Automated Categorization of Parkinsonian Syndromes Using Magnetic Resonance Imaging in a Clinical Setting

Lydia Chougar, Johann Faouzi, Nadya Pyatigorskaya, Lydia Yahia-cherif, Rahul Gaurav, Emma Biondetti, Marie Villotte, Romain Valabregue, Jean-christophe Corvol, Alexis Brice, et al.

Automated Categorization of Parkinsonian Syndromes Using Magnetic Resonance Imaging in a Clinical Setting

Lydia Chougar, MD,[1,2,3,4*] Johann Faouzi, PhD,[1,5] Nadya Pyatigorskaya, MD, PhD,[1,2,3,4] Lydia Yahia-Cherif, PhD,[1,3] Rahul Gaurav,[1,2,3] Emma Biondetti, PhD,[1,2,3] Marie Villotte, MD,[6] Romain Valabrègue, PhD,[1,3] Jean-Christophe Corvol, MD, PhD,[1,7] Alexis Brice, MD, PhD,[1,8] Louise-Laure Mariani, MD, PhD,[1,7,9] Florence Cormier, MD,[9] Marie Vidailhet, MD, PhD,[1,2,9] Gwendoline Dupont, MD,[10,11] Ines Piot, MD,[4] David Grabli, MD, PhD,[9] Christine Payan, MD,[12,13] Olivier Colliot, PhD,[1,5] Bertrand Degos, MD, PhD,[14,15] and Stéphane Lehéricy, MD, PhD[1,2,3,4]

[1]*Paris Brain Institute–ICM, INSERM U 1127, CNRS UMR 7225, Sorbonne Université, UMR S 1127, CNRS UMR 7225, Paris, France*
[2]*ICM, "Movement Investigations and Therapeutics" Team (MOV'IT), Paris, France*
[3]*ICM, Centre de NeuroImagerie de Recherche–CENIR, Paris, France*
[4]*Department of Neuroradiology, Pitié-Salpêtrière University Hospital, APHP, Paris, France*
[5]*INRIA, Aramis Team, Paris, France*
[6]*Faculté de Médecine, Université Denis Diderot, Paris, France*
[7]*ICM, Centre d'Investigation Clinique Neurosciences, Paris, France*
[8]*ICM, Team Neurogénétique Fondamentale et Translationnelle, Paris, France*
[9]*Clinique des Mouvements Anormaux, Département des Maladies du Système Nerveux, Hôpital Pitié-Salpêtrière, APHP, Paris, France*
[10]*Université de Bourgogne, Dijon, France*
[11]*Centre Hospitalier Universitaire François Mitterrand, Département de Neurologie, Dijon, France*
[12]*BESPIM, Hôpital Universitaire de Nîmes, Nîmes, France*
[13]*Service de Pharmacologie Clinique, Hôpital Pitié-Salpêtrière, APHP, Paris, France*
[14]*Dynamics and Pathophysiology of Neuronal Networks Team, Center for Interdisciplinary Research in Biology, Collège de France, CNRS UMR7241/INSERM U1050, MemoLife Labex, Paris, France*
[15]*Department of Neurology, Avicenne University Hospital, Sorbonne Paris Nord University, Bobigny, France*

ABSTRACT: Background: Machine learning algorithms using magnetic resonance imaging (MRI) data can accurately discriminate parkinsonian syndromes. Validation in patients recruited in routine clinical practice is missing.
Objective: The aim of this study was to assess the accuracy of a machine learning algorithm trained on a research cohort and tested on an independent clinical replication cohort for the categorization of parkinsonian syndromes.
Methods: Three hundred twenty-two subjects, including 94 healthy control subjects, 119 patients with Parkinson's disease (PD), 51 patients with progressive supranuclear palsy (PSP) with Richardson's syndrome, 35 with multiple system atrophy (MSA) of the parkinsonian variant (MSA-P), and 23 with MSA of the cerebellar variant (MSA-C), were recruited. They were divided into a training cohort (n = 179) scanned in a research environment and a replication cohort (n = 143) examined in clinical practice on different MRI systems. Volumes and diffusion tensor imaging (DTI) metrics in 13 brain regions were used as input for a supervised machine learning algorithm. To harmonize data across scanners and reduce scanner-dependent effects, we tested two types of normalizations using patient data or healthy control data.

***Correspondence to:** Dr. Lydia Chougar, Department of Neuroradiology, Pitié-Salpêtrière University Hospital, Boulevard de l'Hôpital, 75651 PARIS Cedex 13, France; E-mail: lydia.chougar@aphp.fr

Funding agencies: This work was supported by grants from Agence Nationale de la Recherche [grant numbers: ANR-11-INBS-0006 (France Life Imaging), ANRMNP 2009 (Nucleipark), ANR-11-INBS-0011 (NeurATRIS), Investissements d'Avenir (ANR-19-P3IA-0001, PRAIRIE 3IA Institute, and ANR-10-IAIHU-06, IHU – Paris Institute of Neurosciences), Association France Parkinson, Ecole Neuroscience de Paris, Électricité de France (Fondation d'Entreprise EDF), Institut National de la Santé et de la Recherche Médicale, DHOS-Inserm (2010, Nucleipark), PSP France, the Fondation Thérèse and René Planiol pour l'étude du Cerveau, European Union (EU) Framework Project 6–GENEPARK, the Programme Hospitalier de Recherche Clinique [grant Numbers: PHRC 2007-A00169-44 (LRRK) and PHRC 2004 (BBBIPPS)], the ICM Big Brain Theory Program (project PredictICD), and the Inria Project Lab Program (project Neuromarkers).

Relevant conflicts of interest/financial disclosures: Nothing to report.

Received: 25 August 2020; **Accepted:** 15 September 2020

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.28348

Results: In the replication cohort, high accuracies were achieved using volumetry in the classification of PD–PSP, PD–MSA-C, PSP–MSA-C, and PD-atypical parkinsonism (balanced accuracies: 0.840–0.983, area under the receiver operating characteristic curves: 0.907–0.995). Performances were lower for the classification of PD–MSA-P, MSA-C–MSA-P (balanced accuracies: 0.765–0.784, area under the receiver operating characteristic curve: 0.839–0.871) and PD–PSP–MSA (balanced accuracies: 0.773). Performance using DTI was improved when normalizing by controls, but remained lower than that using volumetry alone or combined with DTI.

Conclusions: A machine learning approach based on volumetry enabled accurate classification of subjects with early-stage parkinsonism, examined on different MRI systems, as part of their clinical assessment. © 2020 International Parkinson and Movement Disorder Society

Key Words: Parkinson's disease; progressive supranuclear palsy; multiple system atrophy; multimodal magnetic resonance imaging; machine learning algorithm

Diagnosis of idiopathic Parkinson's disease (PD) and atypical parkinsonism, whose most frequent types are progressive supranuclear palsy (PSP) and multiple system atrophy (MSA), relies on clinical criteria.[1–3] The diagnostic accuracy of PD is greatly improved when made by clinical experts in movement disorders with a sensitivity of 91.1% and a specificity of 98.4%.[1,4] However, on initial presentation, a correct diagnosis of atypical parkinsonism may be difficult, and clinical uncertainty is high. Accurate diagnosis is crucial to assess the prognosis, enroll patients in adequate care systems, and allow their inclusion in appropriate therapeutic trials.

Degeneration of dopaminergic neurons within the substantia nigra *pars compacta* is the hallmark of neurodegenerative parkinsonian syndromes.[5–7] Although patients with PD exhibit only nigral abnormalities in a limited number of small brain stem nuclei at the early stage of the disease,[5–8] patients with PSP show a larger involvement of the midbrain, dentate nucleus, and superior cerebellar peduncles (SCPs),[6,7,9–11] and patients with MSA are characterized by damage particularly affecting the posterior putamen in the parkinsonian variant (MSA-P) and the pons, middle cerebellar peduncles, and cerebellum in the cerebellar variant (MSA-C).[6,7,12]

Multimodal magnetic resonance imaging (MRI) can detect these different patterns of brain damage.[13–17] Atrophy is visible on T1-weighted images, tissue microstructure alterations are detected using diffusion-weighted images, and iron deposition can be evidenced using iron-sensitive sequences.[13–17] Using MRI data, machine learning algorithms can accurately differentiate between parkinsonian syndromes.[18–25] Most studies have used a single type of MRI data, either volumetry[18–20] or diffusion-weighted data.[25] Some studies have combined volumetry and diffusion,[21] or have included R2* relaxometry[22] or spectroscopy.[23] Two studies have included large cohorts of 1002 subjects[25] or 464 subjects,[19] whereas most studies have investigated smaller samples.[18,20–23] Only two studies have included subjects with PD, PSP, MSA-P, and MSA-C,[19,23] while other studies have not differentiated between MSA-P and MSA-C,[18] have included only patients with MSA-P[25] or MSA-C,[24] or have not included patients with MSA[20,21] or PSP.[22] Moreover, these studies have been mainly designed in a research environment and have been tested without an independent replication cohort.[19–22] To transpose an automated MRI classification approach based on machine learning to clinical practice, it is necessary to evaluate this approach on large cohorts scanned in clinical environments and including the different types of parkinsonian syndromes.

Our objective was to assess the predictive performance of machine learning algorithms for the categorization of parkinsonian syndromes, including PSP, MSA-C, and MSA-P, compared with PD and healthy control subjects (HCs). Such algorithms were trained on a research cohort and tested on an independent replication cohort scanned on different MRI scanners in clinical conditions in a neuroradiology department, using volumetry and diffusion tensor imaging (DTI).

Materials and Methods

Population

Two populations of participants were included: a training cohort to train and validate the algorithms and a replication cohort to independently evaluate each algorithm's performance. The training cohort was constituted of research studies conducted between 2007 and 2012 at the Paris Brain Institute (ICM): Genepark (LSHB-CT-2006-037544), BBBIPPS (DGS 2006/0524), and Nucleipark (RCB 2009-A00922-55). Inclusion criteria for patients were a diagnosis of PD, PSP, or MSA established by movement disorders specialists according to published consensus criteria for PD[1] with no or minimal cognitive disturbances with Mini Mental State Examination score >24, PSP with Richardson's syndrome,[26] or MSA.[3]

Participants in the replication cohort were consecutively enrolled between 2013 and 2019 in the movement disorders clinic of the Pitié-Salpêtrière Hospital, Paris. Diagnosis of probable PD, PSP, or MSA was retrospectively established in 2019 by movement disorders specialists according to the aforementioned clinical criteria based on all available clinical data.

The clinical examination included the Unified Parkinson's Disease Rating Scale Part III (UPDRS III) scores. Baseline MRI scans were obtained on the same days as the clinical examination. For both cohorts, HCs without a history of neurological or psychiatric disease were included. Subjects were excluded if they had any additional neurological disorder.

Local institutional review boards approved the studies (Genepark: CPP Paris II, 2007-A00208-45; BBBIPPS: CPP Paris VI, P040410–65-06; Nucleipark: CPP Paris VI, 65–09; Park Atypique: CPP Ile-de-France VI, 08012015). Written informed consent was obtained from all participants.

Image Acquisition

Participants in the training cohort were scanned at the Paris Brain Institute using a 3T Siemens Trio system (Siemens Healthineers, Erlangen, DE) with a 32-channel head coil. Participants in the replication cohort were scanned in clinical conditions for diagnostic purposes in the Neuroradiology Department of the hospital using two MRI systems: (1) a 3T GE Signa HDxt (GE Healthcare, Chicago, IL) with an 8-channel head coil, and (2) a 3T Siemens Skyra using a 64-channel head coil. HCs in the replication cohort were scanned twice during different sessions on Siemens Skyra and GE Signa systems.

All participants were scanned using a standardized protocol including a high-resolution T_1-weighted gradient-recalled echo sequence (magnetization-prepared rapid acquisition with gradient-recalled echo or spoiled gradient recalled acquisition in steady state) and DTI with 30 (Siemens Skyra and GE Signa), 60, or 64 (Siemens Trio) diffusion directions. Acquisition parameters are provided in Supporting Information Table S1. Quality control was performed by visual inspection; T_1-weighted and diffusion-weighted images with significant motion artifacts or image distortions were excluded.

Data Processing and Analysis

Image processing and analysis were performed using Matlab (R2017b; The MathWorks, Inc., Natick, MA). T_1-weighted images were automatically segmented using FreeSurfer 6.0 (http://freesurfer.net/; MGH, Boston, MA, USA).[27] DTI preprocessing was performed using the FMRIB Software Library v5.0 (FMRIB, Oxford, United Kingdom). Motion and eddy currents were corrected using the eddycor function. Fractional anisotropy (FA) and diffusivity maps were computed using the DTIfit function for the entire brain volume. The diffusion maps were coregistered to the three-dimensional T_1-weighted volume using the Statistical Parametric Mapping coregister function (SPM 12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/).

We included 13 regions of interest known for being involved in parkinsonian syndromes in the gray matter (midbrain, pons, putamen, posterior putamen, caudate, thalamus, pallidum, precentral cortex, insular cortex), white matter (SCPs, cerebellum white matter including the middle cerebellar peduncles), and ventricles (third ventricle, fourth ventricle). Volumes, average values of FA, mean diffusivity, axial diffusivity, and radial diffusivity (RD) were calculated in all segmented regions of interest and used as input features for the algorithms. To remove interindividual variability, all volumes were corrected by the total intracranial volume for each participant. Means of the volumes and DTI metrics were used for bilateral regions.

Normalization Procedures

Two independent normalization procedures were investigated and performed in both cohorts. "Normalization 1" scaled the features so that they were normally distributed. For a given between-group comparison (eg, PSP-PD), each variable of each participant was normalized using the mean and standard deviation (SD) of this variable in all patients in these two groups of the training cohort (eg, all patients with PSP and PD in the training cohort). For example, for the "PD-PSP" classification in either the training or the replication cohort, each variable of each patient with PD and PSP in the corresponding cohort was normalized using the following formula: (Variable – mean of PD and PSP in the training cohort)/SD of PD and PSP in the training cohort.

"Normalization 2" aimed to harmonize data across scanners, reduce scanner-dependent effects in the images, and find a good trade-off between standard normal distribution and same distribution of the features in both cohorts. Normalization 2 used the mean and SD of the controls scanned using the same scanner as the patient, according to the formula: (Variable – mean of controls scanned using the same scanner)/SD of controls scanned using the same scanner.

Machine Learning Algorithms

Using the scikit-learn package,[28] four supervised machine learning algorithms were used: logistic regression, support vector machine (SVM) with a linear kernel, SVM with a radial basis function kernel, and random forest. They were trained and validated on the training cohort, then tested on the replication cohort.

The cross-validation procedure on the training cohort included two nested loops: an outer loop with repeated stratified random splits with 50 repetitions evaluating the classification performances and an inner loop with 5-fold cross-validation used to optimize the hyperparameters of the algorithms. One model was created for each split, leading to 50 models. We selected the model with the highest mean balanced accuracies among the models with frequencies of appearance greater than 10% (Supporting Information Fig. S1).

Correlation between the different input features was investigated in both cohorts (Supporting Information Tables S2 and S3). Ridge regularization was used to avoid overfitting and to deal with possible correlation between the features. Three models were evaluated: "volumetry only," "DTI only," and "volumetry + DTI" with age and sex as covariates. DTI was missing in 13 patients of the training cohort and 14 patients of the replication cohort. Missing values were imputed with the mean values of the features from the training cohort. The algorithms were retrained and retested after removing subjects with missing DTI images. In addition, analysis with UPDRS III scores alone was performed, and performances were compared with the other models.

We evaluated binary classification tasks (PD–PSP, PD–MSA-P, PD–MSA-C, PSP–MSA-P, PSP–MSA-C, MSA-P–MSA-C, PD-atypical parkinsonism) and the multiclass classification task (PD-–SP–MSA, with patients with MSA merged in one group because of their small number).

Receiver operating characteristic curves were generated, and balanced accuracy (BA), area under the curve (AUC), sensitivity, and specificity were calculated to evaluate the algorithm performances. BA was defined as the average sensitivity and specificity in each group. BA avoids overestimation of classification performance because of imbalanced group sizes.[19]

Weighting Factors

Weighting factors of the volume and DTI metrics of each brain region were extracted from the logistic regression training after normalization 2 (corresponding to the best model). They reflected the contribution of each feature to group differentiation. A rescaling to a range of −1 to +1 was applied to highlight the relative importance of each feature: the higher the absolute value, the bigger the contribution of the feature. This assertion relied on the assumption that each feature had the same scale, which was a reasonable assumption because all features were standardized. When the coefficient was positive, the algorithm favored the first group if the value of the feature was high or the second group if the value was low, and vice versa.[19]

Statistical Analyses

Participant Characteristics

Statistical analyses were performed using R software (R Core Development Team, 2017). Group demographic and clinical scores within each cohort were analyzed. Statistical difference in sex distribution was evaluated using Fisher's exact test, followed by pairwise comparisons with *P* adjustment (Holm's method). Age, disease duration, and UPDRSIII scores were compared using the Kruskal–Wallis test, followed by pairwise comparisons with Holm–Bonferroni correction. Intercohort comparisons were also performed by assessing group-wise differences using Fisher's exact test for sex and Wilcoxon's rank sum test with continuity correction for age, disease duration, and the UPDRS III scores.

Comparison of Normalization Procedures, Biomarkers, and Algorithms

We compared BAs between normalization, biomarker, and classification methods in the replication cohort using multiple factor analysis (MFA) as a data-driven exploratory technique and repeated-measures analysis of variance (RM-ANOVA). For the MFA, a first analysis was run by grouping the data by normalization × biomarkers generating six blocks (2 normalizations × 3 biomarkers) of four variables (four algorithms) each to determine which "normalization × biomarker" blocks induced the same structures. A second analysis was run on four blocks (four algorithms) of six variables (2 normalizations × 3 biomarkers) to identify similarities/dissimilarities between algorithms performances. As a first step, individual principal component analysis was performed on each block, which was then normalized by the corresponding first eigenvalue. The obtained matrices were merged to form a global matrix, and a global principal component analysis was performed. The individual observations were then projected onto the global space.

We used RM-ANOVA with a $2 \times 4 \times 3$ factorial design [normalization (1 and 2) × algorithm (logistic regression, linear SVM, radial SVM, random forest) × biomarker (volumetry, volumetry + DTI, DTI)] after testing data for normality with Shapiro–Wilk test ($P > 0.05$). Sphericity assumption was checked using Mauchly's test and, if necessary, corrected using Greenhouse–Geisser correction. Pairwise comparisons with Bonferroni correction were then performed.

Results

Participants' Clinical and Demographic Characteristics

In total, 322 subjects were analyzed, divided into a training cohort (n = 179) and a replication cohort (n = 143) (Table 1). In the training cohort, there was a significant difference in sex distribution (Fisher's test, $P = 0.031$) due to a difference between HCs and patients with PD ($P = 0.040$). There was no significant difference in age and disease duration between groups. Patients with PD had lower UPDRS III scores than other patient groups ($P < 0.002$). In the replication cohort, there was a difference in age (Kruskal–Wallis, $P < 0.002$), with patients with PSP being older than other groups. There were no differences in sex distribution, disease duration, and UPDRS III scores between groups.

When comparing both cohorts, there was a significant difference in sex distribution for patients with PSP ($P = 0.042$) and age, because HCs ($P = 0.007$), patients with PD ($P = 0.002$), and patients with PSP ($P = 0.019$) were older in the replication cohort. Disease duration was shorter in the replication cohort for patients with PD ($P < 0.001$), MSA-P ($P < 0.001$), and MSA-C ($P = 0.020$), but not for patients with PSP ($P = 0.055$). UPDRS III scores were lower for MSA-P ($P = 0.003$) and MSA-C ($P = 0.002$) in the replication cohort.

Comparison of Normalization Methods

The first MFA showed that the first two components with eigenvalues greater than 1 explained 56% and 20% of the total variance. Both normalizations with volumetry and normalization 2 with volumetry + DTI strongly correlated with the first dimension, and thus showed the same profiles. Normalization 2 with DTI also correlated with the first dimension but less strongly. Normalization 1 with DTI and normalization 1 with volumetry + DTI had a different profile and did not perform as well as the other combinations (Supporting Information Figs. S2 and S3).

RM-ANOVA showed a significant effect of the normalization factor with higher performance of normalization 2 ($P < 0.001$) and no interactions with the biomarker and algorithm factors. Thus, only results with normalization 2 are provided in the following paragraphs (see Supporting Information Table S4 for results with normalization 1).

Comparisons of Biomarkers

The highest mean BA was observed for volumetry (mean BA = 0.803) followed by volumetry + DTI (mean BA = 0.756) and DTI (mean BA = 0.631). RM-ANOVA with sphericity correction showed a significant effect of biomarkers ($p = 0.03$) and a trend for an algorithm × biomarkers interaction ($P = 0.050$). Paired t tests confirmed that volumetry had similar performances to that of volumetry + DTI ($P = 0.1$), and that both performed better than DTI alone ($p < 0.001$). Therefore, DTI underperformed compared with volumetry and did not improve classification when combined with volumetry. Additional MFA did not reveal any significant difference between analyses performed with and without patients with missing DTI. Therefore, imputing the data did not have a significant effect on the overall analysis (Supporting Information Table S6 and Fig. S4).

Performances obtained with UPDRS III scores alone were low in both cohorts (BA: 0.397–0.861, AUC: 0.473–0.899 in the training cohort; BA: 0.316–0.656, AUC: 0.567–0.659 in the replication cohort) (Supporting Information Table S7).

Comparison of Algorithms

The second MFA showed that logistic regression, linear SVM, and random forest had similar performances and outperformed radial SVM (Supporting Information Fig. S5). The highest mean BA was observed for logistic regression (mean BA = 0.768), followed by random forest (mean BA = 0.754), linear SVM (mean BA = 0.745), and radial SVM (mean BA = 0.652). RM-ANOVA with sphericity correction showed a trend for an algorithm effect ($P = 0.055$) and a trend for a normalization × algorithm interaction ($P = 0.051$). Post hoc t tests showed that radial SVM underperformed compared with the other three algorithms that had similar performances. In the following paragraphs, only the results obtained with the logistic regression are reported because it provided the highest mean BA. Results with the three other algorithms are provided in Supporting Information Table S5.

Results per Group Comparison With Normalization 2 and Logistic Regression

In the training cohort, using volumetry, the best classification performances were obtained in decreasing order for PSP–MSA-C, PD–MSA-C, PD–PSP, PD-typical parkinsonian syndromes, and PD–MSA-P (BA: 0.892–0.963, AUC: 0.928–0.997). The performances were lower for the classification of PD–PSP–MSA (BA: 0.807), PSP–MSA-P, and MSA-C–MSA-P (BA: 0.668–0.732, AUC: 0.757–0.858). Combining volumetry and DTI did not tend to improve BAs (0.622–0.942) or AUCs (0.723–0.998), while performances with DTI alone were lower (BA: 0.373–0.880, AUC: 0.243–0.967) (Table 2).

In the replication cohort, using volumetry, performances remained high for the classifications of PD–MSA-C, PSP–MSA-C, PD–PSP, and PD-atypical parkinsonism (BA: 0.840–0.983, AUC: 0.907–0.995). Better accuracies than in the training cohort were obtained

TABLE 1. Demographic and clinical characteristics of the population

Groups	Training Cohort (n = 179)					Replication Cohort (n = 143)				
	HC	PD	PSP	MSA-P	MSA-C	HC	PD	PSP	MSA-P	MSA-C
n	72	63	21	11	12	22	56	30	24	11
Sex ratio (M/F)	24/48[a]	37/26	10/11[b]	4/7	7/5	12/10	36/20	23/7	18/6	7/4
Age at MRI scan (yr)	60.8 ± 8.2[c]	60.7 ± 9.7[c]	65.6 ± 9.1[c]	62.8 ± 7.1	60.3 ± 7.4	64.7 ± 7.3	66.6 ± 11.4	71.7 ± 5.7[d]	63.0 ± 7.0	58.3 ± 7.7
Disease duration (yr)	—	6.0 ± 4.0[e]	4.2 ± 1.9	5.0 ± 1.9 [e]	5.2 ± 1.8[e]	—	4.2 ± 3.4	3.4 ± 2.0	3.8 ± 1.8	2.4 ± 1.1
UPDRS III scores[f]	0.4 ± 0.7	21.0 ± 13.1[g]	38.5 ± 15.7	42.2 ± 13.9[h]	47.4 ± 14.4[h]	0.1 ± 0.3	20.6 ± 9.6	29.6 ± 17.6	26.3 ± 10.8	16.7 ± 7.7

Group demographics and clinical scores within each cohort were compared. Statistical difference in sex distribution was evaluated using Fisher's exact test, followed by pairwise comparisons with *P* adjustment (Holm's method). Age, disease duration, and UPDRS III scores were compared using Kruskal–Wallis test, followed by pairwise comparisons with Holm–Bonferroni correction. Intercohorts comparisons were also performed by assessing groupwise differences using Fisher's exact test for sex and Wilcoxon's rank sum test with continuity correction for age, disease duration, and UPDRS III scores.
Abbreviations: HC, healthy control subjects; PD, Parkinson's disease; PSP, progressive supranuclear palsy; MSA-P, parkinsonian variant of multiple system atrophy; MSA-C, cerebellar variant of multiple system atrophy; M, male; F, female; MRI, magnetic resonance imaging; UPDRS III, Unified Parkinson's Disease Rating Scale Part III.
[a] $P < 0.05$ for HCs vs patients with PD in the training cohort.
[b] $P < 0.05$ for patients with PSP between both cohorts.
[c] $P < 0.05$ for HCs, patients with PD, and patients with PSP between both cohorts.
[d] $P < 0.05$ for patients with PSP vs HCs and patients with PD; $P < 0.001$ for patients with PSP vs MSA-P and MSA-C in the replication cohort.
[e] $P < 0.001$ for PD and MSA-P groups between both cohorts; $P < 0.05$ for patients with MSA-C.
[f] UPDRS III scores were missing in 20.2% of the participants.
[g] $P < 0.002$ for PD vs PSP, MSA-P, and MSA-C in the training cohort.
[h] $P < 0.05$ for patients with MSA-P and MSA-C between both cohorts.

for the comparisons of PSP–MSA-P (BA: 0.896, AUC: 0.968) and MSA-C–MSA-P (BA: 0.784, AUC: 0.871). Performances were lower for the classifications of PD–MSA-P (BA: 0.765, AUC: 0.839) and PD–PSP–MSA (BA: 0.773). Classification accuracy using DTI alone was significantly lower (BA: 0.560–0.917, AUC: 0.731–0.982) than volumetry alone ($P < 0.001$), whereas volumetry + DTI (BA: 0.777–0.982, AUC: 0.833–0.994) did not differ from volumetry alone ($P = 0.1$) (Table 2).

Weighting Factors

The best features to differentiate PD and PSP were the volumes of the midbrain (1) and third ventricle (−0.94), FA in the SCP (0.72), and the volumes of the globus pallidus (0.71) and the putamen (0.63). For discriminating subjects with MSA-C, the most relevant features were the midbrain/pons ratio (vs PD and PSP: −1, MSA-P: 1) and the pons atrophy (vs PD: 0.61, MSA-P: −0.47, PSP: 0.31). The other features were the volumes of the fourth ventricle (vs PD: −0.46) and the cerebellum (vs PD: 0.49, PSP: 0.22, MSA-P: −0.36). MSA-P differentiation relied mostly on the putamen volume (0.46–1), with the contribution of DTI metrics in the putamen being low (Table 3 and Fig. 1).

Discussion

Our study demonstrates the feasibility of an automated classification of parkinsonian syndromes in a clinical setting using a large cohort of patients.[19,25] The classification algorithms were tested on a large independent replication cohort comprising patients recruited in a movement disorder clinic and scanned using different MRI systems as part of their routine diagnostic workup. Overall, patients in the replication cohort had a shorter disease duration than those in the training cohort, suggesting that the algorithm could differentiate between patients with early to moderately advanced parkinsonism. A further strength is that patients with both parkinsonian and cerebellar subtypes of MSA were included in addition to PD and PSP, which was only done in two previous studies.[19,23]

Our results are in agreement with previous studies using machine learning to differentiate parkinsonian syndromes and reporting BAs between from 69 and 89%,[19] and AUC greater than 93%,[25] and 95%.[18] Nevertheless, algorithm performances are difficult to compare across studies given differences in terms of input data, diseases studied, type of classification and performance indices used. Performances of the logistic regression using Volumetry were also equivalent to those obtained with automated methods using clinical measurements such as the Magnetic Resonance Parkinsonism Index (MRPI) for the differentiation of PSP-RS

TABLE 2. Performances of the logistic regression classification in both cohorts using normalization 2, with volumetry, DTI, and the combination of volumetry and DTI

		Training Cohort			Replication Cohort		
Group Comparisons	Metric	Volumetry	DTI	Volumetry + DTI	Volumetry	DTI	Volumetry + DTI
PD vs PSP	BA	0.892 (0.090)	0.794 (0.103)	0.892 (0.094)	0.840	0.644	0.879
	AUC	0.977 (0.033)	0.885 (0.104)	0.974 (0.033)	0.968	0.739	0.943
	Se	0.850 (0.189)	0.695 (0.197)	0.840 (0.187)	0.733	0.467	0.900
	Sp	0.934 (0.066)	0.892 (0.091)	0.945 (0.062)	0.946	0.821	0.857
PD vs MSA-P	BA	0.845 (0.127)	0.800 (0.190)	0.880 (0.133)	0.765	0.679	0.744
	AUC	0.928 (0.096)	0.863 (0.167)	0.975 (0.044)	0.839	0.749	0.847
	Se	0.909 (0.069)	0.931 (0.081)	0.969 (0.044)	0.821	0.857	0.946
	Sp	0.780 (0.251)	0.670 (0.373)	0.790 (0.269)	0.708	0.500	0.542
PD vs MSA-C	BA	0.963 (0.088)	0.798 (0.177)	0.928 (0.112)	0.982	0.902	0.982
	AUC	0.991 (0.026)	0.948 (0.070)	0.992 (0.023)	0.995	0.966	0.994
	Se	0.995 (0.024)	0.977 (0.045)	0.995 (0.018)	0.964	0.804	0.964
	Sp	0.930 (0.175)	0.620 (0.358)	0.860 (0.227)	1.000	1.000	1.000
PSP vs MSA-C	BA	0.938 (0.084)	0.880 (0.100)	0.942 (0.092)	0.983	0.917	0.967
	AUC	0.997 (0.016)	0.967 (0.058)	0.998 (0.012)	0.991	0.982	0.994
	Se	0.970 (0.082)	0.900 (0.152)	0.990 (0.049)	0.967	0.833	0.933
	Sp	0.907 (0.166)	0.860 (0.179)	0.893 (0.184)	1.000	1.000	1.000
PSP vs MSA-P	BA	0.732 (0.181)	0.833 (0.155)	0.828 (0.156)	0.896	0.750	0.833
	AUC	0.858 (0.149)	0.966 (0.066)	0.962 (0.075)	0.968	0.814	0.892
	Se	0.784 (0.180)	0.896 (0.135)	0.896 (0.147)	0.833	0.833	0.833
	Sp	0.680 (0.331)	0.770 (0.307)	0.760 (0.307)	0.958	0.667	0.833
MSA-C vs MSA-P	BA	0.668 (0.152)	0.373 (0.172)	0.622 (0.179)	0.784	0.725	0.788
	AUC	0.757 (0.203)	0.243 (0.185)	0.723 (0.217)	0.871	0.811	0.871
	Se	0.630 (0.282)	0.320 (0.346)	0.590 (0.345)	0.750	0.542	0.667
	Sp	0.707 (0.239)	0.427 (0.337)	0.653 (0.260)	0.818	0.909	0.909
PD vs atypical parkinsonism	BA	0.879 (0.073)	0.742 (0.089)	0.880 (0.077)	0.853	0.658	0.839
	AUC	0.948 (0.051)	0.832 (0.097)	0.947 (0.053)	0.907	0.731	0.905
	Se	0.903 (0.083)	0.840 (0.098)	0.934 (0.062)	0.875	0.732	0.893
	Sp	0.856 (0.113)	0.644 (0.174)	0.827 (0.139)	0.831	0.585	0.785
PD vs PSP vs MSA[a]	BA	0.807 (0.071)	0.711 (0.100)	0.812 (0.078)	0.773	0.560	0.777
	Se (PD)	0.906 (0.080)	0.842 (0.100)	0.912 (0.081)	0.875	0.750	0.911
	Se (MSA)	0.684 (0.145)	0.672 (0.218)	0.708 (0.171)	0.743	0.629	0.686
	Se (PSP)	0.830 (0.176)	0.620 (0.201)	0.815 (0.172)	0.700	0.300	0.733

For the training cohort, we have reported means and standard deviations (in parentheses) of performance metrics across all repetitions during the cross-validation procedure.

Abbreviations: DTI, diffusion tensor imaging; PD, Parkinson's disease; PSP, progressive supranuclear palsy; BA, balanced accuracy; AUC, area under the curve; Se, sensitivity; Sp, specificity; MSA-P, parkinsonian variant of multiple system atrophy; MSA-C, cerebellar variant of multiple system atrophy; HC, healthy control subjects.

[a]For the multiclass classification (PD vs PSP vs MSA), in addition to BAs, we have reported sensitivities, that is, the proportion of subjects who have been accurately classified for each class. For instance, Se(PD) = 0.80 means that 80% of the patients with PD have been classified as PD by the algorithm.

from non-PSP participants.[29] However, we found lower accuracy for the classification of MSA-P versus PD and MSA-C and for the multiclass classification. Indeed, MSA-P and MSA-C patients often have overlapping features, a clinical and brain imaging continuum existing between both variants, which makes them more difficult to distinguish.[5,12] The lower performance in categorizing MSA-P and MSA-C may also be explained by the fact we included mixed participants with both patterns of MSA in the MSA-P group, and by insufficient algorithm training due to the relative small number of MSA patients in the training cohort.

We confirmed that morphometric measurements are robust MRI biomarkers for the discrimination of parkinsonism.[18–20] In line with previous pathological and imaging studies,[6,7,13,17,30] the best features for the differentiation between PD and PSP were the midbrain and third ventricle volumes. The midbrain to pons volume ratio was the most relevant feature for the discrimination between PSP and both MSA variants as reported with manual morphometric indices such as the midbrain to pons area ratio[31,32] and the MRPI[33]. The putamen volume was highly discriminant between MSA-P and PD, MSA-P being characterized by a prominent putamen atrophy as compared to PD,[14,17,30,34] but less relevant for the differentiation with PSP30 and MSA-C patients[12].

Diffusion measurements in the putamen had low contributions to differentiate MSA-P from other disease groups in our study. Several studies have shown increased diffusivity in the posterior putamen in MSA-P versus PD patients,[35–37] Results are less clear between

TABLE 3. Weighting factors extracted from the logistic regression training for the different group comparisons using normalization 2

Feature \ Group	PD vs PSP	PD vs MSA-P	PD vs MSA-C	PSP vs MSA-C	PSP vs MSA-P	MSA-C vs -P	PD vs atypical parkinsonism
Midbrain_vol	1.00	0.28	0.34	-0.15	-0.46	-0.11	1.00
Pons_vol	0.42	0.29	0.61	0.31	0.14	-0.47	0.74
Midbrain/Pons_vol	0.32	-0.27	-1.00	-1.00	-1.00	1.00	-0.41
SCP_vol	0.37	0.16	0.25	0.04	-0.04	-0.13	0.48
V3_vol	-0.94	-0.18	-0.06	0.37	0.64	-0.12	-0.69
V4_vol	-0.46	-0.41	-0.50	-0.25	-0.10	0.41	-0.71
Cerebellum_vol	0.49	0.35	0.49	0.22	0.07	-0.36	0.81
Thalamus_vol	0.59	0.22	0.07	-0.21	-0.24	0.18	0.36
Caudate_vol	0.12	0.33	0.02	-0.04	0.20	0.24	0.17
Putamen_vol	0.63	1.00	0.20	-0.09	0.46	0.57	0.98
Pallidum_vol	0.71	0.51	0.17	-0.17	0.03	0.36	0.83
Insula_vol	-0.02	0.04	0.03	0.04	0.05	-0.05	-0.15
Precentral_vol	0.24	0.25	-0.01	-0.12	0.03	0.23	0.06
Midbrain_FA	0.40	0.00	0.13	-0.05	-0.20	-0.06	0.24
Pons_FA	0.30	0.15	0.18	0.05	0.06	-0.03	0.20
SCP_FA	0.72	0.17	0.11	-0.22	-0.44	0.06	0.56
Putamen_FA	-0.25	-0.23	-0.11	0.00	-0.11	-0.09	-0.43
Posteriorputamen_FA	-0.13	-0.14	-0.11	-0.04	-0.04	0.06	-0.25
Pallidum_FA	0.01	0.00	0.02	0.00	-0.03	0.00	0.13
Thalamus_FA	0.15	-0.11	-0.03	-0.09	-0.33	-0.11	-0.15
Caudate_FA	-0.01	0.08	0.07	0.05	0.00	-0.08	0.07
Cerebellum_FA	0.21	-0.11	0.13	0.02	-0.23	-0.21	-0.04
Insula_FA	0.01	0.03	0.11	0.06	-0.03	-0.14	-0.02
Precentral_FA	0.25	0.07	0.10	-0.03	-0.21	-0.11	0.26
Midbrain_MD	-0.33	0.01	-0.11	0.04	0.20	0.10	-0.20
Pons_MD	-0.04	-0.08	-0.19	-0.18	-0.26	0.12	-0.05
SCP_MD	-0.22	-0.56	-0.31	-0.21	-0.52	-0.01	-0.43
Putamen_MD	-0.06	0.07	-0.05	-0.01	0.06	0.06	-0.07
Posteriorputamen_MD	0.00	-0.01	-0.13	-0.09	-0.05	0.11	-0.12
Pallidum_MD	0.04	0.21	0.08	0.05	0.17	0.02	0.11
Thalamus_MD	-0.03	0.18	0.10	0.09	0.23	0.03	0.20
Caudate_MD	-0.20	-0.08	-0.01	0.05	0.13	0.02	-0.21
Cerebellum_MD	-0.08	-0.23	-0.27	-0.24	-0.34	0.18	-0.23
Insula_MD	-0.27	0.03	-0.02	0.08	0.25	0.08	-0.13
Precentral_MD	-0.21	0.07	0.02	0.10	0.26	0.04	-0.06
Midbrain_AD	-0.21	0.11	-0.07	0.04	0.28	0.16	-0.08
Pons_AD	0.07	0.00	-0.18	-0.22	-0.26	0.18	0.07
SCP_AD	0.14	-0.41	-0.22	-0.29	-0.64	0.03	-0.08
Putamen_AD	-0.12	0.01	-0.08	-0.02	0.03	0.04	-0.19
Pallidum_AD	0.07	0.27	0.12	0.07	0.20	0.02	0.22
Thalamus_AD	0.00	0.20	0.11	0.08	0.22	0.03	0.22
Caudate_AD	-0.18	-0.06	0.01	0.06	0.12	0.01	-0.16
Cerebellum_AD	0.01	-0.21	-0.20	-0.21	-0.37	0.10	-0.15
Insula_AD	-0.25	0.05	0.01	0.10	0.25	0.05	-0.09
Precentral_AD	-0.14	0.11	0.07	0.10	0.22	0.00	0.05
Midbrain_RD	-0.34	0.01	-0.10	0.09	0.24	-0.02	-0.26
Pons_RD	-0.06	0.05	-0.16	-0.11	-0.07	0.09	0.00
SCP_RD	-0.46	-0.43	-0.23	0.01	-0.03	-0.08	-0.46
Putamen_RD	-0.03	0.10	0.07	0.10	0.07	-0.17	0.09
Pallidum_RD	0.05	0.30	0.18	0.14	0.20	-0.11	0.34
Thalamus_RD	-0.08	0.24	0.10	0.13	0.27	-0.04	0.16
Caudate_RD	-0.28	-0.02	-0.02	0.11	0.22	-0.02	-0.28
Cerebellum_RD	-0.22	-0.23	-0.43	-0.29	-0.25	0.30	-0.45
Insula_RD	-0.36	-0.04	0.02	0.19	0.25	-0.17	-0.25
Precentral_RD	-0.39	0.00	0.07	0.25	0.42	-0.16	-0.21

Weighting factors were scaled to a range of −1 to +1. Higher absolute values indicate greater contribution of the feature. When the coefficient was positive, the algorithm favored the first disease group if the value of the feature was high or the second disease group if the value was low. Conversely, when the coefficient was negative, the algorithm favored the second group if the value of the feature was high, or the first group if the value was low. Positive values are highlighted in shades of red while negative values are highlighted in shades of blue.

Abbreviations: PD, Parkinson's disease; PSP, progressive supranuclear palsy; vol, volume; MSA-P, parkinsonian variant of multiple system atrophy; MSA-C, cerebellar variant of multiple system atrophy; SCP, superior cerebellar peduncles; V3, third ventricle; V4, fourth ventricle; FA, fractional anisotropy; MD, mean diffusivity; AD, axial diffusivity; RD, radial diffusivity.

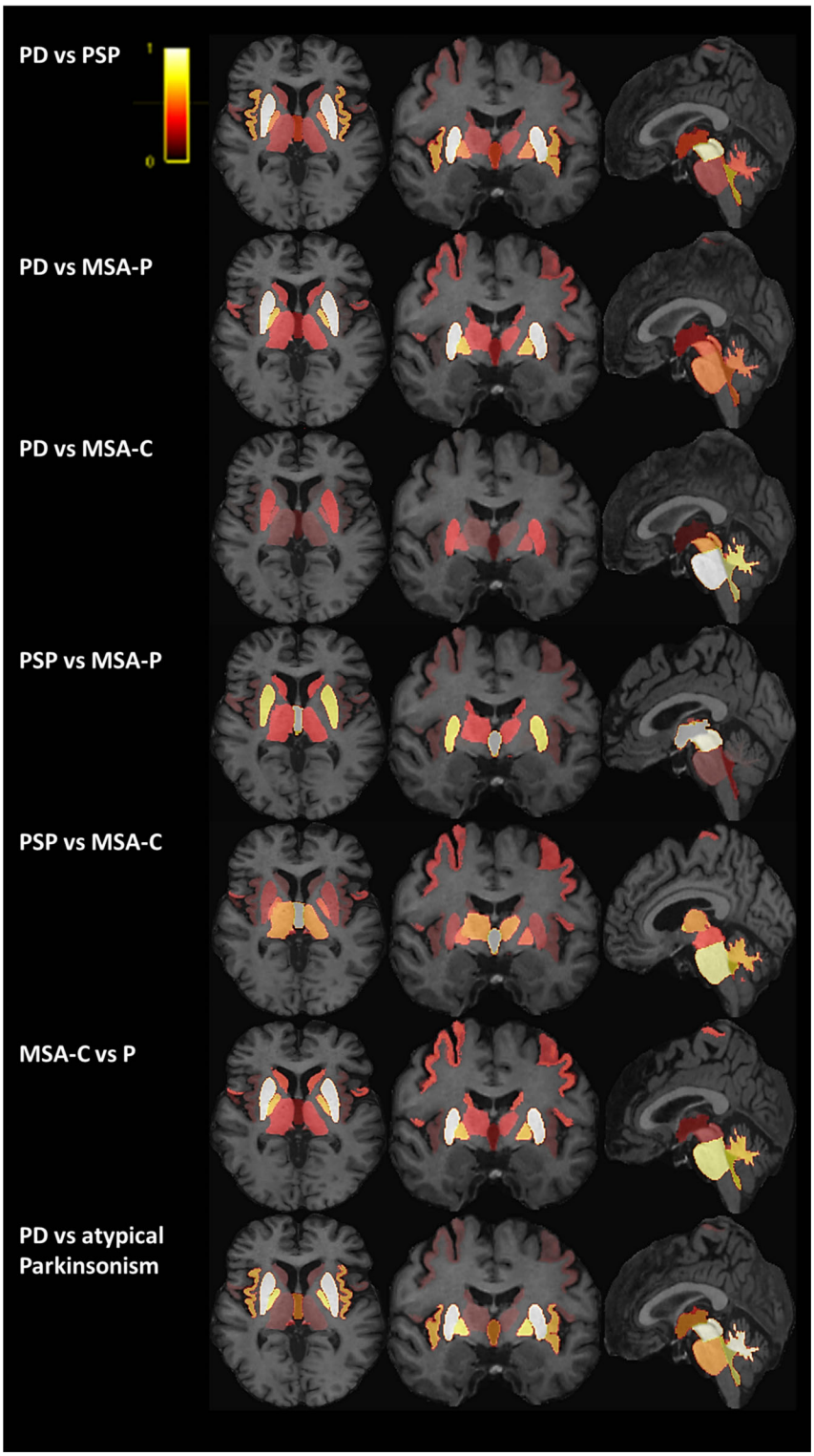

FIG. 1. Contribution of each region of interest for the group classification using volumetry. A color shade was attributed to the weighting factor corresponding to the volume for each brain region, depending on its relevance for the group differentiation. A color bar in the top left corner of the figure indicates the shades, with brighter color representing greater weight. Weighting factors were rescaled to the range of 0 to 1. Weighting factors corresponding to diffusion tensor imaging metrics are not represented. The most relevant regions were Parkinson's disease (PD) versus progressive supranuclear palsy (PSP): midbrain and putamen; PD versus parkinsonian variant of multiple system atrophy (MSA-P): putamen; PD versus cerebellar variant of multiple system atrophy (MSA-C): pons; PSP versus MSA-P: midbrain and third ventricle; PSP versus MSA-C: pons and third ventricle; MSA-P versus MSA-C: putamen and pons; and PD versus atypical parkinsonism: putamen, midbrain, cerebellum, and fourth ventricle. [Color figure can be viewed at wileyonlinelibrary.com]

MSA-P and PSP patients with overlapping diffusivity[38–40] and iron content[40,41] reported in the entire putamen. Measurements in the posterior putamen may provide better results.[37] In our study, co-registration inaccuracies between T1-weighted images and diffusion maps due to echo planar imaging distortions and susceptibility artifacts could have contributed to reducing the accuracy of the diffusion measurements and the performance of the algorithms.

In the replication cohort, normalization using control data reduced the scanner effect and improved categorization performances using DTI. However, combining DTI to Volumetry did not improve performances while DTI alone had lower performances, as assessed by the overall low weighting factors associated with DTI measurements. The heterogeneity of DTI data in the replication cohort, which included images acquired using two MRI systems and heterogeneous acquisition parameters (voxel size, number of directions, geometric distortions), probably explained the lower categorization accuracy. Standardization of DTI acquisition parameters may improve classification performances as suggested previously.[25]

Our study has several limitations. First, there was no neuropathological confirmation of parkinsonian diagnosis, which is the case in most neurodegenerative studies. Second, our study focused on PSP-RS while the differential diagnosis between PD and PSP-Parkinsonism is also challenging. The recruitment of PSP variants will be the scope of future studies. Third, DTI-derived white matter tract microstructure could have improved categorization performance, especially since these large tracts could be less susceptible to changes in DTI scan resolution. Future work could also involve free-water and free-water-corrected measurements, which may improve classification accuracy.[25] Furthermore, the automated segmentation and machine learning approach that we used are time-consuming and not yet available in clinical routine unlike morphometric methods such as the MRPI. However, there is a huge interest in machine learning approaches, with a high potential for translation in clinical practice.

To conclude, our study showed that automated categorization of parkinsonian syndromes was applicable to patients with early to moderately advanced parkinsonism recruited in a clinical environment, despite variability in scanners and acquisition parameters. Volumetry was the most robust discriminative biomarker. Improvement in acquisition and analysis of diffusion data and inclusion of iron measurements may improve classification. Implementing a machine learning algorithm in the clinical workflow may thus be relevant to help clinicians improve diagnosis of parkinsonism. ●

Acknowledgments: This work was supported by grants from Agence Nationale de la Recherche [grant numbers: ANR-11-INBS-0006 (France Life Imaging), ANRMNP 2009 (Nucleipark), ANR-11-INBS-0011 (NeurATRIS), Investissements d'Avenir (ANR-19-P3IA-0001, PRAIRIE 3IA Institute, and ANR-10-IAIHU-06, IHU – Paris Institute of Neurosciences), Association France Parkinson, Ecole Neuroscience de Paris, Électricité de France (Fondation d'Entreprise EDF), Institut National de la Santé et de la Recherche Médicale, DHOS-Inserm (2010, Nucleipark), PSP France, the Fondation Thérèse and René Planiol pour l'étude du Cerveau, European Union (EU) Framework Project 6–GENEPARK, the Programme Hospitalier de Recherche Clinique [grant Numbers: PHRC 2007-A00169-44 (LRRK) and PHRC 2004 (BBBIPPS)], the ICM Big Brain Theory Program (project PredictICD), and the Inria Project Lab Program (project Neuromarkers).

References

1. Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain 2002;125:861–870.
2. Höglinger GU, Respondek G, Stamelou M, et al. Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria. Mov Disord 2017;32:853–864.
3. Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. Neurology 2008;71:670–676.
4. Berardelli A, Wenning GK, Antonini A, et al. EFNS/MDS-ES/ENS [corrected] recommendations for the diagnosis of Parkinson's disease. Eur J Neurol 2013;20:16–34.
5. Dickson DW. Neuropathology of Parkinson disease. Parkinsonism Relat Disord 2018;46(suppl 1):S30–S33.
6. Dickson DW. Parkinson's disease and parkinsonism: neuropathology. Cold Spring Harb Perspect Med [online serial] 2012;2. Accessed at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3405828/. Accessed March 16, 2019.
7. Levin J, Kurz A, Arzberger T, Giese A, Höglinger GU. The differential diagnosis and treatment of atypical parkinsonism. Dtsch Arztebl Int 2016;113:61–69.
8. Kalia LV, Lang AE. Parkinson's disease. The Lancet 2015;386:896–912.
9. Steele JC, Richardson JC, Olszewski J. Progressive supranuclear palsy: a heterogeneous degeneration involving the brain stem, Basal Ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. Semin Neurol 2014;34:129–150.
10. Hauw JJ, Daniel SE, Dickson D, et al. Preliminary NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). Neurology 1994;44:2015–2019.
11. Pyatigorskaya N, Yahia-Cherif L, Gaurav R, et al. Multimodal magnetic resonance imaging quantification of brain changes in progressive supranuclear palsy. Mov Disord 2020;35:161–170.
12. Castellani R. Multiple system atrophy. Am J Pathol 1998;153:671–676.
13. Whitwell JL, Höglinger GU, Antonini A, et al. Radiological biomarkers for diagnosis in PSP: where are we and where do we need to be? Mov Disord 2017;32:955–971.
14. Heim B, Krismer F, Seppi K. Structural imaging in atypical parkinsonism. Int Rev Neurobiol 2018;142:67–148.
15. Lehericy S, Vaillancourt DE, Seppi K, et al. The role of high-field magnetic resonance imaging in parkinsonian disorders: pushing the boundaries forward. Mov Disord 2017;32:510–525.
16. Rizzo G, Zanigni S, De Blasi R, et al. Brain MR contribution to the differential diagnosis of Parkinsonian syndromes: an update. Parkinsons Dis 2016;2016:2983638.
17. Kassubek J. MRI-based neuroimaging: atypical parkinsonisms and other movement disorders. Curr Opin Neurol 2018;31:425–430.
18. Scherfler C, Göbel G, Müller C, et al. Diagnostic potential of automated subcortical volume segmentation in atypical parkinsonism. Neurology 2016;86:1242–1249.
19. Huppertz H-J, Möller L, Südmeyer M, et al. Differentiation of neurodegenerative parkinsonian syndromes by volumetric magnetic resonance imaging analysis and support vector machine classification. Mov Disord 2016;31:1506–1517.
20. Salvatore C, Cerasa A, Castiglioni I, et al. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy. J Neurosci Methods 2014;222:230–237.

21. Cherubini A, Morelli M, Nisticó R, et al. Magnetic resonance support vector machine discriminates between Parkinson disease and progressive supranuclear palsy. Mov Disord 2014;29:266–269.

22. Péran P, Barbagallo G, Nemmi F, et al. MRI supervised and unsupervised classification of Parkinson's disease and multiple system atrophy. Mov Disord 2018;33:600–608.

23. Morisi R, Manners DN, Gnecco G, et al. Multi-class parkinsonian disorders classification with quantitative MR markers and graph-based features using support vector machines. Parkinsonism Relat Disord 2018;47:64–70.

24. Du G, Lewis MM, Kanekar S, et al. Combined diffusion tensor imaging and apparent transverse relaxation rate differentiate Parkinson disease and atypical parkinsonism. AJNR Am J Neuroradiol 2017;38:966–972.

25. Archer DB, Bricker JT, Chu WT, Burciu RG, McCracken JL. Development and validation of the automated imaging differentiation in parkinsonism (AID-P): a multicentre machine learning study. Lancet Digit Health 2019;1(5):e222–e231.

26. Litvan I, Agid Y, Calne D, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. Neurology 1996;47:1–9.

27. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341–355.

28. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–2830.

29. Nigro S, Antonini A, Vaillancourt DE, et al. Automated MRI classification in progressive supranuclear palsy: a large international cohort study. Mov Disord 2020;35(6):976–983.

30. Messina D, Cerasa A, Condino F, et al. Patterns of brain atrophy in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. Parkinsonism Relat Disord 2011;17:172–176.

31. Mangesius S, Hussl A, Krismer F, et al. MR planimetry in neurodegenerative parkinsonism yields high diagnostic accuracy for PSP. Parkinsonism Relat Disord 2018;46:47–55.

32. Möller L, Kassubek J, Südmeyer M, et al. Manual MRI morphometry in Parkinsonian syndromes. Mov Disord 2017;32:778–782.

33. Quattrone A, Nicoletti G, Messina D, et al. MR imaging index for differentiation of progressive supranuclear palsy from Parkinson disease and the Parkinson variant of multiple system atrophy. Radiology 2008;246:214–221.

34. Whitwell JL, Master AV, Avula R, et al. Clinical correlates of white matter tract degeneration in progressive supranuclear palsy. Arch Neurol 2011;68:753–760.

35. Seppi K, Schocke MFH, Prennschuetz-Schuetzenau K, et al. Topography of putaminal degeneration in multiple system atrophy: a diffusion magnetic resonance study. Mov Disord 2006;21:847–852.

36. Pellecchia MT, Barone P, Mollica C, et al. Diffusion-weighted imaging in multiple system atrophy: a comparison between clinical subtypes. Mov Disord 2009;24:689–696.

37. Tsukamoto K, Matsusue E, Kanasaki Y, et al. Significance of apparent diffusion coefficient measurement for the differential diagnosis of multiple system atrophy, progressive supranuclear palsy, and Parkinson's disease: evaluation by 3.0-T MR imaging. Neuroradiology 2012;54:947–955.

38. Nicoletti G, Lodi R, Condino F, et al. Apparent diffusion coefficient measurements of the middle cerebellar peduncle differentiate the Parkinson variant of MSA from Parkinson's disease and progressive supranuclear palsy. Brain 2006;129:2679–2687.

39. Seppi K, Schocke MFH, Esterhammer R, et al. Diffusion-weighted imaging discriminates progressive supranuclear palsy from PD, but not from the parkinson variant of multiple system atrophy. Neurology 2003;60:922–927.

40. Focke NK, Helms G, Pantel PM, et al. Differentiation of typical and atypical Parkinson syndromes by quantitative MR imaging. Am J Neuroradiol 2011;32:2087–2092.

41. Sjöström H, Granberg T, Westman E, Svenningsson P. Quantitative susceptibility mapping differentiates between parkinsonian disorders. Parkinsonism Relat Disord 2017;44:51–57.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

SGML and CITI Use Only
DO NOT PRINT

Author Roles

Lydia Chougar designed and conceptualized the study, analyzed the data, and drafted the manuscript for intellectual content. Johann Faouzi designed and conceptualized the study, analyzed the data, and revised the manuscript for intellectual content. Nadya Pyatigorskaya designed and conceptualized the study and revised the manuscript for intellectual content. Lydia Yahia-Cherif performed data analysis and revised the manuscript. Rahul Gaurav performed data analysis and revised the manuscript. Emma Biondetti performed data analysis and revised the manuscript. Marie Villotte performed data collection and analysis and revised the manuscript. Romain Valabrègue performed data analysis and revised the manuscript. Jean-Christophe Corvol revised the manuscript. Alexis Brice revised the manuscript. Louise-Laure Mariani revised the manuscript. Florence Cormier revised the manuscript. Marie Vidailhet designed and conceptualized the study, and revised the manuscript for intellectual content. Gwendoline Dupont performed data collection and revised the manuscript. Ines Piot performed data collection and revised the manuscript. David Grabli designed and conceptualized study and revised the manuscript for intellectual content. Christine Payan performed data collection and revised the manuscript. Olivier Colliot designed and conceptualized the study, analyzed the data, and revised the manuscript for intellectual content. Bertrand Degos designed and conceptualized study and revised the manuscript for intellectual content. Stéphane Lehéricy designed and conceptualized study and revised the manuscript for intellectual content.

Financial Disclosures

Lydia Chougar reports no disclosures. Johann Faouzi reports no disclosures. Nadya Pyatigorskaya reports no disclosures. Lydia Yahia-Cherif reports no disclosures. Rahul Gaurav reports no disclosures related to the present work. Competing financial interests unrelated to the present work: received a grant from Biogen Inc. Emma Biondetti reports no disclosures related to the present work. Competing financial interests unrelated to the present work: received grant funding from Biogen Inc. Marie Villotte reports no disclosures. Romain Valabrègue reports no disclosures. Jean-Christophe Corvol reports no disclosures related to the present work; has served in scientific advisory boards for Biogen, Denali, Ever Pharma, Isdorsia, Prevail Therapeutics, and UCB; and has received grants from Sanofi, the Michael J Fox Foundation, ANR, France Parkinson, the French Ministry of Health. Alexis Brice reports no disclosures. Louise-Laure Mariani reports no disclosures related to the present work. Competing financial interests unrelated to the present work: received research support grants from INSERM, JNLF, The L'Oreal Foundation; speech honoraria from CSL, Sanofi-Genzyme, Lundbeck, Teva; consultant for Alzprotect, Bionure, Digitsole and travel funding from the Movement Disorders Society, ANAINF, Merck, Merz, Medtronic, Teva, and AbbVie, outside the submitted work. Florence Cormier reports no disclosures. Marie Vidailhet reports no disclosures. Gwendoline Dupont reports no disclosures. Ines Piot reports no disclosures. David Grabli reports no disclosures. Christine Payan reports no disclosures. Olivier Colliot reports no disclosures related to the present work. Competing financial interests unrelated to the present work: received consulting fees from AskBio (2020), received fees for writing a lay audience short paper from Expression Santé (2019), received speaker fees for a lay audience presentation from Palais de la découverte (2017). His laboratory received grants (paid to the institution) from Air Liquide Medical Systems (2011–2016) and Qynapse (2017–present). Members from his laboratory have cosupervised a PhD thesis with myBrainTechnologies (2016–present). Colliot's spouse is an employee of myBrainTechnologies (2015–present). Colliot has submitted a patent to the International Bureau of the World Intellectual Property Organization (PCT/IB2016/0526993, J.-B. Schiratti, S. Allassonniere, O. Colliot, S. Durrleman, "A method for determining the temporal progression of a biological phenomenon and associated methods and devices," 2016). Bertrand Degos reports no disclosures related to the present work. Competing financial interests unrelated to the present work: received research support grants from Fondation de France, Inserm, and ANR; speech honoraria from Ipsen, Merz Pharma, and Orkyn; and received travel funding from Merz Pharma, Elivie, and Orkyn. Stéphane Lehéricy reports no disclosures related to the present work. Competing financial interests unrelated to the present work: received grants from 'Investissements d'avenir' (grant numbers ANR-10-IAIHU-06 and ANR-11-INBS-0006) and Biogen Inc.