, F. De Carli, et al., Prediction of cognitive worsening in de novo Parkinson's disease: clinical use of biomarkers, Mov. Disord. 32 (2017) 1738–1747.
[23] I. Fyfe, Prediction of cognitive decline in PD, Nat. Rev. Neurol. 14 (2018) 213–317.
[24] C. Gao, H. Sun, et al., Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease, Sci. Rep. 8 (1) (2018) 1–21.
[25] M. Salmanpour, M. Shamsaei, et al., Machine learning methods for optimal prediction of outcome in Parkinson's disease, in: IEEE Nucl. Sci. Symp. Conf. Record, Sydney, 2018.
[26] M. Salmanpour, M. Shamsaei, e. al, Machine learning methods for optimal prediction of motor outcome in Parkinson's disease, Phys. Med. 69 (2020) 233–240.
[27] M. salmanpour, M. Shamsaei, e. al, Optimized machine learning methods for prediction of cognitive outcome in Parkinson's disease, Comput. Biol. Med. 111 (2019) 1–8.
[28] K. Leung, M. Salmanpour, et al., Using deep-learning to predict outcome of patients with Parkinson's disease, in: IEEE Nucl. Sci. Symp. Conf. Record, Sydney, 2018.
[29] G. Ramani, G. Sivagami, et al., Feature relevance analysis and classification of Parkinson's disease TeleMonitoring data through data mining", Int. J. Adv. Res. Comput. Sci. Software Eng. 2 (3) (2012) 298–304.
[30] M. Nilashi, O. Ibrahim, A. Ahani, Accuracy Improvement for Predicting Parkinson's Disease Progression, Scientific Reports, 2016.

[31] B. Post, J. Speelman, et al., Clinical heterogeneity in newly diagnosed Parkinson's disease, J. Neurol. 255 (5) (2008) 716–722.
[32] S. Parashos, S. Luo, et al., Measuring disease progression in early Parkinson disease: the national institutes of health exploratory trials in Parkinson disease (NET-PD) experience, JAMA Neurol 6 (71) (2014) 710–716.
[33] A. Noyce, a. Schrag, et al., Subtle motor disturbances in PREDICT-PD participants, J. Neurol. Neurosurg. Psychiatry 88 (2017) 212–217.
[34] P. Lui, H. Wang, et al., Parkinson's disease diagnosis using neostriatum radiomic features based on T2-weighted magnetic resonance imaging, Front. Neurol. 11 (2020) 1–9.
[35] J. Tang, B. Yang, et al., Artificial neural network based outcome prediction in DAT SPECT imaging of Parkinson's Disease, J. Nucl. Med. 58 (supplement 1) (2017), 292-292.
[36] A. Rahmim, Y. Salimpour, et al., Application of texture analysis to DAT SPECT imaging: relationship to clinical assessments, Neuroimage: Clinical 12 (2016) e1–e9.
[37] Z. Cheng, J. Zhang, et al., Radiomic features of the nigrosome region of the substantia nigra: using quantitative susceptibility mapping to assist the diagnosis of idiopathic Parkinson's disease, Front. Aging Neurosci. 11 (2019) 1–11.
[38] R. Erro, S. Schneide, et al., What do patients with scans without evidence of dopaminergic deficit (SWEDD) have? New evidence and continuing controversies, J. Neurol. Neurosurg. Psychiatr. 87 (3) (2015) 319–323.
[39] A. Marco, T. Utiumi, et al., Dopamine transporter imaging in clinically unclear cases of parkinsonism and the importance of Scans without Evidence of Dopaminergic Deficit (SWEDDs), Arq. Neuro. Psiquiatr. 70 (9) (2010) 667–673.
[40] Y. Wu, H. Jiang, et al., Use of radiomic features and support vector machine to distinguish Parkinson's disease cases from normal controls, Ann. Transl. Med. 7 (23) (2019) 1–16.
[41] P. Huang, N. Shenkov, et al., Radiomics analysis of longitudinal DaTscan images for improved progression tracking in Parkinson's disease, in: Journal of Nuclear Medicine, 2017.
[42] T. Foltynie, C. Brayne, R. Barker, The heterogeneity of idiopathic Parkinson's disease, J. Neurol. 2 (249) (2002) 138–145.
[43] D. Berg, R. Postuma, e. al, Ime to redefine PD? Introductory statement of the MDS task force on the definition of Parkinson's disease, Mov. Disord. 29 (2014) 454–462.
[44] S. Lewis, T. Foltynie, et al., Heterogeneity of Parkinson's disease in the early clinical stages using a datan driven approach, J. Neurol. Neurosurg. Psychiatry 76 (3) (2005) 343–348.
[45] C. Eggers, D. Pedrosa, e. al, Parkinson subtypes progress differently in clinical course and imaging pattern, PloS One 7 (2012), e46813.
[46] S. Fereshtehnejad, Y. Zeighami, e. al, Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression, Brain 140 (7) (2017) 1959–1976.
[47] J. Graham, H. Sagar, A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: identification of three distinct subtypes, Mov. Disord. 14 (1) (1999) 10–20.
[48] R. Erro, C. Vitale, et al., The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients, PloS One 8 (8) (2013) 1–8.
[49] E. Gasparoli, D. Delibori, et al., Clinical predictors in Parkinson's, Neurol. Sci. 23 (2002) s77–s78.
[50] M. Poletti, D. Frosini, et al., The association between motor subtypes and alexithymia in de novo Parkinson's disease, J. Neurol. 258 (6) (2010) 1042–1045.
[51] H. Sitte, C. Pifl, et al., Dopamine and noradrenaline, but not serotonin, in the the human claustrum are greatly reduced in patients with Parkinson's disease: possible functional implications, Eur. J. Neurosci. 45 (2017) 192–197.
[52] G. Hodge, L. Butcher, Pars compacta of the substantia nigra modulates motor activity but is not involved importantly in regulating food and water intake, Naunyn-Schmiedeberg's Arch. Pharmacol. 313 (1) (1980) 51–67.
[53] S. Greffard, M. Verny, et al., Motor score of the Unified Parkinson Disease Rating Scale as a good predictor of Lewy body-associated neuronal loss in the substantia nigra, Arch. Neurol. 63 (4) (2006) 584–588.
[54] R. Nandhagopal, M. McKeown, A. Stoessl, Functional imaging in Parkinson disease, Neurology 15 (2008) 1478–1488.
[55] G. Grosch, J. Winkler, Z. Kohl, Early degeneration of both dopaminergic and serotonergic axons – a common mechanism in Parkinson's disease, Front. Cell. Neurosci. 10 (2016) 1–8.
[56] M. Hoehn, M. Yahr, Parkinsonism: onset, progression and mortality, Neurology 5 (1967) 427–442.
[57] V. Venkatesh, R. Swarnalatha, Review and assessment OF the rating scales OF Parkinson's disease, Int. J. Appl. Sci. 5 (2018) 1–7.
[58] S. Happe, P. Baier, Association of daytime sleepiness with nigrostriatal dopaminergic degeneration in early Parkinson's disease, J. Neurol. 254 (8) (2007) 1037–1043.
[59] T. Wakamori, T. Agari, et al., Cognitive functions in Parkinson's disease: relation to disease severity and hallucination, Park. Relat. Disord. 20 (4) (2014) 415–420.
[60] K. Szewczyk-Krolikowski, P. Tomlinson, et al., The influence of age and gender on motor and non-motor features of early Parkinson's disease: initial findings from the Oxford Parkinson Disease Center (OPDC) discovery cohort, Park. Relat. Disord. 20 (1) (2014) 99–105.
[61] A. Schrag, N. Quinn, B. Ben-Shlomo, Heterogeneity of Parkinson's disease, J. Neurol. Neurosurg. Psychiatry 77 (2006) 275–284.
[62] S. van Rooden, F. Colas, et al., Clinical subtypes of Parkinson's disease, Mov. Disord. 27 (8) (2011) 996–1003.
[63] J. Reijnders, U. Ehrt, e. al, The association between motor subtypes and psychopathology in Parkinson's disease, Park. Relat. Disord. 15 (5) (2009) 379–382.
[64] M. Lawton, Y. Shlomo, e. al, Developing and validating Parkinson's disease subtypes and their motor and cognitive progression, J. Neurol. Neurosurg. Psychiatry 89 (12) (2018) 1279–1288.
[65] P. Liu, T. Feng, et al., Clinical heterogeneity in patients with early-stage Parkinson's disease: a cluster analysis, Biomedicine & Biotechnology) 12 (9) (2011) 694–703.
[66] M. Lawton, F. Baigb, et al., Parkinson's disease subtypes in the oxford Parkinson disease centre (OPDC) discovery cohort, J. Parkinsons Dis. 5 (2015) 269–279.
[67] J. Mu, K. Chaudhuri, et al., Parkinson's disease subtypes identified from cluster Analysis of motor and non-motor symptoms, Front. Aging Neurosci. 9 (2017) 1–10.
[68] T. Mestre, C. Eberly, et al., Reproducibility of data-driven Parkinson's disease subtypes for clinical research, Park. Relat. Disord. 56 (2018) 102–106.
[69] Y. Singh, P.K. Bhatia, O. Sangwan, A review OF studies ON machine learning techniques, Int. J. Comput. Sci. Secur. 1 (1) (2007) 70–84.
[70] X. Wu, H. Cai, et al., Recent progress of imaging agents for Parkinson's disease, Curr. Neuropharmacol. 12 (6) (2014) 551–563.
[71] D. Brooks, Assessment of Parkinson's disease with imaging, Park. Relat. Disord. 13 (2007) S268–S275.
[72] S. Ashrafinia, Quantitative Nuclear Medicine Imaging Using Advanced Image Reconstruction and Radiomics, Ph.D. Dissertation, Johns Hopkins University, 2019.
[73] A. Zwanenburg, S. Leger, et al., Image Biomarker Standardisation Initiative, 2016 arXiv preprint arXiv:1612.07003.
[74] A. Zwanenburg, M. Vallières, et al., The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping, Radiology 295 (2) (2020) 328–338.
[75] M. McNitt-Gray, S. Napel, et al., Standardization in quantitative imaging: a multi-center comparison of radiomics features from different software packages on digital reference objects and patient datasets, Tomography 6 (2020) 118–128.
[76] S. Ashrafinia, D. Pejman, et al., Application of texture and radiomics analysis to clinical myocardial perfusion SPECT imaging, J. Nucl. Med. 59 (Supplement 1) (2018), 94-94.
[77] I. Klyuzhin, M. Gonzalez, et al., Exploring the use of shape and texture descriptors of positron emission tomography tracer distribution in imaging studies of neurodegenerative disease, J. Cerebr. Blood Flow Metabol. 36 (6) (2016) 1122–1134.
[78] Parkinson Progression Marker Initiative, The Parkinson progression marker initiative (PPMI), Prog. Neurobiol. 95 (2011) 629–635.
[79] B. Mwangi, T. Siva Tian, J. Soares, A review of feature reduction techniques in neuroimaging, Neuroinformatics 12 (2014) 229–244.
[80] M. Reif, F. Shafait, Efficient feature size reduction via predictive forward selection, Pattern Recogn. 47 (2014) 1664–1673.
[81] M. Patel, A. Khalaf, H. Aizenstein, Studying depression using imaging and machine learning methods, Neuroimage: Clinical 10 (2016) 115–123.
[82] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1–3) (1987) 37–52.
[83] Z. Wenming, Z. Cairong, Z. Li, An improved algorithm for kernel principal component analysis, Neural Process. Lett. 22 (1) (2005) 49–56.
[84] L. Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.
[85] H. Tinsley, D. Tinsley, Uses of factor analysis in counseling psychology research, J. Counsel. Psychol. 34 (4) (1987) 414–424.
[86] J.W. Sammon, A nonlinear mapping for data structure analysis, IEEE Trans. Comput. 5 (May 1969).
[87] J. Sun, M. Crowe, C. Fyfe, Extending Sammon Mapping with Bregman Divergences, vol. 187, Elsevier, March 2012.
[88] J.B. Tenenbaum, V.d. Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[89] H. Shi, B. Yin, et al., Robust L-isomap with a novel landmark selection method, Hindawi Mathematical Problems in Engineering (2017) 1–12, 2017.
[90] M. Belkin, P. Niyogi, Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering, 2001, pp. 586–691.
[91] M. Lewandowski, J.M.-d. Rincon, e. al, Temporal Extension of Laplacian Eigenmaps for Unsupervised Dimensionality Reduction of Time Series, ICPR, 2010.
[92] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[93] A. Mead, Review of the development of multidimensional scaling methods, Journal of the Royal Statistical Society. Series D (The Statistician) 41 (1) (1992) 27–39.
[94] R. Coifman, S. Lafon, et al., Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps, Proc. Natl. Acad. Sci. U. S. A. 102 (21) (2005) 7426–7431.
[95] R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (1) (2006) 5–30.
[96] D. Agrafiotis, Stochastic proximity embedding, J. Comput. Chem. 24 (10) (2003) 1215–1221.
[97] P. Li, S. Chen, A review on Gaussian process latent variable models, CAAI Transactions on Intelligence Technology 1 (4) (2016) 366–376.
[98] N. Lawrence, Learning for larger datasets with the Gaussian process latent variable model, J. Mach. Learn. Res. 2 (2007) 243–250.

[99] G. Hinton, S. Roweis, Stochastic neighbor embedding, Adv. Neural Inf. Process. Syst. 15 (2003) 857–864.
[100] K. Nam, H. Je, S. Choi, Fast stochastic neighbor embedding: a trust-region algorithm, in: IEEE International Joint Conference on Neural Networks, Budapest, 2004.
[101] G. Hinton, R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[102] A. Jain, R. Duin, J. Mao, Statistical pattern recognition: a review, IEEE 22 (1) (2000) 4–37.
[103] M. Jain, M. Murty, P. Flynn, Data clustering: a review, ACM Computing 31 (3) (1999) 264–323.
[104] M. Rodriguez1, C. Comin, et al., Clustering algorithms: a comparative approach, PloS One 14 (1) (2019) 1–34.
[105] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Transactions on Neural Networks and Learning Systems 11 (3) (2000) 586–600.
[106] D. Frey, D. Dueck, Clustering by passing messages between data points, Science 315 (5814) (2007) 972–976.
[107] F. Murtagh, P. Legendre, Ward's hierarchical agglomerative clustering method: J. Classif. 31 (2014) 274–295.
[108] D. Krznaric, C. Levcopoulos, Optimal algorithms for complete linkage clustering in d dimensions, Theor. Comput. Sci. 286 (2002) 139–149.
[109] S. Michener, "A Statistical Method for Evaluating Systematic Relationships," University of Kansas Science Bulletin, vol. vol. 38, pp. 1409-1438, 38..
[110] T. Kanungo, D. Mount, et al., An efficient k-means clustering algorithm: analysis and implementation, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 881–892.
[111] O. Oyelade, O. Oladipupo, I. Obagbuwa, Application of k-means clustering algorithm for prediction of students' academic performance, Int. J. Comput. Sci. Inf. Secur. 7 (1) (2010) 292–295.
[112] P. Arora, D. Dr, S. Varshney, Analysis of K-means and K-medoids algorithm for big data, Procedia Computer Science 28 (2016) 507–512.
[113] A. Bhat, K-MEDOIDS clustering using partitioning around medoids for performing face recognition, International Journal of Soft Computing, Mathematics and Control 3 (3) (2014) 1–12.
[114] H. Sahbi, A particular Gaussian mixture model for clustering and its application to image retrieval, Soft Computing 12 (2008) 667–676.
[115] M. RodneyOD, P. Goodman, Decision tree design using information theory, Knowl. Acquis. 2 (1990) 1–19.
[116] S. Chourasia, Survey paper on improved methods of ID3 decision tree, International Journal of Scientific and Research Publications 3 (12) (2013) 1–2.
[117] D. Denison, C. Holmes, et al., Bayesian Methods for Nonlinear Classification and Regression, John Wiley and Sons, New York, 2002.
[118] C. Chung Chang, C. Jen Lin, LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (3) (2011) 1–27.
[119] C. Cortes, V. Vapnik, Support-Vector networks, Mach. Learn. 20 (1995) 273–297.
[120] K. Crammer, Y. Singer, On the algorithmic implementation of multiclass kernel-based vector machines, J. Mach. Learn 2 (2001) 265–292.
[121] D. Berrar, Bayes' theorem and naive Bayes classifier, Encyclopedia of Bioinformatics and Computational Biology 1 (2019) 403–412.
[122] S. Taheri, M. Mammadov, Learning the naive Bayes classifier with optimization models, Int. J. Appl. Math. Comput. Sci. 23 (4) (2013) 787–795.
[123] N. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Statistician 46 (3) (1992) 175–185.
[124] N. Suguna, K. Thanushkodi, An improved k-nearest neighbor classification using genetic algorithm, IJCSI International Journal of Computer Science Issues 7 (4) (2010) 18–21.
[125] J. Talbot, B. Lee, et al., EnsembleMatrix: interactive visualization to support machine learning with multiple classifiers, in: CHI '09: CHI Conference on Human Factors in Computing Systems, Boston MA, New York, 2009.
[126] J. Shan, H. Zhang, et al., Online active learning ensemble framework, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS 30 (2) (2018) 486–498.
[127] Q. Lu, X. Qiao, Sparse Fisher's linear discriminant analysis for partially labeled data, Stat. Anal. Data Min. 11 (1) (2018) 17–31.
[128] G. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, John Wiley & Sons, New Jersey, 1992.
[129] M. Kusy, R. Zajdel, Probabilistic neural network training procedure based on Q (0)-learning algorithm in medical data classification, Appl. Intell. 41 (2014) 837–854.
[130] D. Specht, Probabilistic neural networks and the polynomial adaline as complementary techniques for classification, IEEE Trans. Neural Network. 1 (1) (1990) 111–121.
[131] H. Joutsijoki, M. Haponen, et al., Error-correcting Output codes in classification of of human induced pluripotent stem cell colony images, BioMed Res. Int. 2016 (2016) 1–14.
[132] R. Hamming, Error detecting and error correcting codes, Bell System Technical Journal 29 (3) (1950) 147–160.
[133] S. Alsmadi, M. khalil, et al., Back propagation algorithm: the best algorithm, IJCSNS International Journal of Computer Science and Network Security 9 (4) (2009) 378–383.
[134] D. Rumelhart, E. Geoffrey, et al., Leaner Representations by back-Propagating errors, Nature 323 (9) (1986) 533–536.
[135] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[136] A. Jehad, R. Khan, N. Ahmad, Random forests and decision Trees, IJCSI International Journal of Computer Science Issues 9 (5) (2012) 272–278.
[137] A. Townley, M. Ilchmann, et al., Existence and learning of oscillations in recurrent neural networks, IEEE Trans. Neural Network. 11 (1) (2000) 205–214.
[138] N. Maknickiene, V. Rutkauskas, et al., Investigation of financial market prediction by recurrent neural network, Innovative Infotechnologies for Science, Business and Education 11 (2) (2011) 3–8.
[139] Y. Arora, A. Singhal, A. Bansal, A study of applications of RBF network, Int. J. Comput. Appl. 94 (2) (2014) 17–20.
[140] O. Nelles, A. Fink, R. Isermann, Local linear model Trees (LOLIMOT) toolbox for nonlinear system identification, science Direct (IFAC System Identification) 33 (15) (2000) 845–850.
[141] J. Martínez-Morales, E. Palacios, Modeling of internal combustion engine emissions by, SciVerse Science Direct 3 (2012) 251–258.
[142] R. Rossi, Mathematical Statistics : an Introduction to Likelihood Based Inference, John Wiley & Sons, New York, 2018.
[143] I. Myung, Tutorial on maximum likelihood estimation, J. Math. Psychol. 47 (1) (2003) 90–100.
[144] S. Zenzo, R. Bernstein, et al., Gaussian maximum likelihood and contextual classification algorithms for multicrop classification, IEEE Trans. Geosci. Rem. Sens. GE-25 (6) (1987) 805–814.
[145] K. Wang, B. Wang, L. Peng, CVAP: validation for cluster Analyses, Data Sci. J. 8 (2009) 88–93.
[146] R. Sharan, A. Maron-Katz, R. Shamir, CLICK and EXPANDER: a system for clustering and visualizing gene expression data, Bioinformatics 19 (14) (2003) 1787–1799.
[147] P. Secchi, A. Stamm, S. Vantini, Inference for the mean of large p small n data: a finite-sample high-dimensional generalization of Hotelling's theorem, Electronic Journal of Statistics 7 (2013) 2005–2031.
[148] E. Qian, Y. Huang, Subtyping of Parkinson's disease - where are we up to? Aging and Disease 10 (5) (2019) 1130–1139.
[149] C. Marras, P. Rochom, A.E. Lang, Predicting motor decline and disability in Parkinson disease, Arch. Neurol. 59 (2002) 1724–1728.
[150] S. Wang, R. Liu, et al., Technetium-99m hexamethylpropylene amine oxime single photon emission tomography of the brain in early Parkinson's disease: correlation with dementia and lateralization, Eur. J. Nucl. Med. 20 (4) (1993) 339–344.
[151] R. Lui, K. Lin, et al., Cognition and 99Tcm-HMPAO SPECT in Parkin's Disease, Nucl. Med. Commun. 13 (1992) 744–748.
[152] R. Marie, L. Barre, et al., Relationships between striatal dopamine denervation and frontal executive tests in Parkinson's disease, Neurosci. Lett. 260 (1999) 77–80.
[153] J. Rinne, J. Rummukainen, et al., Dementia in parhson's disease is related to neuronal loss in the medial substantia nigra, Ann. Neurol. 26 (1) (1988) 47–50.
[154] A. Owen, M. James, et al., FRONTO-STRIATAL cognitive deficits at different stages OF Parkinson's disease, Brain 115 (1992) 1727–1751.
[155] C. Marras, K. Chaudhuri, Nonmotor features of Parkinson's disease subtypes, Mov. Disord. 31 (8) (2016) 1095–1102.
[156] W. Paulus, K. Jellinger, The neuropathologic basis of different clinical subgroups of Parkinson's disease, J. Neuropathol. Exp. Neurol. 50 (6) (1991) 743–755.
[157] K. Jellinger, Recent developments in the pathology of Parkinson's disease, J. Neural. Transm. Suppl. 62 (2002) 347–376.
[158] M. Selikhova, D. Williams, et al., A clinico-pathological study of subtypes in Parkinson's disease, Brain 132 (2009) 2947–2957.
[159] S. Rahman, H. Griffin, et al., Quality of life in Parkinson's disease: the relative importance of the symptoms, Mov. Disord. 10 (23) (2008) 1428–1434.
[160] X. Zhang, J. Chou, et al., Data-driven subtyping of Parkinson's disease using longitudinal clinical records: a cohort study, Sci. Rep. 9 (1) (2019) 1–12.
[161] J. Marinus, J. van der Heeden, J. van Hilten, Calculating clinical progression rates in Parkinson's disease: methods matter, Park. Relat. Disord. 20 (2014) 1263–1267.