Development and validation of the automated imaging differentiation in parkinsonism (AID-P): a multicentre machine learning study

Derek B Archer, Justin T Bricker, Winston T Chu, Roxana G Burciu, Johanna L McCracken, Song Lai, Stephen A Coombes, Ruogu Fang, Angelos Barmpoutis, Daniel M Corcos, Ajay S Kurani, Trina Mitchell, Mieniecia L Black, Ellen Herschel, Tanya Simuni, Todd B Parrish, Cynthia Comella, Tao Xie, Klaus Seppi, Nicolaas I Bohnen, Martijn LTM Müller, Roger L Albin, Florian Krismer, Guangwei Du, Mechelle M Lewis, Xuemei Huang, Hong Li, Ofer Pasternak, Nikolaus R McFarland, Michael S Okun, David E Vaillancourt

Summary

Background Development of valid, non-invasive biomarkers for parkinsonian syndromes is crucially needed. We aimed to assess whether non-invasive diffusion-weighted MRI can distinguish between parkinsonian syndromes using an automated imaging approach.

Methods We did an international study at 17 MRI centres in Austria, Germany, and the USA. We used diffusion-weighted MRI from 1002 patients and the Movement Disorders Society Unified Parkinson's Disease Rating Scale part III (MDS-UPDRS III) to develop and validate disease-specific machine learning comparisons using 60 template regions and tracts of interest in Montreal Neurological Institute space between Parkinson's disease and atypical parkinsonism (multiple system atrophy and progressive supranuclear palsy) and between multiple system atrophy and progressive supranuclear palsy. For each comparison, models were developed on a training and validation cohort and evaluated in an independent test cohort by quantifying the area under the curve (AUC) of receiving operating characteristic curves. The primary outcomes were free water and free-water-corrected fractional anisotropy across 60 different template regions.

Findings In the test cohort for disease-specific comparisons, the diffusion-weighted MRI plus MDS-UPDRS III model (Parkinson's disease *vs* atypical parkinsonism had an AUC 0·962; multiple system atrophy *vs* progressive supranuclear palsy AUC 0·897) and diffusion-weighted MRI only model had high AUCs (Parkinson's disease *vs* atypical parkinsonism AUC 0·955; multiple system atrophy *vs* progressive supranuclear palsy AUC 0·926), whereas the MDS-UPDRS III only models had significantly lower AUCs (Parkinson's disease *vs* atypical parkinsonism 0·775; multiple system atrophy *vs* progressive supranuclear palsy 0·582). These results indicate that a non-invasive imaging approach is capable of differentiating forms of parkinsonism comparable to current gold standard methods.

Interpretations This study provides an objective, validated, and generalisable imaging approach to distinguish different forms of parkinsonian syndromes using multisite diffusion-weighted MRI cohorts. The diffusion-weighted MRI method does not involve radioactive tracers, is completely automated, and can be collected in less than 12 min across 3T scanners worldwide. The use of this test could positively affect the clinical care of patients with Parkinson's disease and parkinsonism and reduce the number of misdiagnosed cases in clinical trials.

Funding National Institutes of Health and Parkinson's Foundation.

Introduction

Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy are neurodegenerative disorders that are challenging to differentiate because of shared and overlapping motor and non-motor features.[1–3] Misdiagnosis of Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy is frequent, especially early in disease. Diagnosis accuracy in early Parkinson's disease (<5 years duration) is about 58%, and 54% of misdiagnosed patients have multiple system atrophy or progressive supranuclear palsy.[3–7] Although dopamine transporter imaging can identify the nigrostriatal denervation that leads to dopaminergic deficiency, it cannot distinguish between the different forms of parkinsonism, because they all exhibit this characteristic.[8] The absence of a clinically reliable non-invasive biomarker to distinguish different parkinsonian syndromes is a major hindrance to improved diagnosis accuracy and, therefore, improved categorisation in clinical trials; however, diffusion-weighted MRI shows particular promise in addressing this shortfall, because it is sensitive to detecting microstructural differences between forms of parkinsonism.

Diffusion-weighted MRI facilitates in-vivo quantification of brain microstructure associated with histology using a measure called fractional anisotropy.[9] However, fractional anisotropy can be susceptible to partial volume effects, because both tissue and fluid are contained in its

Lancet Digital Health 2019; 1: e222–31

Published Online August 27, 2019 http://dx.doi.org/10.1016/S2589-7500(19)30105-0

See Comment page e196

Laboratory for Rehabilitation Neuroscience, Department of Applied Physiology and Kinesiology (D B Archer PhD, J T Bricker MS, W T Chu BS, J L McCracken BS, S A Coombes PhD, T Mitchell MS, M L Black MPH, Prof D E Vaillancourt PhD), **J Crayton Pruitt Family Department of Biomedical Engineering** (W T Chu, R Fang PhD, A Barmpoutis PhD, Prof D E Vaillancourt), **Department of Radiation Oncology** (Prof S Lai PhD), **CTSI Human Imaging Core** (Prof S Lai), **Fixel Institute for Neurological Disease** (N R McFarland MD, Prof M S Okun MD, Prof D E Vaillancourt), **and Digital Worlds Institute** (A Barmpoutis), **University of Florida, Gainesville, FL, USA; Department of Kinesiology and Applied Physiology, University of Delaware, Newark, DE, USA** (R G Burciu PhD); **Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA** (D M Corcos PhD, Prof C Comella MD); **Department of Radiology** (A S Kurani PhD, Prof T B Parrish PhD) **and Department of Neurology** (E Herschel BS, Prof T Simuni MD), **Northwestern University Feinberg School of Medicine, Chicago, IL, USA; Department of Neurology, University of Chicago, Chicago, IL, USA** (T Xie MD); **Neuroimaging Research Core Facility, Department of Neurology,**

Medical University Innsbruck, Innsbruck, Austria (Prof K Seppi MD, F Krismer MD); **Department of Radiology** (Prof N I Bohnen MD, M L T M Müller PhD) **and Department of Neurology** (Prof N I Bohnen, Prof R L Albin MD), **University of Michigan, Ann Arbor, MI, USA; Neurology Service and Geriatrics Research, Education, and Clinical Center, VA Ann Arbor Healthcare System, Ann Arbor, MI, USA** (Prof N I Bohnen, Prof R L Albin); **Udall Center of Excellence for Parkinson's Disease Research, University of Michigan, Ann Arbor, MI, USA** (Prof N I Bohnen, M L T M Müller, Prof R L Albin); **Department of Neurology** (G Du MD, M M Lewis PhD, Prof X Huang MD), **Department of Pharmacology** (M M Lewis, Prof X Huang), **Department of Neurosurgery** (Prof X Huang), **Department of Radiology** (Prof X Huang), **and Department of Kinesiology** (Prof X Huang), **Penn State Milton S Hershey Medical Center, Hershey, PA, USA; Department of Public Health Sciences, Medical College of South Carolina, Charleston, SC, USA** (H Li PhD); **Department of Psychiatry and Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA** (O Pasternak PhD); **and Department of Neurology** (N R McFarland, Prof M S Okun, Prof D E Vaillancourt), **and Department of Neurosurgery** (Prof M S Okun), **McKnight Brain Institute, University of Florida, Gainesville, FL, USA**

Correspondence to: Prof David E Vaillancourt, Laboratory for Rehabilitation Neuroscience, Department of Applied Physiology and Kinesiology, University of Florida, Gainesville, FL 32611-8205, USA
vcourt@ufl.edu

Research in context

Evidence before this study

Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy are neurodegenerative disorders that are challenging to differentiate in a clinical setting, because they often share motor and non-motor symptoms. Although dopamine transporter imaging can detect nigrostriatal denervation, it cannot distinguish between different forms of parkinsonism. We searched PubMed for original articles published in English, with the terms "Parkinsonism", "Parkinson's disease", "multiple system atrophy", and "progressive supranuclear palsy" combined with "diffusion MRI". Previous studies have shown that diffusion-weighted MRI might be used to distinguish Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy, but they used small samples and were only tested at one imaging site. The main challenge in neuroimaging studies is in evaluating a procedure in multicentre cohorts. Using diffusion-weighted MRI as a method offers unique clinical importance, because it can be done on most 3T scanners worldwide, does not need a contrast drug, and the data can be acquired within a 12 min scan.

Added value of this study

In this study, we created an automated imaging analysis procedure that we tested with data from 1002 patients across 17 MRI centres, making this the largest cohort of parkinsonism evaluated to date. The inputted regions were relevant to parkinsonism, and consisted of regions within the basal ganglia, cerebellum, and cortex. Using a fully automated approach, we found that diffusion-weighted MRI is capable of differentiating Parkinson's disease from multiple system atrophy or progressive supranuclear palsy and multiple system atrophy from progressive supranuclear palsy, with high accuracy across 17 MRI sites. The ten regions with greatest relative importance to the model were those previously shown to be pathologically involved in Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy. In a subset of five cases, the diffusion-weighted MRI diagnosis matched pathological diagnosis. This study developed a region of interest template and three white matter tractography templates, all of which are available to the public, which will expedite future studies in parkinsonism.

Implications of all the available evidence

This study provides an objective, validated, and generalisable imaging approach to distinguish different forms of Parkinsonian syndromes using geographically diverse diffusion-weighted MRI cohorts. Our results are relevant in the clinical setting because they indicate that diffusion-weighted MRI might provide a biomarker for physicians to use in considering a patient to have atypical parkinsonism or Parkinson's disease, and in distinguishing between multiple system atrophy from progressive supranuclear palsy. The outcome of this study suggests that the imaging and machine learning model might function well using data from new sites.

calculation. One study[10] using a single-centre cohort has shown that free-water imaging, a method that allows for the separation of the fluid (ie, free water) and tissue (free-water-corrected fractional anisotropy, FA_T) components in diffusion-weighted MRI can detect unique microstructural changes in different forms of parkinsonism. Several studies[10,31,32,33] have shown elevated free water within the posterior substantia nigra in patients with Parkinson's disease and in widespread, but distinct, networks in patients with multiple system atrophy or progressive supranuclear palsy. Although these studies show promise for detecting objective and unique diffusion measurements in parkinsonism, they used manual region delineation, small cohorts, and diffusion-weighted MRI data acquired from one MRI scanner. Development of a fully automated, generalisable procedure that differentiates parkinsonian syndromes by use of diffusion measurements from pathologically relevant regions is a crucial need for the field. Therefore, we aimed to create and validate an objective and generalisable biomarker to differentiate parkinsonian syndromes.

For **FMRIB Software Library** see http://www.fmrib.ox.ac.uk/fsl/

Methods

Data sources

We did an international study at 17 MRI centres in Austria, Germany, and the USA. Imaging data were obtained from eight cohorts using 17 different MRI scanners. Patients with parkinsonism had been diagnosed by a movement disorder specialist using standard diagnostic criteria.[1,2,7,11] Patients with probable diagnosis, available diffusion-weighted MRI data, and available Movement Disorders Society Unified Parkinson's Disease Rating Scale part III (MDS-UPDRS III) data were included. All progressive supranuclear palsy subtypes were included, but we did not delineate subtypes. The controls reported no history of neuropsychiatric or neurological problems. Disease severity was assessed using the MDS-UPDRS III or the Unified Parkinson's Disease Rating Scale (UPDRS). UPDRS scores were converted to MDS-UPDRS III scores using established guidelines.[12] Patients provided written informed consent for all procedures and post-mortem analyses, which were approved by institutional review boards.

Data preprocessing and normalisation

The primary outcomes were free water and free-water-corrected fractional anisotropy across 60 different template regions. FMRIB Software Library and custom UNIX shell scripts were used to preprocess the data.[13] Data preprocessing with custom MATLAB scripts was done for all datasets to obtain free water and FA_T images for each individual.[14] Preprocessing was identical to

previous work from our group.[10] Quality control was done by visually inspecting each individual free-water and FA_T map. If the field of view did not encompass the whole brain or was distorted they were not included in this study (<1% of the total data). Because partial brains and distortions were excluded from our analysis, missing values did not need to be imputed for any region or tract used in the analysis.

We did a literature review and analysis (appendix pp 3–6) that determined an optimal, automated normalisation pipeline for diffusion-weighted MRI data. We compared free-water values obtained from hand-drawn regions of interest to template-derived free-water values in a subset of patients from the University of Florida (n=104, including controls, Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy) by use of four different normalisation pipelines. The Advanced Normalization Tools software provided the highest intraclass correlation coefficients between hand-drawn and template-derived free-water values. Thus, free water and FA_T were normalised by use of this automated pipeline for all subsequent analyses.[15]

Region and tracts of interest

We created a Parkinson's disease region of interest template in Montreal Neurological Institute (MNI) space (figure 1A), which included 17 regions in the basal ganglia (anterior substantia nigra, posterior substantia nigra, subthalamic nucleus, globus pallidus, putamen, and caudate nucleus), midbrain, thalamus, and cortex (premotor corpus callosum, prefrontal corpus callosum, pedunculopontine nucleus, red nucleus, and thalamus), and cerebellum (middle cerebellar peduncle, superior cerebellar peduncle, inferior vermis, dentate nucleus, cerebellar lobule V, and cerebellar lobule VI).

43 white matter tracts were also used (figure 1B). We did probabilistic tractography identically to our previous work[16] with the Human Connectome Project dataset to characterise the subthalamo-pallidal, nigrostriatal, and corticostriatal tracts. We also incorporated several existing tractography templates, which include the sensorimotor area tract template (SMATT),[16] the transcallosal tractography template (TCATT), and a cerebellar white matter atlas.[17] The SMATT includes six different sensorimotor tracts, including the tracts descending from the primary motor cortex, dorsal premotor cortex, ventral premotor cortex, supplemental motor area, pre-supplemental motor area, and somatosensory cortex. The TCATT includes five parietal, six occipital, six frontal, and 12 prefrontal commissural tracts.[18] We also used the superior and middle cerebellar tracts from the cerebellar probabilistic white matter atlas.[17] Free water and FA_T were calculated for each region and tract separately for each participant.

Machine learning approach

The variables available in this study include diffusion-weighted MRI values (free water and FA_T in 17 regions and 43 tracts), MDS-UPDRS III, sex, and age. Three different combinations of variables were created: diffusion-weighted MRI plus MDS-UPDRS III, diffusion-weighted MRI only, and MDS-UPDRS III only. Age and sex were included in all analyses, because they are variables that would be readily available for input in regular practice. Diffusion-weighted MRI variables include free-water and FA_T values in each region and tract, resulting in a total of 60 FA_T measurements and 60 free-water measurements for each individual. Each combination of variables was used in the training and validation of a support vector machine (SVM) learning algorithm using a linear kernel in the scikit-learn package in Python.[19] We chose the SVM algorithm, because it is a widely accepted and robust machine learning model for classification. Compared

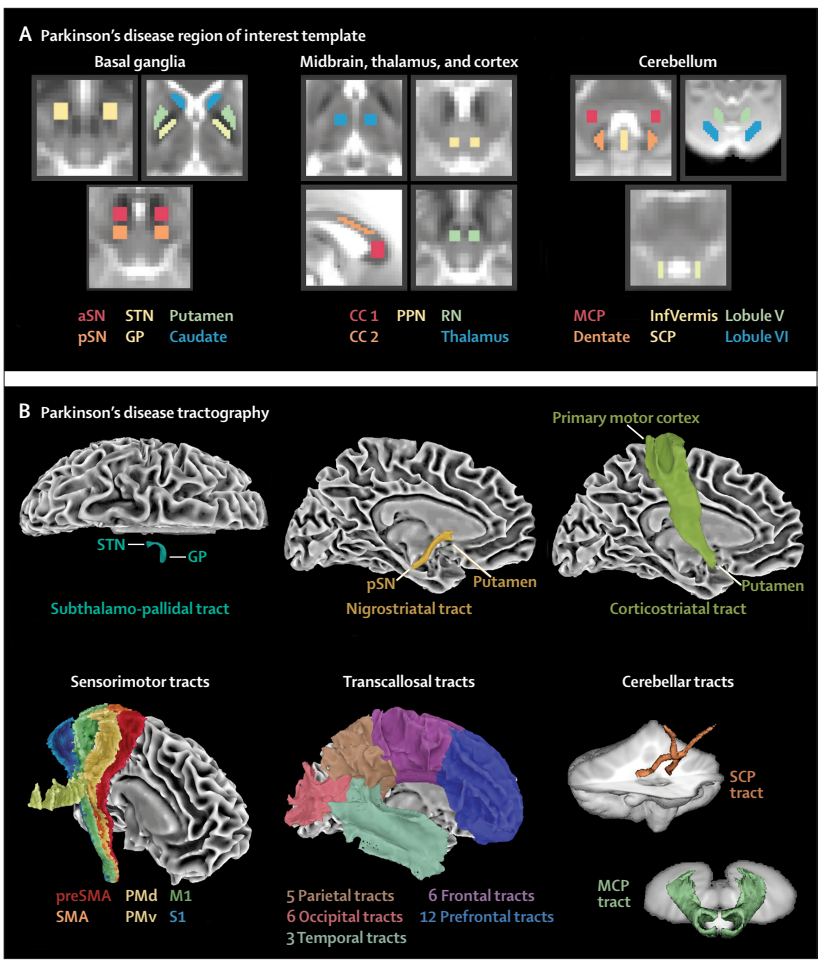

Figure 1: **Regions and tracts of interest**
(A) Regions of interest covered many areas of the basal ganglia (column 1), midbrain, thalamus, and cortex (column 2), and cerebellum (column 3). (B) Tracts of interest included a subthalamo-pallidal tract, nigrostriatal tract, and corticostriatal tract as well as six sensorimotor tracts from the sensorimotor area tract template, 32 transcallosal tracts from the transcallosal tractography template, and two cerebellar tracts. aSN=anterior subtantia nigra. CC 1=prefrontal corpus callosum. CC 2=premotor corpus callosum. GP=globus pallidus. InfVermis=inferior cerebellar vermis. M1=primary motor cortex. MCP=middle cerebellar peduncle. PMd=dorsal premotor area. PMv=ventral premotor area. PPN=pedunculopontine nucleus. preSMA=pre-supplemental motor area. pSN=posterior subtantia nigra. RN=red nucleus. S1=somatosensory cortex. SCP=superior cerebellar peduncle. SMA=supplemental motor area. STN=subthalamic nucleus.

See Online for appendix

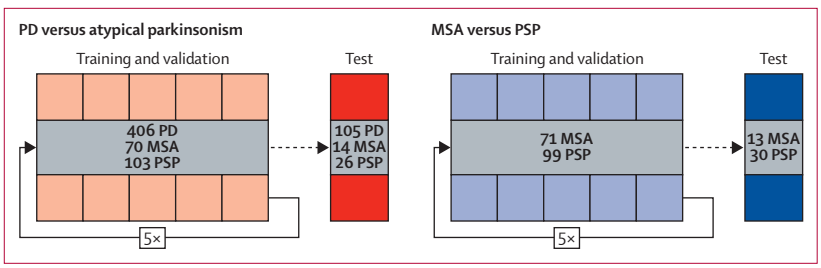

Figure 2: **Machine learning procedure**
The PD versus atypical parkinsonism and MSA versus PSP comparisons were first split into training and validation (80%) and test (20%) cohorts using random sampling. Five-fold cross-validation was done independently for both comparisons to tune the hyperparameters in the support vector machine learning analysis. Following hyperparameter tuning, this model was applied to the test cohort to evaluate performance. PD=Parkinson's disease. MSA=multiple system atrophy. PSP=progressive supranuclear palsy.

with deep learning methods, such as convolutional neural networks, and conventional machine learning algorithms, such as decision trees and random forests, SVM has the advantage of high performance at low computational cost. It also does not need a large amount of labelled data. SVM works by separating the data using a hyperplane in the feature space. A linear kernel indicates that the data points are represented using its original feature space, instead of being projected into a high-dimensional space for classification. We chose to use a linear kernel because it could linearly separate the data in this study with good performance compared with comparable approaches. Furthermore, the development of SVM models using a linear kernel allowed us to extract the coefficient of each region and tract of interest to determine its relative importance to the model. Other more complex kernels, such as polynomial and radial basis function kernels, increased the computational time required to train the models and did not provide additional accuracy. Disease-specific comparisons were made to predict diagnosis (Parkinson's disease *vs* atypical parkinsonism and multiple system atrophy *vs* progressive supranuclear palsy). Training and validation sets for each disease-specific comparison consisted of 80% of the total relevant data, and the remaining 20% was reserved for a test dataset. Patients were randomly assigned to the training and validation set or the test dataset by use of stratified sampling to ensure that the training and validation and test dataset group proportions were equal to the total dataset group proportions (figure 2). In the training and validation cohort, the data was randomly split into five subgroups for five-fold cross-validation. The purpose of the five-fold cross-validation was to optimise the F1 score (ie, the harmonic mean of the precision and recall) by optimising the penalty parameter (a measure that directly represents the tolerance for error) across the five distinct folds. This penalty parameter was used to train the machine learning model with the training and validation dataset and the performance of this optimised model was evaluated on the test dataset.

To evaluate the performance of the machine learning models in the training and validation cohort and the test cohort, we did receiver operating characteristic analyses using the trained models. The area under the curve (AUC) was calculated for each model (diffusion-weighted MRI plus MDS-UPDRS III, diffusion-weighted MRI only, and MDS-UPDRS III only) for each comparison (Parkinson's disease *vs* atypical parkinsonism and multiple system atrophy *vs* progressive supranuclear palsy). The models were statistically evaluated with Delong's test to compare AUCs between receiver operating characteristics.[20] We also calculated several measures from the confusion matrix, including accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. We also evaluated the pathophysiological relevance of the models by relating feature importance to the absolute value of the coefficients of the hyperplane that defines the optimised SVM model. To determine whether between-site effects had an effect on the machine learning performance, we also did secondary machine learning analyses, in which we harmonised the diffusion-weighted MRI data with the ComBat batch-effect correction tool, because it has been shown to be effective in correcting multisite diffusion-weighted MRI data.[21]

Comparison of diagnoses

We obtained data from five patients (three from University of Florida and two from Penn State Hershey Medical Center) who received in-vivo diffusion-weighted MRI and post-mortem neuropathological examination. The diffusion-weighted MRI data for these patients was inputted into the automated imaging differentiation in parkinsonism (AID-P; Parkinson's disease *vs* atypical parkinsonism and multiple system atrophy *vs* progressive supranuclear palsy) to determine group probabilities. These probabilities were used to classify patients and compared with neuropathological diagnosis.

Role of the funding source

The funders in this study had no role in the study design or data collection, analysis, or interpretation. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Data on 1002 individuals from between Dec 9, 2008, and Nov 29, 2018, were obtained. 278 (28%) of these individuals were healthy controls, 511 (51%) had Parkinson's disease, 84 (8%) had multiple system atrophy, and 129 (13%) had progressive supranuclear palsy (table 1). Several phenotypes of multiple system atrophy exist, including parkinsonian and cerebellar. 80 (95%) of 84 individuals with multiple system atrophy cohort had parkinsonian multiple system atrophy.

The means and SDs for each machine learning input feature for each group (control, Parkinson's disease, multiple system atrophy, and progressive supranuclear

	University of Florida cohort 1, n=169	University of Florida cohort 2, n=100	Penn State Hershey Medical Center, n=104	Medical University Innsbruck, n=180	Northwestern University, n=55	University of Michigan, n=206	Parkinson's Progression Marker's Initiative, n=150	4 Repeat Tauopathy Neuroimaging Initiative, n=38	Total, n=1002
Cohort composition, n (%)									
Control	46 (27%)	11 (11%)	36 (35%)	48 (27%)	12 (22%)	76 (37%)	49 (33%)	..	278 (28%)
Parkinson's disease	59 (35%)	72 (72%)	35 (34%)	91 (51%)	24 (44%)	130 (63%)	100 (67%)	..	511 (51%)
Multiple system atrophy	32 (19%)	9 (9%)	15 (14%)	20 (11%)	7 (13%)	..	1 (<1%)	..	84 (8%)
Progressive supranuclear palsy	32 (19%)	8 (8%)	18 (17%)	21 (12%)	12 (22%)	..	..	38 (100%)	129 (13%)
Age (years), mean (SD)	65·24 (9·20)	65·22 (7·92)	70·68 (8·15)	63·46 (10·57)	67·25 (7·58)	65·61 (9·34)	59·96 (9·69)	69·76 (7·60)	65·05 (9·65)
Sex, n (%)									
Male	98 (58%)	58 (58%)	68 (65%)	106 (59%)	30 (55%)	130 (63%)	99 (66%)	19 (50%)	608 (61%)
Female	71 (42%)	42 (42%)	36 (35%)	74 (41%)	25 (45%)	76 (37%)	51 (34%)	19 (50%)	394 (39%)
Recruitment start date	Jan 31, 2012	April 13, 2017	Nov 24, 2009	July 27, 2011	Nov 16, 2017	Dec 9, 2008	Oct 27, 2010	June 21, 2011	..
Recruitment end date	Sep 18, 2014	Nov 29, 2018	Aug 25, 2015	April 4, 2013	Nov 29, 2018	June 18, 2014	May 28, 2013	Dec 15, 2014	..
Disease duration (years), mean (SD)									
Parkinson's disease	3·22 (2·34)	2·56 (2·39)	3·37 (3·61)	6·02 (4·27)	4·05 (2·44)	6·00 (4·21)	0·60 (0·59)	..	3·87 (3·84)
Multiple system atrophy	3·73 (3·05)	2·17 (1·49)	3·60 (3·27)	2·03 (2·26)	1·86 (1·85)	..	0·20	..	2·94 (2·77)
Progressive supranuclear palsy	3·15 (2·95)	2·15 (1·59)	2·83 (2·53)	1·74 (1·22)	3·60 (3·13)	..	..	5·22 (3·88)	3·45 (3·16)
Movement Disorders Society Unified Parkinson's Disease Rating Scale part III, mean (SD)									
Control	3·07 (2·82)	6·00 (3·10)	6·08 (7·59)	3·00 (7·17)	2·92 (1·38)	4·34 (4·96)	0·35 (0·69)	..	3·42 (5·26)
Parkinson's disease	26·19 (10·98)	30·72 (11·68)	38·17 (28·42)	36·75 (12·59)	31·00 (13·13)	32·60 (14·27)	19·9 (8·92)	..	30·15 (15·02)
Multiple system atrophy	46·25 (16·88)	64·22 (14·00)	49·60 (19·54)	52·48 (10·97)	59·71 (21·87)	..	..	..	51·13 (17·17)
Progressive supranuclear palsy	45·19 (18·56)	55·63 (19·40)	45·44 (24·20)	34·02 (18·10)	46·67 (15·75)	..	..	33·16 (17·42)	40·65 (19·79)
Sites	1	1	1	1	1	1	10	1	17
MRI strength and vendor	3T, Philips	3T, Siemens	3T, Siemens	3T, Siemens	3T, Siemens	3T, Philips	3T, Siemens	3T, Siemens	..
Directions	64	64	42	20	64	15	64	41	..
B-values (s/mm^2)	0–1000	0–1000	0–1000	0–1000	0–1000	0–800	0–1000	0–1000	..
Resolution (mm × mm × mm)	2 × 2 × 2	2 × 2 × 2	2 × 2 × 2	2 × 2 × 3	2 × 2 × 2	1·75 × 1·75 × 1·75	2 × 2 × 2	2·7 × 2·7 × 2·7	..
Echo time (ms)	86	58	82	83	58	67	86*	82	..
Repetition time (ms)	7748	6400	8300	8200	6400	8044	2223*	9200	..

*Because the Parkinson's Progression Marker's Initiative cohort had variable repetition times, the median is reported.

Table 1: **Baseline characteristics for the eight cohorts**

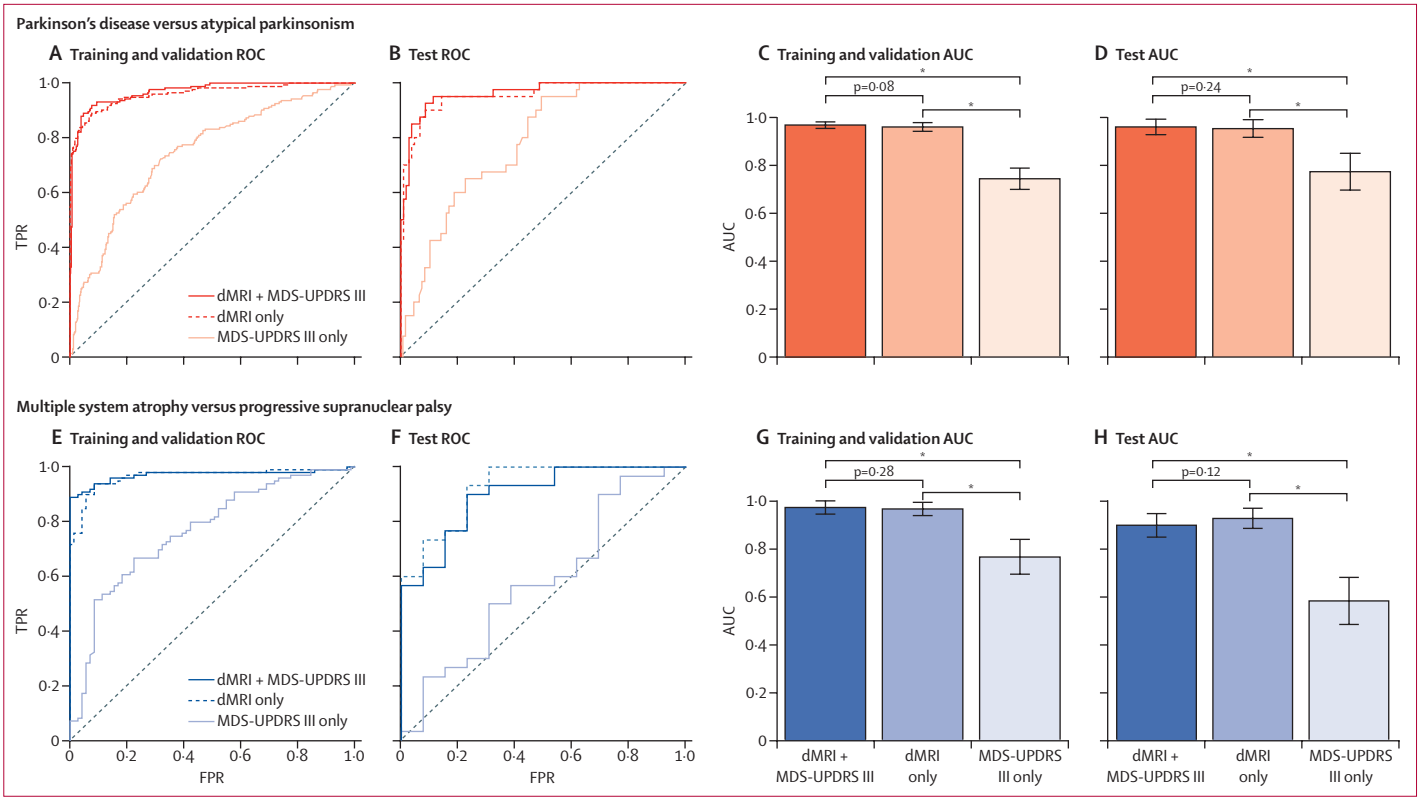

Figure 3: **Machine learning performance**
Receiver operating characteristic analyses (A, B) and corresponding AUC (C, D) for each model in the Parkinson's disease versus atypical parkinsonism comparison for the training and validation and test cohorts are shown. ROC analyses (E, F) and corresponding areas under the curve (G, H) for each model in the multiple system atrophy versus progressive supranuclear palsy comparison for the training and validation and test cohorts are also shown. For each comparison, Delong's test was done to determine between-model differences. Bars represent mean with 95% CI error bars. AUC=area under the curve. dMRI=diffusion-weighted MRI. FPR=false positive rate. MDS-UPDRS III=Movement Disorders Society Unified Parkinson's Disease Rating Scale part III. TPR=true positive rate. ROC=receiver operating characteristic. *p<0·0001.

	Diffusion-weighted MRI plus MDS-UPDRS III		Diffusion-weighted MRI only		MDS-UPDRS III only	
	Training and validation, mean (95% CI)	Test	Training and validation, mean (95% CI)	Test	Training and validation, mean (95% CI)	Test
Parkinson's disease versus atypical parkinsonism						
Area under the curve	0·969 (0·955–0·982)	0·962	0·961 (0·943–0·979)	0·955	0·745 (0·701–0·789)	0·775
Accuracy (%)	85·06% (82·57–87·55)	91·49	85·21% (83·43–86·98)	90·24	63·00% (59·32–66·68)	66·61
Sensitivity (%)	79·23% (74·16–84·30)	92·5	78·05% (74·65–81·45)	90·00	41·53% (32·90–50·16)	67·50
Specificity (%)	90·89% (88·38–93·40)	90·48	92·37% (89·28–95·45)	90·48	84·47% (78·46–90·48)	65·71
Positive predictive value (%)	89·79% (87·16–92·14)	90·67	91·29% (88·07–94·51)	90·43	73·47% (66·24–80·70)	66·32
Negative predictive value (%)	81·64% (77·92–85·36)	92·35	90·91% (88·55–93·27)	90·05	59·36% (56·13–62·59)	66·91
Multiple system atrophy versus progressive supranuclear palsy						
Area under the curve	0·971 (0·943–0·998)	0·897	0·965 (0·937–0·992)	0·926	0·765 (0·693–0·838)	0·582
Accuracy (%)	87·99% (84·23–91·74)	80·13	84·85% (79·65–90·05)	81·79	69·97% (62·07–77·87)	50·90
Sensitivity (%)	85·89% (79·50–92·28)	83·33	83·89% (77·52–90·26)	86·67	66·79% (57·57–76·01)	63·33
Specificity (%)	90·10% (86·96–93·24)	76·92	85·81% (80·11–91·51)	76·92	73·14% (62·24–84·04)	38·46
Positive predictive value (%)	89·70% (86·65–92·75)	78·31	85·64% (80·18–91·10)	78·97	72·35% (62·10–82·60)	50·72
Negative predictive value (%)	86·83% (81·57–92·09)	82·19	84·41% (78·79–90·03)	85·23	68·96% (61·15–76·77)	51·19

MDS-UPDRS III=Movement Disorders Society Unified Parkinson's Disease Rating Scale part III.

Table 2: **Support vector machine learning classification discriminative measures**

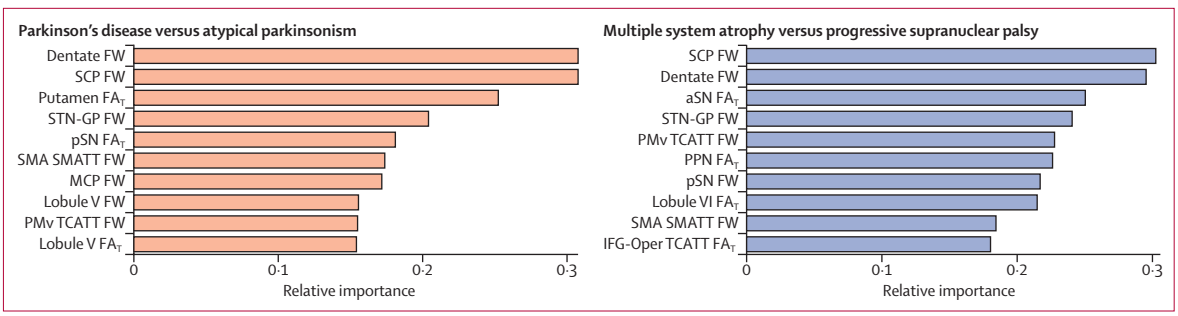

Figure 4: **Machine learning pathophysiological relevance**

The ten regions with the greatest relative importance to the diffusion-weighted MRI only model for the Parkinson's disease versus atypical parkinsonism and multiple system atrophy versus progressive supranuclear palsy comparisons. Relative importance is the absolute value of the attributes weight in the orthogonal vector that defines the hyperplane separating positive and negative class predictions. An attribute with higher relative importance thus contributes more when determining the classification of a new data point. aSN=anterior substantia nigra. FA_T=free-water-corrected fractional anisotropy. FW=free water. IFG-Oper TCATT=inferior frontal gyrus pars opercularis transcallosal tract template. MCP=middle cerebellar peduncle. PMv TCATT=ventral premotor cortex transcallosal tract template. PPN=pedunculopontine nucleus. SCP=superior cerebellar peduncle. pSN=posterior substantia nigra. SMA SMATT=supplemental motor area descending motor tract template. STN-GP=subthalamo-pallidal tract.

palsy) can be found in the appendix (pp 9–11). Although we did machine learning on both unharmonised and harmonised diffusion-weighted MRI data, the receiver operating characteristic curves were similar (appendix p 14).

To evaluate the performance of the Parkinson's disease versus atypical parkinsonism machine learning models (diffusion-weighted MRI plus MDS-UPDRS III, diffusion-weighted MRI only, and MDS-UPDRS III only), we did receiver operating characteristic analyses in the training and validation (figure 3A,C) and test cohorts (figure 3B,D). The diffusion-weighted MRI plus MDS-UPDRS III model show large AUCs in both the training and validation (0·969, 95% CI 0·955–0·982) and test (0·962) cohorts (table 2). Similarly, the diffusion-weighted MRI only model showed large AUCs in the training and validation (0·961, 95% CI 0·943–0·979) and test (0·955) cohorts; however, the MDS-UPDRS III only model had substantially smaller AUCs (training and validation 0·745, 95% CI 0·701–0·789; test 0·775; table 2). In both the training and validation and test cohorts, significant differences existed between the diffusion-weighted MRI plus MDS-UPDRS III and MDS-UPDRS III only models and between the diffusion-weighted MRI only and MDS-UPDRS III only models ($p<0·0001$). The diffusion-weighted MRI plus MDS-UPDRS III and diffusion-weighted MRI only models did not differ. An analysis of variable relative importance in the diffusion-weighted MRI only model revealed that the ten features with the greatest relative importance to the model included many regions previously shown to be pathologically involved in parkinsonism (figure 4).[22] Additional machine learning performance metrics, including accuracy, sensitivity, specificity, positive predictive value, and negative predictive value, were calculated (table 2). Of note, one contributing cohort (University of Michigan) used a smaller number of directions than the other cohorts, was the only centre that had a b-value of 800 s/mm², and did not have any participants with multiple system atrophy or progressive supranuclear palsy; thus, this cohort could have skewed results in this analysis. For this reason, we removed this cohort and did an identical machine learning analysis. Even with the removal of this group, we obtained comparable AUCs in the diffusion-weighted MRI plus MDS-UPDRS III (training and validation 0·956, 95% CI 0·937–0·975; test 0·926), diffusion-weighted MRI only (training and validation 0·945, 0·923–0·968; test 0·899), and MDS-UPDRS only (training and validation 0·763, 0·719–0·807; test 0·666) models.

Identical to the Parkinson's disease versus atypical parkinsonism models, the multiple system atrophy versus progressive supranuclear palsy models (figure 3E–H) had large AUCs in both the diffusion-weighted MRI plus MDS-UPDRS III (training and validation 0·971, 95% CI 0·943–0·998; test 0·897) and diffusion-weighted MRI only models (training and validation 0·965, 95% CI 0·937–0·992; test 0·926), whereas the MDS-UPDRS III only model exhibited lower AUCs (training and validation 0·765, 95% CI 0·693–0·838; test 0·582; table 2). The models were compared with Delong's test, in which significant differences existed between the diffusion-weighted MRI plus MDS-UPDRS III and MDS-UPDRS III only and between the diffusion-weighted MRI only and MDS-UPDRS III only models ($p<0·0001$). The diffusion-weighted MRI plus MDS-UPDRS III and diffusion-weighted MRI only models did not differ. An analysis of relative importance in the diffusion-weighted MRI model revealed that the ten variables with the greatest relative importance to the model included many regions shown to be pathologically involved in atypical parkinsonism (figure 4).[22] Additional machine learning performance metrics, including accuracy, sensitivity, specificity, positive predictive value, and negative predictive value, were calculated (table 2).

Secondary disease-specific comparisons were also developed, including control versus parkinsonism, multiple system atrophy versus Parkinson's disease

	Clinical	Pathological	AID-P: Parkinson's disease *vs* atypical parkinsonism (group probability)	AID-P: multiple system atrophy *vs* progressive supranuclear palsy (group probability)
1	Progressive supranuclear palsy	Progressive supranuclear palsy	Atypical parkinsonism (0·962)	Progressive supranuclear palsy (0·805)
2	Progressive supranuclear palsy	Progressive supranuclear palsy	Atypical parkinsonism (1·000)	Progressive supranuclear palsy (0·685)
3	Progressive supranuclear palsy	Progressive supranuclear palsy	Atypical parkinsonism (0·955)	Progressive supranuclear palsy (0·947)
4	Multiple system atrophy	Progressive supranuclear palsy	Atypical parkinsonism (0·885)	Progressive supranuclear palsy (0·927)
5	Multiple system atrophy	Multiple system atrophy	Atypical parkinsonism (0·602)	Multiple system atrophy (0·778)

Table 3: **Application of AID-P to pathological diagnosis in five patients**

or progressive supranuclear palsy, and progressive supranuclear palsy versus Parkinson's disease or multiple system atrophy (appendix p 8). Performance metrics for these machine learning comparisons and Delong's comparison of model performance (diffusion-weighted MRI plus MDS-UPDRS III, diffusion-weighted MRI only, and MDS-UPDRS III only) for each disease-specific model are in the appendix (pp 12–13).

The diffusion-weighted MRI data in 5 patients who were followed to post mortem were used to compare clinical, pathological, and machine learning (ie, diffusion-weighted MRI only AID-P) ability to predict disease state (table 3). In four of five patients, the clinical, pathological, and AID-P diagnosis matched. In one patient, the clinical diagnosis was multiple system atrophy, but a pathological examination of this individual revealed a progressive supranuclear palsy diagnosis. The AID-P matched the pathological examination and predicted this patient to have a progressive supranuclear palsy diagnosis. Thus, the AID-P accurately predicted five of five cases at pathology, but more cases are needed to generalise these findings.

Discussion

This study provides a completely automated, objective, validated, and generalisable imaging approach to distinguish different forms of parkinsonian syndromes using geographically diverse diffusion-weighted MRI cohorts. The strength of this approach was that it was effective across a range of MRI platforms using standard diffusion-weighted MRI sequences, indicating the effect it could have in clinical trials and clinical care for parkinsonism.

A difficult problem confronting neurologists in parkinsonism is the correct classification into idiopathic Parkinson's disease or atypical parkinsonism. Misclassification is common, particularly early in the disease,[5] which can reduce the efficacy of clinical trials that only recruit patients with Parkinson's disease. Pathological examination is required for final confirmation of a clinical diagnosis and cannot be obtained in vivo, when participants are being recruited and selected for clinical trials. This study suggests that diffusion-weighted MRI, specifically free-water imaging, could be used as a marker for detecting structural deficits and is accurate in differentiating degenerative parkinsonisms. Furthermore, our results indicate that diffusion-weighted MRI outperformed the MDS-UPDRS III in differentiating parkinsonian syndromes. Our findings in this large multicentre dataset were also capable of distinguishing degenerative parkinsonian syndromes with high accuracy, sensitivity, and specificity.

Several machine learning studies[23–26] have been done using a variety of neuroimaging methods, including structural T1, ^{18}F-fluorodeoxyglucose-PET, and diffusion-weighted MRI. One study[27] did blood-based neurofilament light chain analyses and successfully separated Parkinson's disease from atypical parkinsonism with a fair amount of accuracy, but this measure was unable to separate multiple system atrophy and progressive supranuclear palsy. Our automated pipeline (the AID-P) builds upon these studies and incorporates several novel features, including the largest cohort of patients to date, a procedure that is translatable across a large number of scanners using a single pipeline, a completely automated approach, and the inclusion of a distinct test cohort. In the test cohort, we obtained excellent classification (classification accuracies, sensitivities, and specificities comparable to that of current gold standards in parkinsonism) of both Parkinson's disease versus atypical parkinsonism and multiple system atrophy versus progressive supranuclear palsy using only diffusion-weighted MRI variables as input. The performance in the test dataset was comparable to that of the training and validation dataset, suggesting that our cross-validation approach did not overfit the data and could be generalisable to new datasets. Because site was not added as a covariate in the model, the data reported here might generalise to new cohorts on a patient-by-patient basis, because the model would not need substantial data from a specific site to produce a predictive value. One caveat is that the diffusion-weighted MRI data collection parameters would need to be consistent with those in table 1, but these scanning parameters are available on most current 3T MRI scanners as a routine sequence.

A potential limitation of this study is that expert clinical diagnosis, not pathological diagnosis, was used to classify patients; however, because the patients with parkinsonism had relatively advanced disease severity, misdiagnosis is less probable. Three cohorts (University of Michigan, Parkinson's Progression Marker's Initiative, and 4 Repeat Tauopathy Neuroimaging Initiative) included radiotracer

scans confirming that a dopaminergic deficit exists. Future evaluation of this method should involve patients with lower disease severity and other diseases that are often misdiagnosed as parkinsonisms, including dementia with Lewy bodies and corticobasal degeneration.[3] A longitudinal assessment of the AID-P would be particularly useful in determining how well it predicts subsequent pathology in a large cohort. Moreover, the sample size of multiple system atrophy and progressive supranuclear palsy patients will need to be larger in future machine learning models, which should increase the accuracy of the method. The relative ease of obtaining diffusion-weighted MRI data will facilitate these kinds of studies. In the interim, if the AID-P tool is used to classify patients in clinical research, the recommended approach would be to use it in conjunction with expert clinical evaluations.

In each machine learning model, a weight was given to each of the regions or tracts. The regions and tracts with the highest weights contribute the most to classifying patients in the model. How the regions identified by the model might compare with previous neuropathological evidence in the published literature needs to be determined. In the Parkinson's disease versus atypical parkinsonism model, the features with the greatest relative importance to the model included free water in the dentate nucleus, superior cerebellar peduncle, subthalamo-pallidal tract, and FA_T in the putamen. In the multiple system atrophy versus progressive supranuclear palsy model, features with the greatest relative importance to the model were free water in the superior cerebellar peduncle, dentate nucleus, and subthalamo-pallidal tract, and FA_T in the anterior substantia nigra. These results largely agree with pathological reports in multiple system atrophy and progressive supranuclear palsy.[28,29] In multiple system atrophy, neuropathology is exhibited as glial cytoplasmic inclusions in the nigrostriatal and olivopontocerebellar pathways;[29] whereas, neuropathology in progressive supranuclear palsy is exhibited as neurofibrillary tangles or atrophy in regions that include the basal ganglia, frontal cortex, midbrain, and the superior cerebellar peduncle.[28] Although we didn't directly evaluate the neuropathological characteristics in all patients in our cohort, we tested models created from a large dataset of patients with parkinsonism on five patients with post-mortem confirmed diagnosis. The AID-P performed with 100% accuracy, sensitivity, and specificity in these cases (table 3). Furthermore, in one patient, the AID-P agreed with the post-mortem diagnosis, whereas the clinical diagnosis did not. This example illustrates the clinical utility of the AID-P.

In summary, the AID-P provides an automated, objective, validated, and generalisable imaging approach to distinguish different forms of parkinsonian syndromes. By use of diffusion-weighted MRI datasets obtained from 17 different MRI scanners and eight geographically diverse cohorts in conjunction with 60 anatomical region or tract templates in MNI space, this study provides strong evidence that diffusion-weighted MRI alone can assist in the diagnosis and differentiation of different forms of parkinsonism. Future work is needed to assess the AID-P from other cohorts across the globe, and the creation of a software platform using cloud computing will facilitate international use of the AID-P. Furthermore, future studies could implement automated quality control steps, which would eliminate the need to visually inspect each patient's diffusion-weighted MRI map in MNI space.[30] This imaging method does not involve radioactive tracers, and the scan can be collected in less than 12 min on 3T scanners worldwide. The results of this study suggest that the AID-P might function well using data from new sites that have incorporated pulse sequences consistent with the MRI scanners used in this study.

Contributors

DBA and DEV did the literature search, designed the study, and produced figures. JLM, SL, ASK, MLB, EH, TS, TBP, CC, TX, KS, NIB, MLTMM, RLA, FK, GD, MML, XH, HL, NRM, and MSO collected the data. DBA, JTB, WTC, RGB, JLM, SAC, RF, AB, DMC, TM, HL, OP, and DEV analysed and interpreted the data. DBA, JTB, WTC, RGB, SAC, RF, AB, DMC, ASK, TM, TS, TBP, CC, TX, KS, NIB, MLTMM, RLA, FK, GD, MML, XH, HL, OP, NRM, MSO, and DEV wrote the manuscript.

Declaration of interests

DBA reports grants from National Institutes of Health (NIH) and Parkinson's Foundation, during the conduct of the study. JTB reports personal fees from NIH, during the conduct of the study. EH reports grants and personal fees from NIH for payment of salary, during the conduct of the study. TS reports consulting fees from Acadia, AbbVie, Amneal, Allergan, Acorda Therapeutics, Aptinyx, Denali, General Electric, Neuroderm, Neurocrine, Sanofi, Sinopia, Sunovion, TEVA, Takeda, Voyager, and US World Meds, during the conduct of the study. CC reports grants from NIH, Dystonia Medical Research Foundation, Merz Pharmaceutical, Revance Therapeutic, Retrophin, Acorda Therapeutics, and Parkinson's Foundation, during the conduct of the study, and personal fees from Acorda Therapeutics, Allergan, Lundbeck, Merz Pharmaceuticals, Acadia Pharmaceuticals, Jazz Pharmaceuticals, Neurocrine Biosciences, Revance Therapeutic, Sunovion, and AEON Biopharma, outside the submitted work. CC also received royalties from Cambridge and Wolters Kluwer. KS reports grants from FWF Austrian Science Fund, Michael J Fox Foundation, and International Parkinson and Movement Disorder Society and personal fees from Teva, UCB, Lundbeck, AOP Orphan Pharmaceuticals, AbbVie, Roche, and Grünenthal, outside the submitted work. NIB reports grants from NIH, Department of Veterans Affairs, Michael J Fox Foundation, Eisai, Axovant Sciences, and Chase Pharmaceuticals, outside the submitted work. MLTMM reports grants from NIH, Michael J Fox Foundation, and Department of Veterans Affairs, during the conduct of the study. FK reports grants from MSA Coalition, outside the submitted work. GD reports grants from National Institute of Neurological Disorders and Stroke (NINDS), NIH, and Michael J Fox Foundation, during the conduct of the study, and grants from NINDS, National Institute of Environmental Health Sciences (NIEHS), Department of Defense, and Michael J Fox Foundation, outside the submitted work. GD also has a patent pending (US 62/638,628). MML reports grants from NINDS, NIH, and Michael J Fox Foundation, during the conduct of the study, and grants from NINDS, NIEHS, Department of Defense, Michael J Fox Foundation, Biogen, and Pfizer outside the submitted work. In addition, MML also has a patent pending (US 62/638,628). XH reports grants from NINDS, NIH, and Michael J Fox Foundation, during the conduct of the study, and grants from NINDS, NIEHS, Department of Defense, Michael J Fox Foundation, Biogen, and Pfizer, outside the submitted work. XH also has a patent pending (US 62/638,628). OP reports grants from NIH and personal fees from University of Florida, during the conduct of the study. NRM reports personal fees from AbbVie, outside

the submitted work. MSO serves as a consultant for the National Parkinson Foundation and has received research grants from NIH, Parkinson's Foundation, Michael J Fox Foundation, Parkinson Alliance, Smallwood Foundation, Bachmann-Strauss Foundation, Tourette Syndrome Association, and UF Foundation. MSO's deep brain stimulation research is supported by grants from NIH (R01 NR014852 and R01NS096008). MSO has previously received honoraria, but in the past 60 months has received no support from industry. MSO has received royalties for publications with Demos, Manson, Amazon, Smashwords, Books4Patients, and Cambridge (movement disorders books). MSO is an associate editor for New England Journal of Medicine Journal Watch Neurology. MSO has participated in Continuing Medical Education and educational activities on movement disorders in the last 36 months sponsored by PeerView, Prime, QuantiaMD, WebMD, Medicus, MedNet, Henry Stewart, and Vanderbilt University. MSO's institution receives grants from Medtronic, AbbVie, Allergan, and ANS-St Jude, and MSO has no financial interest in these grants. MSO has participated as a site principal investigator or co-investigator for several NIH and industry sponsored trials over the years but has not received honoraria. DEV reports grants from NIH (R01 NS075012, R01 NS052318, U01 NS102038, T32 NS082169, and R01 NS058487), outside the submitted work. DEV also has a patent pending for diffusion imaging in Parkinson's disease and parkinsonism (WO2018194778A1) pending. All other authors declare no competing interests.

Data sharing

Study protocol and de-identified data are available from the investigators upon request. Requests should be sent to the corresponding author by email.

Acknowledgments

This study was funded by National Institutes of Health (U01 NS102038, R01 NS052318) and Parkinson's Foundation (PF-FBS-1778).

References

1 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.

2 Litvan I, Goetz CG, Jankovic J, et al. What is the accuracy of the clinical diagnosis of multiple system atrophy? A clinicopathologic study. *Arch Neurol* 1997; **54:** 937–44.

3 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:** 861–70.

4 Beach TG, Adler CH. Importance of low diagnostic accuracy for early Parkinson's disease. *Mov Disord* 2018; **33:** 1551–54.

5 Adler CH, Beach TG, Hentz JG, et al. Low clinical diagnostic accuracy of early vs advanced Parkinson disease: clinicopathologic study. *Neurology* 2014; **83:** 406–12.

6 Jankovic J, Rajput AH, McDermott MP, Perl DP. The evolution of diagnosis in early Parkinson disease. *Arch Neurol* 2000; **57:** 369–72.

7 Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992; **55:** 181–84.

8 Perlmutter JS, Norris SA. Neuroimaging biomarkers for Parkinson disease: facts and fantasy. *Ann Neurol* 2014; **76:** 769–83.

9 Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. 1996. *J Magn Reson* 2011; **213:** 560–70.

10 Planetta PJ, Ofori E, Pasternak O, et al. Free-water imaging in Parkinson's disease and atypical parkinsonism. *Brain* 2016; **139:** 495–508.

11 Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* 2008; **71:** 670–76.

12 Goetz CG, Stebbins GT, Tilley BC. Calibration of unified Parkinson's disease rating scale scores to Movement Disorder Society-unified Parkinson's disease rating scale scores. *Mov Disord* 2012; **27:** 1239–42.

13 Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. *Neuroimage* 2012; **62:** 782–90.

14 Pasternak O, Sochen N, Gur Y, Intrator N, Assaf Y. Free water elimination and mapping from diffusion MRI. *Magn Reson Med* 2009; **62:** 717–30.

15 Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 2009; **46:** 786–802.

16 Archer DB, Vaillancourt DE, Coombes SA. A template and probabilistic atlas of the human sensorimotor tracts using diffusion MRI. *Cereb Cortex* 2018; **28:** 1685–99.

17 van Baarsen KM, Kleinnijenhuis M, Jbabdi S, Sotiropoulos SN, Grotenhuis JA, van Cappellen van Walsum AM. A probabilistic atlas of the cerebellar white matter. *Neuroimage* 2016; **124:** 724–32.

18 Archer DB, Coombes SA, McFarland NR, DeKosky ST, Vaillancourt DE. Development of a transcallosal tractography template and its application to dementia. *Neuroimage* 2019; **200:** 302–12.

19 Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011; **12:** 2825–30.

20 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44:** 837–45.

21 Fortin JP, Parker D, Tunc B, et al. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 2017; **161:** 149–70.

22 Dickson DW. Parkinson's disease and parkinsonism: neuropathology. *Cold Spring Harb Perspect Med* 2012; **2:** a009258.

23 Scherfler C, Gobel G, Muller C, et al. Diagnostic potential of automated subcortical volume segmentation in atypical parkinsonism. *Neurology* 2016; **86:** 1242–49.

24 Tang CC, Poston KL, Eckert T, et al. Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis. *Lancet Neurol* 2010; **9:** 149–58.

25 Tripathi M, Tang CC, Feigin A, et al. Automated differential diagnosis of early parkinsonism using metabolic brain networks: a validation study. *J Nucl Med* 2016; **57:** 60–66.

26 Barbagallo G, Sierra-Pena M, Nemmi F, et al. Multimodal MRI assessment of nigro-striatal pathway in multiple system atrophy and Parkinson disease. *Mov Disord* 2016; **31:** 325–34.

27 Hansson O, Janelidze S, Hall S, et al. Blood-based NfL: a biomarker for differential diagnosis of parkinsonian disorder. *Neurology* 2017; **88:** 930–37.

28 Dickson DW, Rademakers R, Hutton ML. Progressive supranuclear palsy: pathology and genetics. *Brain Pathol* 2007; **17:** 74–82.

29 Koga S, Aoki N, Uitti RJ, et al. When DLB, PD, and PSP masquerade as MSA: an autopsy study of 134 patients. *Neurology* 2015; **85:** 404–12.

30 Bastiani M, Cottaar M, Fitzgibbon SP, et al. Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *Neuroimage* 2019; **184:** 801–12.

31 Ofori E, Pasternak O, Planetta PJ, et al. Longitudinal changes in free-water within the substantia nigra of Parkinson's disease. *Brain* 2015; **138:** 2322–31.

32 Ofori E, Pasternak O, Planetta PJ, et al. Increased free water in the substantia nigra of Parkinson's disease: a single-site and multi-site study. *Neurobiol Aging* 2015; **36:** 1097–104.

33 Burciu RG, Ofori E, Archer DB, et al. Progression marker of Parkinson's disease: a 4-year multi-site imaging study. *Brain* 2017; **140:** 2183–92.