CLINICAL INVESTIGATIVE STUDY

JOURNAL OF
NEUROIMAGING
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF NEUROIMAGING

WILEY

Differentiation of multiple system atrophy subtypes by gray matter atrophy

Anna Campabadal[1,2] | Alexandra Abos[1] | Barbara Segura[1,2,3] |
Gemma Monte-Rubio[1] | Alexandra Perez-Soriano[3,4,5] | Darly Milena Giraldo[3,4,5] |
Esteban Muñoz[2,3,4,5] | Yaroslau Compta[2,3,4,5] | Carme Junque[1,2,3] |
Maria Jose Marti[2,3,4,5]

[1] Medical Psychology Unit, Department of Medicine, Institute of Neuroscience, University of Barcelona, Barcelona, Spain

[2] Institute of Biomedical Research August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[3] Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Barcelona, Spain

[4] Movement Disorders Unit, Neurology Service, Hospital Clínic de Barcelona, Barcelona, Spain

[5] Institute of Neuroscience, University of Barcelona, Barcelona, Spain

Correspondence
Carme Junque, Medical Psychology Unit, Department of Medicine, Institute of Neuroscience, University of Barcelona, Casanova 143, 08036 Barcelona, Spain.
Email: cjunque@ub.edu

Abstract

Background and Purpose: Multiple system atrophy(MSA) is a rare adult-onset synucleinopathy that can be divided in two subtypes depending on whether the prevalence of its symptoms is more parkinsonian or cerebellar (MSA-P and MSA-C, respectively). The aim of this work is to investigate the structural MRI changes able to discriminate MSA phenotypes.

Methods: The sample includes 31 MSA patients (15 MSA-C and 16 MSA-P) and 39 healthy controls. Participants underwent a comprehensive motor and neuropsychological battery. MRI data were acquired with a 3T scanner (MAGNETOM Trio, Siemens, Germany). FreeSurfer was used to obtain volumetric and cortical thickness measures. A Support Vector Machine (SVM) algorithm was used to assess the classification between patients' group using cortical and subcortical structural data.

Results: After correction for multiple comparisons, MSA-C patients had greater atrophy than MSA-P in the left cerebellum, whereas MSA-P showed reduced volume bilaterally in the pallidum and putamen. Using deep gray matter volume ratios and mean cortical thickness as features, the SVM algorithm provided a consistent classification between MSA-C and MSA-P patients (balanced accuracy 74.2%, specificity 75.0%, and sensitivity 73.3%). The cerebellum, putamen, thalamus, ventral diencephalon, pallidum, and caudate were the most contributing features to the classification decision ($z > 3.28$; $p < .05$ [false discovery rate]).

Conclusions: MSA-C and MSA-P with similar disease severity and duration have a differential distribution of gray matter atrophy. Although cerebellar atrophy is a clear differentiator between groups, thalamic and basal ganglia structures are also relevant contributors to distinguishing MSA subtypes.

KEYWORDS
cognition, cortical thickness, multiple system atrophy, machine learning, neuroimaging

INTRODUCTION

Multiple system atrophy (MSA) is a rare adult-onset synucleinopathy characterized by progressive autonomic dysfunction combined with parkinsonian and cerebellar features.[1] It can be divided into two subtypes depending on whether the initial symptoms were predominantly parkinsonian or cerebellar (MSA-P and MSA-C, respectively). MSA is neuropathologically characterized by α-synuclein glial cytoplasmatic inclusions and striatonigral or/and olivopontocerebellar neurodegeneration.[2] Although the neuropathological features are mainly located in subcortical structures, including the pons, the putamen, the brainstem, and the cerebellum, involvement of cortical motor areas in the atrophy process has also been described.[3,4]

In this context, several abnormalities on conventional MRI, such as atrophy of the putamen, middle cerebellar peduncle, pons, or cerebellum, are included as additional features in the current diagnostic criteria for MSA.[5] In the last decade, there has been an increasing interest in detecting specific MRI changes able to distinguish MSA subtypes. Most investigations have focused on structural differences with healthy controls (HCs) reporting deep gray matter volume loss in MSA-P[6-14] and in MSA-C,[8,9,14-17] as well as cortical atrophy.[6-8,12,13,16,18] One of the major drawbacks of these previous research is that few have directly compared MSA-C and MSA-P,[8,15,19-22] and only some included a control group.[8,15,20,22] The comparison of MRI parameters of atrophy between subtypes would provide scientific evidence on possible MSA phenotypes.

The recent introduction of machine learning algorithms in MRI studies has helped to test the importance of specific measures in discriminating different subtypes of neurodegenerative diseases. In parkinsonism disorders, this approach has been applied to differentiate Parkinson's disease from atypical parkinsonism.[9,20,23-25] Nonetheless, despite some of these previous neuroimaging works included patients with both MSA subtypes, only one multicenter study investigated the structures that better discriminate MSA-C from MSA-P[9] using volumetric MRI measures.

In the current case-control study, we aimed to investigate those cortical and subcortical changes able to distinguish MSA subtypes based on cortical thickness and volumetric MRI data. According to previous literature, we hypothesized that MSA subtypes will present a differential topographical distribution of gray matter atrophy with particular involvement of deep gray matter nuclei. To test this hypothesis, we studied intergroup differences in cortical and subcortical structures, and we also introduced gray matter data into a supervised machine learning algorithm to assess its ability to correctly classify each patient's group membership.

METHODS

Participants

Thirty-eight MSA patients (17 MSA-C and 21 MSA-P) were recruited from Movement Disorders Unit, Hospital Clínic de Barcelona.

MSA variants were diagnosed by an experienced movement disorder specialist, and phenotype was assigned depending on predominant motor symptom at disease onset following clinical consensus criteria.[5] Forty HCs were recruited from patients' spouses or friends who volunteered to participate in the study.

Exclusion criteria consisted of (1) pathological MRI findings other than mild white matter (WM) hyperintensities, (2) MRI movement artifacts, and (3) significant neurological, systemic, or psychiatric comorbidity in the HC group.

Six MSA patients were excluded for excessive movement. One MSA patient was excluded for MRI artifacts. One HC was excluded for abundant WM hyperintensities. The final sample therefore consisted of 39 HC and 31 MSA patients (15 MSA-C and 16 MSA-P).

Disease severity was evaluated using the Unified Multiple System Atrophy Rating Scale (UMSARS) for MSA patients.

The study was approved by the Ethics Committee of the University of Barcelona and the Hospital Clinic (IRB00003099 and HCB/2015/0798, respectively). All participants provided written informed consent to participate after full explanation of the procedures involved.

Clinical and neuropsychological assessment

Participants were evaluated with a comprehensive neuropsychological battery. Attention and working memory domains were assessed with the Trail Making Test (TMT, parts A and B) (in seconds), Digit Span Forward and Backward, the Stroop Color-Word Test, and the Symbol Digits Modalities Test (SDMT)—Oral version. Executive functions were evaluated with phonemic (words beginning with the letter "p" in 1 minute) and semantic (animals in 1 minute) fluencies. Language was evaluated by the total number of correct responses in the short version of the Boston Naming Test (BNT). For memory domain, we used the Rey's Auditory Verbal Learning Test (RAVLT). We recorded total learning recall (sum of correct responses from trial I to trial V) and delayed recall (total recall after 20 minutes). Visuospatial and visuoperceptual (VS/VP) domains were assessed with Benton's Judgement of Line Orientation (BJLO), Visual Form Discrimination (VFD), and Facial Recognition tests.[26]

Neuropsychiatric symptomatology was measured by means of the Neuropsychiatric Inventory (NPI),[27] the Beck Depression Inventory II (BDI),[28] and the Starkstein's Apathy Scale (AS).[29]

MRI acquisition and preprocessing

MRI data were acquired with a 3T scanner (MAGNETOM Trio, Siemens, Germany). The scanning protocol included high-resolution 3-dimensional T1-weighted images acquired in the sagittal plane (repetition time [TR] = 2300 ms, echo time [TE] = 2.98 ms, inversion time = 900 ms, 240 slices, field of view = 256 mm; 1 mm isotropic voxel) and an axial fluid-attenuated inversion recovery sequence (TR = 9000 ms, TE = 96 ms).

Structural MRI preprocessing was performed using the automated FreeSurfer software (version 5.1; available at: https://surfer.nmr.mgh.harvard.edu/). Independent steps were performed: removal of nonbrain tissue, automated Talairach transformation, intensity normalization,[30] tessellation of the gray matter/white matter boundary, automated topology correction,[31] and accurate surface deformation to optimally place the gray matter/white matter and gray matter/cerebrospinal fluid boundaries.[32] The output of each step (registration, skull stripping, segmentation, and cortical surface reconstruction) was visually inspected to guarantee correct and accurate preprocessing.

Automated subcortical segmentation performed with FreeSurfer was used to obtain deep gray matter nuclei volumetry. Estimated Total Intracranial Volume (eTIV) was obtained to correct volumetric data for interindividual differences in brain sizes.

Machine learning classification

Machine learning analysis was performed using NeuroMiner software, version 1.05 (http://proniapredictors.eu/neurominer/index.). To differentiate between classes, supervised classification using a linear Support Vector Machine (SVM) was performed within a repeated nested cross-validation (CV) framework. In both inner and outer CVs, a 10-fold CV cycle was applied. We carried out a repeated nested CV at the outer CV cycle by randomly permuting the participants within their groups (10 permutations) and repeating the CV cycle for each of these permutations. Deep gray matter structures, namely, the brainstem, accumbens, amygdala, caudate, cerebellum, hippocampus, pallidum, putamen, thalamus, and ventral diencephalon volumes, were divided by the corresponding eTIV. Volume ratios and mean cortical thickness, a total of 11 variables per subject, were used as features. To remove age and gender effects, partial correlations were regressed out, and a scale feature wise (from 0 to 1) was applied to the data matrix. The optimization of the C parameter associated with the SVM was carried out using a range of six parameters: 0.001, 0.01, 0.1, 1, 10, and 100.

The reliability of the predictive pattern elements was evaluated using cross-validation ratio (CVR) mapping. Also, the significance of predictive features used by the neuroimaging model was assessed by means of sign-based consistency mapping. This mapping allows knowing the contribution of the features to the classification decision. NeuroMiner has implemented the approach proposed by Gómez-Verdejo et al.,[33] which is based on wrapper-based feature selection strategies. NeuroMiner uses the normal cumulative distribution function to pick the right-tailed p-value corresponding to the respective z-score of the variable importance of each feature (details in ref. 34).

Statistical analyses

Group differences were conducted in demographic, neuropsychological, clinical, and volumetric variables using IBM SPSS Statistics 25.0.0

(2017; IBM Corp, Armonk, NY) by analysis of variance (ANOVA), covariance (ANCOVA) followed by post hoc tests, or Kruskal-Wallis H and Mann-Whitney U tests as appropriate. False discovery rate (FDR) was used for multiple comparison correction. Differences in categorical measures were analyzed by Pearson's chi-squared.

Intergroup cortical thickness comparisons were performed using a vertex-by-vertex general linear model. The model included cortical thickness as a dependent factor and group as an independent factor. All results were corrected for multiple comparisons using a precached cluster-wise Monte Carlo simulation with 10,000 iterations. Reported cortical regions reached a two-tailed corrected significance level of $p < .05$.

Regarding the machine learning classification, the total accuracy (A) and balanced accuracy (BA) are provided. BA accounts for equal weight to the accuracies obtained on each class. Additionally, sensitivity and specificity of the SVM classification as well as the area under the curve (AUC) of the receiver operating characteristic (ROC) curve are additionally reported.

The reliability of the predictive pattern elements was estimated using CVR, and sign-based consistency mappings. To determine the significance of the predictive signatures of the features, z-scores and p-values were derived using permutation analysis; 100 permutations were set. The obtained p-values were corrected using FDR, and the corrected significance threshold was defined at $\alpha = .05$.

RESULTS

Sociodemographic and clinical characteristics

Table 1 summarizes the sociodemographic and clinical characteristics of participants. We found no significant differences between groups in age and sex. Differences in years of education did not survive after multiple comparison corrections.

MSA subtypes were comparable regarding years of disease evolution, severity, and disease stage measured by the UMSARS and Hoehn and Yahr scales. However, MSA-P had higher levels of levodopa equivalent daily dose. Regarding neuropsychiatric scales, groups differed in the NPI, BDI, and AS. Post hoc analysis showed differences between HC and both MSA subtypes for BDI and AS scales.

Neuropsychological results

Table 2 describes neuropsychological results by group. Among those measures surviving FDR multiple comparison correction ($p \leq .017$), both MSA subtypes had significant differences compared with HC in Stroop Word, Stroop Color, Stroop Word and Color, TMTA, TMTB, TMT B-A phonemic and semantic fluency, RAVLT total learning and recall, BJLO, and SDMT. Only the MSA-P significantly differed from HC in MMSE, Digit Span Backward, and VFD. Comparing MSA phenotypes, MSA-P had greater impairment than MSA-C in BNT.

TABLE 1 Clinical and sociodemographic measures

	HC (n = 39)	MSA-C (n = 15)	MSA-P (n = 16)	Test stat/p-value
Age (year)	61.7 (11.5)	61.0 (7.0)	60.9 (9.8)	0.170/.919
Years of education	13.1 (4.2)	11.4 (4.1)	10.0 (3.5)	6.297/.043
Sex (male/female)	17/22	10/5	9/7	2.503/.286
H&Y (1:2:3:4:5)	–	0:4:7:3:1	0:4:4:6:2	2.121/.548
UMSARS	–	41.7 (13.8)	53.3 (19.2)	161.5/.101
LEDD	–	287.0 (395.0)	735.3 (321.2)	201.5/.001*
Years of disease evolution	–	4.1 (2.2)	5.2 (2.9)	146/.318
NPI	7.1 (10.4)	14.0 (13.2)	11.6 (9.6)	6.240/.044
BDI	7.1 (8.2)	15.4 (8.0)	19.4 (11.8)	20.246/<.001*[1,2]
AS	8.5 (5.4)	19.5 (7.1)	18.9 (7.7)	26.068/<.001*[1,2]

Note: Subjects are grouped according to Multiple System Atrophy diagnoses subtype. Data are presented as mean and standard deviation or frequencies. Asterisk (*) refers to significant results surviving false discovery rate multiple comparison correction ($p \leq .001$).
Post hoc Differences between HC and MSA-C[1]; HC and MSA-P[2]; MSA-C and MSA-P[3].
Abbreviations: AS, Starkstein's Apathy Scale; BDI, Beck Depression Inventory II; HC, healthy controls; H&Y, Hoehn and Yahr scale; LEDD, levodopa equivalent daily dose (in mg); MSA-C, multiple system atrophy cerebellar type patient group; MSA-P, multiple system atrophy parkinsonian type patient group; NPI, Neuropsychiatric Inventory; *n*, number of subjects; UMSARS, United Multiple System Atrophy Rating Scale.

TABLE 2 Neuropsychological results

	HC (n = 36)	MSA-C (n = 15)	MSA-P (n = 16)	Test stat/p-value[a]
MMSE	29.25 (0.97)	28.6 (1.9)	27.4 (2.4)	10.848/.004*[2]
Digit Span Forward	5.39 (1.1)	4.7 (0.9)	4.8 (0.8)	5.691/.058
Digit Span Backward	4.0 (0.86)	3.7 (0.8)	3.2 (0.7)	10.945/.004*[2]
Stroop W	92.5 (18.1)	63.7 (22.0)	57.9 (17.6)	23.879/<.001*[1,2]
Stroop C	62.9 (11.7)	44.5 (14.7)	42.2 (14.5)	21.614/<.001*[1,2]
Stroop WC	35.1 (9.8)	26.4 (10.9)	27.3 (9.6)	8.135/.017*[1,2]
TMT A	44.4 (14.8)	73.6 (26.8)	110.4 (59.9)	23.248/<.001*[1,2]
TMT B	114.4 (65.4)	222.1 (99.8)	317.0 (252.2)	18.506/<.001*[1,2]
TMT B-A	69.7 (54.9)	148.5 (85.2)	224.5 (222.4)	15.571/<.001*[1,2]
Phonemic fluency	14.4 (5.2)	10.6 (4.1)	10.3 (4.3)	9.920/.007*[1,2]
Semantic fluency	19.8 (5.2)	16.0 (5.1)	13.6 (4.3)	14.853/.001*[1,2]
BNT	13.6 (1.1)	13.4 (1.1)	12.4 (1.3)	8.640/.013*[2,3]
RAVLT Total	45.6 (7.1)	30.7 (9.5)	36.5 (9.2)	23.525/<.001*[1,2]
RAVLT Recall	9.2 (2.3)	6.5 (3.0)	6.6 (3.0)	11.590/.003*[1,2]
BJLO	24.9 (4.2)	21.9 (4.1)	18.3 (5.4)	16.849/<.001*[1,2]
VFD	30.0 (2.0)	27.9 (3.6)	26.6 (2.7)	15.861/<.001*[2]
FRT	23.4 (4.3)	21.5 (3.1)	21.1 (3.1)	5.200/.074
SDMT	48.4 (10.7)	31.3 (11.9)	28.5 (16.2)	22.999/<.001*[1,2]

Note: Subjects are grouped according to Multiple System Atrophy diagnoses subtype. Data are presented as neuropsychological performance raw score mean and standard deviation. Asterisk (*) refers to significant results surviving false discovery rate multiple comparison correction ($p^a \leq .017$). Post hoc differences between HC and MSA-C[1]; HC and MSA-P[2]; MSA-C and MSA-P[3].
Abbreviations: BJLO, Benton's Judgment of Line Orientation test; BNT, Boston Naming Test; FRT, Facial Recognition test short form; HC, healthy controls; MMSE, Mini-mental state examination; MSA-C, multiple system atrophy cerebellar type patient group; MSA-P, multiple system atrophy parkinsonian type patient group; n, number of subjects; RAVLT, Rey's Auditory Verbal Learning Test; RAVLT Recall, total recall after 20 min; RAVLT Total, sum of correct responses from trial I to trial V; SDMT, Symbol Digits Modalities Test—Oral version; Stroop C, Stroop Color; Stroop W, Stroop Word; Stroop WC, Stroop Word Color; TMT A, Trail Making Test part A; TMT B, Trail Making Test part B; TMT B-A, TMT B minus TMT A; VFD, Visual Form Discrimination.
[a]Kruskal-Wallis *H* test followed by Mann-Whitney *U* test.

TABLE 3 Measures of global atrophy

	HC ($n = 39$)	MSA-C ($n = 15$)	MSA-P ($n = 16$)	Test stat/p-value
Cortical GM volume, mm^3	447,609.5 (36,850.2)	423,858.3 (34,871.6)	421,251.2 (49,019.3)	7.708/0.001[*1,2]
Subcortical GM volume, mm^3	172,946.5 (16,584.2)	141,098.0 (17,630.7)	148,136.6 (21,880.3)	29.083/<0.001[*1,2]
Lateral ventricles, mm^3	10,454.6 (5443.2)	13,571.5 (10,841.3)	16,089.0 (6638.7)	4.569/0.014[*2]
Mean cortical thickness, mm	2.52 (0.1)	2.45 (0.1)	2.45 (0.1)	4.646/0.003[*1,2]

Note: Subjects are grouped according to Multiple System Atrophy diagnoses subtype. Data are presented as volumetric or cortical thickness measures mean and standard deviation. In volumetric analyses, estimated Total Intracranial Volume was introduced as a covariate. Asterisk (*) refers to significant results surviving false discovery rate multiple comparison correction ($p \leq .014$). Post hoc differences between HC and MSA-C[1]; HC and MSA-P[2]; MSA-C and MSA-P[3].
Abbreviations: GM, gray matter; HC, healthy controls; MSA-C, multiple system atrophy cerebellar type patient group; MSA-P, multiple system atrophy parkinsonian type patient group n, number of subjects.

TABLE 4 Differences between healthy controls and multiple system atrophy patients in cortical thickness

	Cluster size (mm^2)	MNI305 space			Cluster-wise p value	Cluster anatomical annotation
		X	Y	Z		
HC > MSA-C						
	LH clusters					
1	7505.8	−40.8	3.2	21.7	<.001	Precentral
2	2689.9	−44.4	−64.2	−2.0	<.001	Inferior temporal
3	1769.0	−8.5	−14.1	39.0	.022	Posterior cingulate
	RH clusters					
1	2126.4	34.6	15.0	24.4	.005	Caudal middle frontal
HC > MSA-P						
	LH clusters					
1	4099.41	−51.5	1.6	4.9	<.001	Precentral
2	2704.97	−49.1	−43.5	1.3	<.001	Banks of superior temporal sulcus
3	1778.48	−18.0	−60.3	21.6	.024	Precuneus
	RH clusters					
1	2988.74	12.4	16.0	39.7	<.001	Superior frontal
2	2444.76	51.1	−10.4	24.6	.001	Postcentral
3	2116.04	45.8	−4.5	14.6	.005	Postcentral
4	1696.26	42.6	10.8	−37.7	.021	Middle temporal

Note: Significant contrasts showing cortical thickness differences between groups. Results after family-wise error correction with Monte Carlo simulation and threshold at $p \leq 0.05$.
Abbreviations: HC, healthy controls; LH, left hemisphere; MSA-C, multiple system atrophy cerebellar type patient group; MSA-P, multiple system atrophy parkinsonian type patient group; RH, right hemisphere.

Measures of global atrophy and cortical thickness

Both MSA subtypes showed significant reduction in volumetric measures of cortical and subcortical gray matter, as well as reduced mean cortical thickness compared to HC (Table 3).

Maps of cortical thickness comparisons showed that MSA-C group had cortical atrophy compared to HC in a cluster extending from the lateral orbitofrontal, to the pars triangularis and opercularis, the precentral, and the caudal middle frontal; in a second cluster including the inferior and middle temporal; in a cluster involving the left posterior cingulate and superior frontal; and in the right caudal middle frontal (corrected $p \leq .05$) (Figure 1 and Table 4).

As for MSA-P, they had larger atrophy than HC in a cluster located in the left precentral and postcentral; in a cluster involving regions from the left superior temporal sulcus and inferior parietal gyrus; in the left precuneus; in the right middle temporal gyrus; in a cluster extending from the right superior frontal to the posterior cingulate and precuneus; and in a cluster including regions from the right precentral, postcentral, and middle temporal gyri (corrected $p \leq 0.05$) (Figure 1 and Table 4). We did not find differences in measures of global atrophy and cortical thickness between MSA-C and MSA-P.
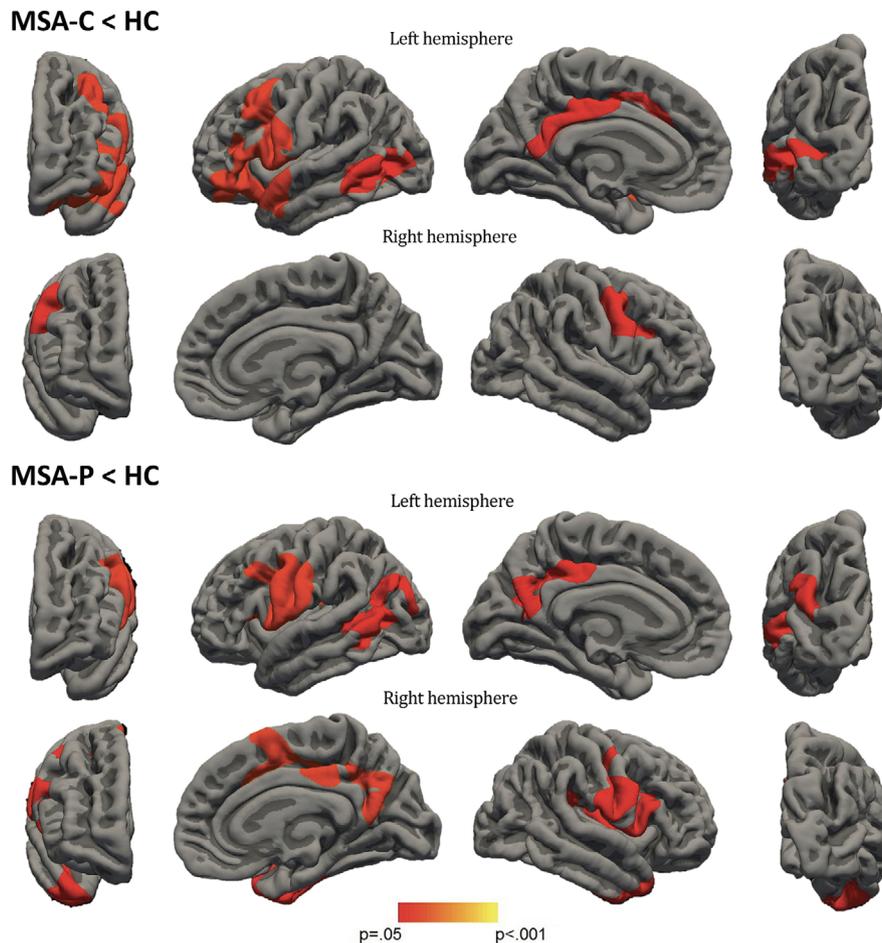
FIGURE 1 Differences between healthy controls and multiple system atrophy subtypes in cortical thickness. Significant clusters are highlighted in warm colors.
HC, healthy controls; MSA-C, multiple system atrophy cerebellar type patient group; MSA-P, multiple system atrophy parkinsonian type patient group.
Results after family-wise error correction with Monte Carlo simulation and threshold at $p \leq .05$. Graphics program: Freeview from FreeSurfer (https://surfer.nmr.mgh.harvard.edu/fswiki/FreeviewGuide) and edited with Microsoft PowerPoint®

Deep gray matter nuclei volumetry

Table 5 includes volumetric data by group. Between group analyses showed significant differences after FDR multiple comparisons correction ($p \leq .017$) in all the structures except the left thalamus. Post hoc analyses showed that MSA-P patients had greater volume reduction in all the significant structures in comparison to HC, whereas MSA-C showed greater volume reduction in brainstem, left hippocampus, pallidum, putamen, and ventral diencephalon, as well as bilaterally in the accumbens and cerebellum. MSA-C patients had decreased volume of the bilateral cerebellum compared to MSA-P (right cerebellum $p = .021$ and left cerebellum $p \leq .001$). By contrast, MSA-P showed reduced gray matter volume compared to MSA-C bilaterally in the pallidum (right pallidum $p = .004$ and left pallidum $p = .006$), putamen (right putamen $p \leq .001$ and left putamen $p = .002$), but also left amygdala ($p = .036$).

SVM classification

Performance of the classification provided a balance accuracy of 74.2% (specificity 75.0%; sensitivity 73.3%). In this case, as the samples were highly balanced, accuracy was also 74.2%. The corresponding AUC was 0.75 (95% confidence interval 0.58–0.93). Figure 2 depicts the classification metrics and ROC curve.

For completeness, the predictive pattern elements evaluated using CVR mapping provided a distinction between those features that more contributed to each class decision (Figure 3A). Also, the significance of the predictive features was assessed by means of sign-based consistency mapping (Figure 3B). Significant predictors, identified by the sign-based consistency mapping ($z > 3.28$; $p < .05$ for FDR), were the cerebellum, putamen, thalamus, ventral diencephalon, pallidum, caudate, accumbens, and brainstem. The z-scores accounting for the predictive power of the features are detailed in Table 6.
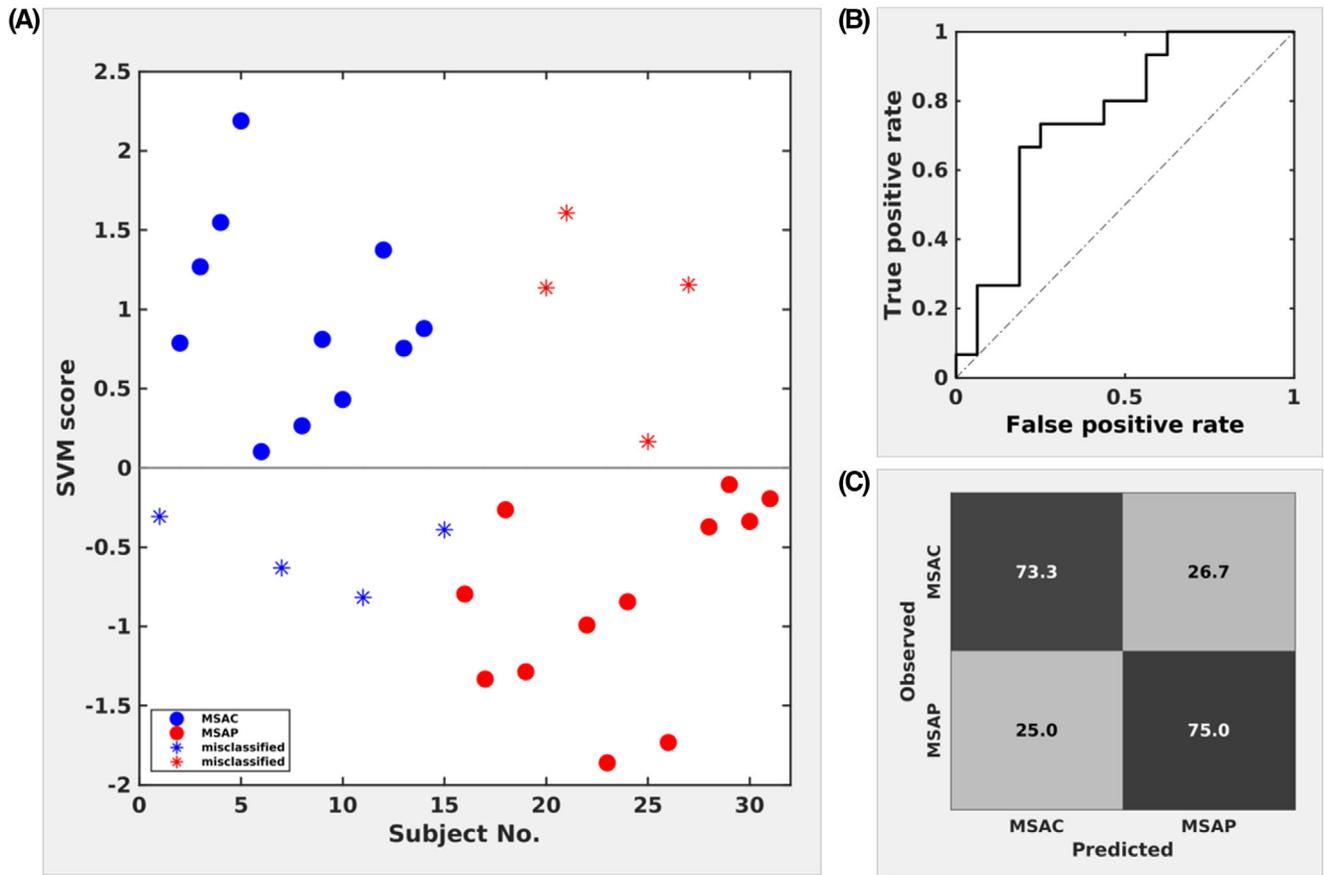
FIGURE 2 Support Vector Machine (SVM) classification performance. (A) SVM classification per MSA subtypes. (B) Receiver operating characteristic curve. (C) Sensitivity, specificity, and predictive values. No., Number; MSA-C, multiple system atrophy cerebellar type patient group; MSA-P, multiple system atrophy parkinsonian type patient group; SVM, Support Vector Machine
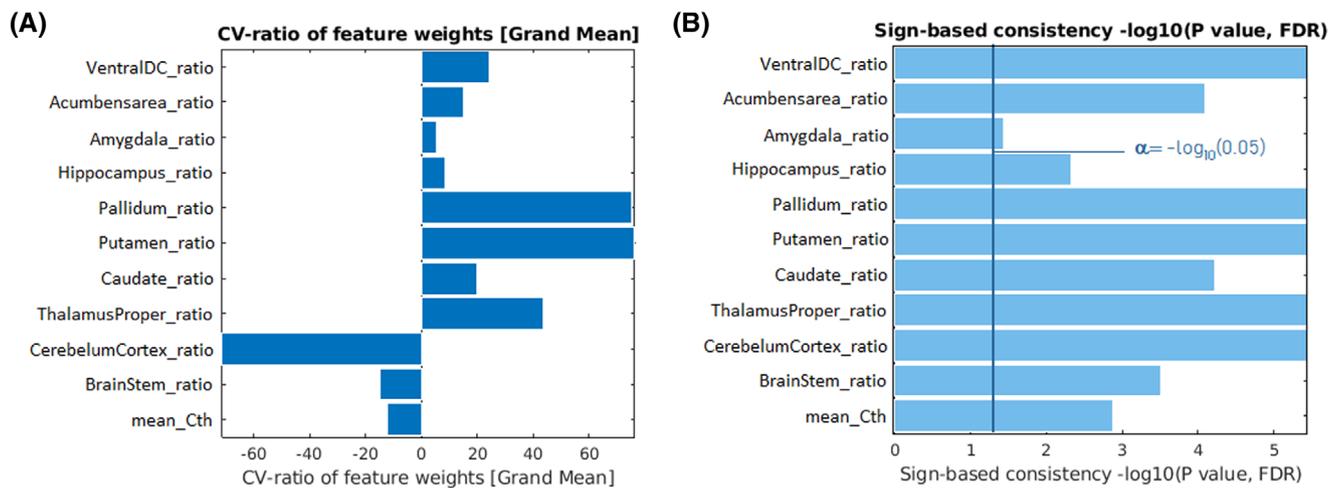
FIGURE 3 Predictive signatures underlying the multiple system atrophy cerebellar type patient group versus multiple system atrophy parkinsonian type patient group model. (A) The cross-validation ratio mapping shows the predictive pattern elements. (B) The sign-based consistency mapping shows the statistical significance of the predictive features. CV, cross-validation; DC, diencephalon; FDR, false discovery rate; Cth, cortical thickness

TABLE 5 Volumetric measures (mm^3) of deep gray matter nuclei

	HC (n = 39)	MSA-C (n = 15)	MSA-P (n = 16)	Test stat/p-value
Brainstem	21,102.9 (2654.9)	15,631.7 (2892.6)	16,696.8 (3267.3)	32.727/<.001*[1,2]
L accumbens	539.0 (103.8)	467.9 (131.8)	427.1 (125.3)	6.137/.004*[1,2]
R accumbens	591.8 (104.8)	513.7 (113.9)	458.6 (122.3)	9.384/<.001*[1,2]
L amygdala	1641.3 (234.4)	1542.0 (203.6)	1391.4 (225.5)	9.355/<.001*[2,3]
R amygdala	1678.3 (235.5)	1577.7 (200.9)	1458.9 (239.3)	5.631/.006*[2]
L caudate	3412.1 (369.5)	3186.9 (370.3)	2952.6 (701.6)	6.539/.003*[2]
R caudate	3435.4 (412.9)	3309.1 (420.9)	3007.0 (662.2)	5.756/.005*[2]
L cerebellum	48,651.6 (5659.0)	36,989.2 (6539.4)	42,333.3 (7668.7)	25.391/<.001*[1,2,3]
R cerebellum	49,889.3 (5701.8)	38,305.4 (7106.7)	43,507.1 (7233.0)	24.178/<.001*[1,2,3]
L hippocampus	4088.7 (451.2)	3791.5 (387.1)	3600.4 (483.9)	8.773/<.001*[1,2]
R hippocampus	4189.4 (593.0)	3870.0 (555.7)	3637.4 (829.0)	5.203/.008*[2]
L pallidum	1658.7 (237.2)	1500.3 (201.8)	1255.7 (307.5)	16.406/<.001*[1,2,3]
R pallidum	1453.1 (177.3)	1413.3 (197.7)	1196.1 (296.0)	9.240/<.001*[2,3]
L putamen	5111.0 (582.4)	4615.3 (831.5)	3695.4 (1198.2)	18.288/<.001*[1,2,3]
R putamen	4779.4 (560.5)	4383.4 (791.4)	3414.0 (1025.9)	21.228/<.001*[2,3]
L thalamus	6435.2 (631.0)	6274.5 (478.2)	6165.2 (889.4)	1.181/.313
R thalamus	6653.9 (636.4)	6519.6 (677.1)	6107.2 (737.1)	4.324/.017*[2]
L ventral DC	3894.0 (473.5)	3607.3 (508.8)	3438.4 (494.8)	7.458/.001*[1,2]
R ventral DC	3741.7 (434.0)	3599.3 (487.5)	3394.2 (497.5)	4.854/.011*[2]

Note: Subjects are grouped according to Multiple System Atrophy diagnoses subtype. Data are presented as volumetric measures mean and standard deviation. Asterisk (*) refers to significant results surviving false discovery rate multiple comparison correction ($p \leq .017$). Post hoc differences between HC and MSA-C[1]; HC and MSA-P[2]; MSA-C and MSA-P[3].

Abbreviations: DC, diencephalon; HC, healthy controls; L, left; MSA-C, multiple system atrophy cerebellar type patient group; MSA-P, multiple system atrophy parkinsonian type patient group; n, number of subjects; R, right.

DISCUSSION

Our results showed that MSA-C and MSA-P patients with similar disease severity and disease duration have structural differences in the cerebellum and deep gray matter structures, with no differences between them in cortical thickness. The SVM classifier demonstrated that subcortical data contribute to the differentiation between MSA phenotypes, with the putamen, pallidum, cerebellum, thalamus, ventral diencephalon, and caudate having the greatest contribution.

As stated in the Introduction, most previous conventional MRI investigations have mainly been focused on studying subtypes separately, and only a few have directly compared MSA-C and MSA-P[8,15,19–22] and include a control group for comparison.[8,15,20,22] In our work, when directly comparing subtypes, MSA-C patients had larger atrophy than MSA-P in the left cerebellum, whereas MSA-P showed reduced gray matter volume mainly in bilaterally in the pallidum and putamen, but also in left amygdala. These results are in line with previous studies reporting the expected atrophy in the cerebellum in MSA-C compared to MSA-P,[8,14,20,24] and volume reduction in the putamen in MSA-P.[20] Our findings, also agree with the presence of classical structural MRI clinical hallmarks involving the cerebellum and striatum that showed high specificity but low or moderate sensitivity to differentiate MSA from other neurodegenerative disorders such as Parkinson's

disease or Progressive Supranuclear Palsy.[35] Indeed, although not limited to, postmortem examination consistently found severely affected striatonigral degeneration and olivopontocerebellar atrophy in of MSA patients, reflecting the presence of parkinsonian features and ataxia.[1] As for cortical gray matter, it has been mainly studied using voxel-based approaches.[8,15] Only Chang et al. found gray matter cortical atrophy in MSA-C and MSA-P compared to controls, in the insula and frontal lobe and in the insula, olfactory lobes, and temporal cortex, respectively.[8] Using cortical thickness technique, we found that both MSA subtypes had cortical thinning compared to age-matched HCs in the precentral, middle temporal gyri, and posterior cingulate. It is worth noting that the pattern of cortical thinning observed in the contrast with controls involved different regions for each subtype. The MSA-C patients had a reduction in orbitofrontal regions, whereas patients with MSA-P showed greater thinning in the postcentral and parietal regions. However, differences between subtypes were probably subtle because the direct comparison did not reach significance. The lack of difference in cortical thickness between MSA phenotypes has been also found in a previous work that used the same approach.[18] Similarly, to our findings they reported significant cortical thinning in the ventromedial prefrontal and bilateral ventrolateral prefrontal cortices, in MSA-C compared to controls, but in their work MSA-P did not differ from controls. We can speculate that increased cortical degeneration in

TABLE 6 Scores of the predictive signatures

Feature	CV-ratio	Sign-based consistency Z-score
Cerebellum	−71.38	4.71
Putamen	76.43	4.71
Thalamus	43.81	4.70
Ventral DC	24.10	4.66
Pallidum	75.21	4.65
Caudate	19.69	3.99
Accumbens	15.14	3.88
Brainstem	−14.85	3.51
Mean cortical thickness	−12.18	3.07
Hippocampus	8.53	2.63
Amygdala	5.31	1.79

Note: Grand mean scaling cross-validation ratio mapping and z-score (false discovery rate [FDR] correction) from sign-based consistency method ($z > 3.28$; $p < .05$ for FDR).
Abbreviations: Cv, cross validation; DC, diencephalon.

prefrontal cortex seen in both studies could be reflecting the contribution of transneuronal degeneration due to the primary cerebellar degeneration.

Based on the differential vulnerability of the deep gray matter nuclei, and the fact that cortical thinning is also present in both MSA phenotypes, we evaluated their discriminating power. We introduced all subcortical measures and the cortical mean thickness into a supervised machine learning algorithm to assess their ability to correctly determine each patient's group membership. Methodological explorations were far from the scope of this work, thus we only applied one single machine learning approach. SVM was chosen because it is versatile for a wide range of applications, also when few labeled observations are available, and provides consistent outcomes. Also, to take the most of the data, a 10-fold CVR method was used to evaluate the model performance. CVR rotates the test sample across the whole dataset and for every test sample, the remaining dataset becomes the training sample. For each split, the test error is computed after fitting the model over the corresponding training sample. The test errors from each split are averaged to obtain the average test error. In this sense, it provides higher performance than if the hold-out method were used, splitting the data into train and test for once. SVM classification provided a consistent classification between MSA-C and MSA-P patients (balanced accuracy 74.2%, specificity 75.0%, and sensitivity 73.3%) using deep gray matter volume ratios and mean cortical thickness as features. The cerebellum, putamen, thalamus, ventral diencephalon, pallidum, caudate, accumbens, and brainstem were the most contributing features to the classification, whereas cortical thickness slightly contributes to the model. Remarkably, the volume ratio of the cerebellum on the one hand and both pallidum and putamen on the other were the features that most contributed to each class decision. This finding gives support to the expected role

of the cerebellum when it comes to discriminating between MSA subtypes, because cerebellar syndrome is a predominant clinical feature in MSA-C, but it also gives new evidence of thalamus and basal ganglia contribution. Furthermore, it is interesting to note that only patients with MSA-P showed volume reduction in limbic structures in comparison to controls, and that such structures also contribute to the classification decision. As far as we know, only one previous work has used machine learning algorithms to differentiate MSA subtypes using MRI T1 data. In their multicentric work, Huppertz et al. obtained similar accuracy (75%) but lower sensitivity (57%) using both gray matter and white matter volumes as features.[9] Their results also highlight the contribution of the striatum and the cerebellum in distinguishing MSA subtypes. On the other hand, accuracy values in our work are lower than those reported in previous investigations focused on distinguishing MSA patients from PD,[9,20,23–25] which may be explained by the fact that the two MSA subtypes share neuropathological substrates.

In our study, both MSA subtypes had neuropsychological impairment in mental processing speed, attention, executive functions, and learning, but in addition MSA-P had impairment in VS/VP and global cognition in comparison to controls. When directly comparing MSA phenotypes, MSA-P had greater impairment than MSA-C in naming. Thus MSA-P had more severe cognitive impairment. Similarly, Kawai et al. reported that in comparison to controls, MSA-P had impaired visuospatial functions, verbal fluency, and executive functions, whereas MSA-C only had visuospatial deficits.[36] These results agree with our MRI findings showing increased cortical atrophy compared with controls in MSA-P.

The main limitation of this study is the small sample of MSA patients. This disease is a rare neurodegenerative disease, and it is important to highlight the difficulty of recruiting larger samples. However, in our study all patients underwent the same MRI and neuropsychological protocol. This avoids the heterogeneity of multicenter MRI studies.

In this work, we found volumetric differences in the cerebellum but also in the putamen and pallidum between MSA subtypes with similar disease severity and duration. Gray matter atrophy discriminates with high accuracy MSA-C and MSA-P patients with special involvement of deep gray matter nuclei and without significant contribution of cortical thinning. In conclusion, although cerebellar atrophy clearly discriminates between subtypes, other subcortical structures have a relevant contribution.

ORCID

Anna Campabadal https://orcid.org/0000-0002-2833-4884

Alexandra Abos https://orcid.org/0000-0002-8179-4177

Barbara Segura https://orcid.org/0000-0002-9673-5479

Gemma Monte-Rubio https://orcid.org/0000-0002-3532-2224

Alexandra Perez-Soriano https://orcid.org/0000-0003-1614-7528

Yaroslau Compta https://orcid.org/0000-0001-6443-0104

Carme Junque https://orcid.org/0000-0002-6381-3063

Maria Jose Marti https://orcid.org/0000-0002-3874-967X

REFERENCES

1. Fanciulli A, Wenning GK. Multiple-system atrophy. N Engl J Med 2015;372:249-63.
2. Lin DJ, Hermann KL, Schmahmann JD. Multiple system atrophy of the cerebellar type: clinical state of the art. Mov Disord 2014;29:294-304.
3. Lantos PL. The definition of multiple system atrophy. J Neuropathol Exp Neurol 1998;57:1099-111.
4. Wenning GK, Quinn NP. Parkinsonism. Multiple system atrophy. Baillieres Clin Neurol 1997;6:187-204.
5. Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. Neurology 2008;71:670-6.
6. Brenneis C, Seppi K, Schocke MF, et al. Voxel-based morphometry detects cortical atrophy in the Parkinson variant of multiple system atrophy. Mov Disord 2003;18:1132-8.
7. Caso F, Canu E, Lukic MJ, et al. Cognitive impairment and structural brain damage in multiple system atrophy-parkinsonian variant. J Neurol 2020;267:87-94.
8. Chang CC, Chang YY, Chang WN, et al. Cognitive deficits in multiple system atrophy correlate with frontal atrophy and disease duration. Eur J Neurol 2009;16:1144-50.
9. 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;31:1506-17.
10. Paviour DC, Price SL, Jahanshahi M, et al. Regional brain volumes distinguish PSP, MSA-P, and PD: MRI-based clinico-radiological correlations. Mov Disord 2006;21:989-96.
11. Planetta PJ, Kurani AS, Shukla P, et al. Distinct functional and macrostructural brain changes in Parkinson's disease and multiple system atrophy. Hum Brain Mapp 2015;36:1165-79
12. Kim JS, Yang J, Lee D-K, et al. Cognitive impairment and its structural correlates in the Parkinsonian subtype of multiple system atrophy. Neurodegener Dis 2015;15:294-300.
13. Tir M, Delmaire C, le Thuc V, et al. Motor-related circuit dysfunction in MSA-P: usefulness of combined whole-brain imaging analysis. Mov Disord 2009;24:863-70.
14. Yang H, Wang N, Luo X, et al. Cerebellar atrophy and its contribution to motor and cognitive performance in multiple system atrophy. NeuroImage Clin 2019;23:101891.
15. Dash SK, Stezin A, Takalkar T, et al. Abnormalities of white and grey matter in early multiple system atrophy: comparison of parkinsonian and cerebellar variants. Eur Radiol 2019;29:716-24
16. Lee MJ, Shin J-H, Seoung J-K, et al. Cognitive impairments associated with morphological changes in cortical and subcortical structures in multiple system atrophy of the cerebellar type. Eur J Neurol 2016;23:92-100.
17. Zhang H, Ji S, Ren S, et al. Cerebellar atrophy in multiple system atrophy (cerebellar type) and its implication for network connectivity. Cerebellum 2020;19:636-44.
18. Kim JS, Youn J, Yang J-J, et al. Topographic distribution of cortical thinning in subtypes of multiple system atrophy. Parkinsonism Relat Disord 2013;19:970-4.
19. Barbagallo G, Sierra-Peña 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.
20. Krismer F, Seppi K, Göbel G, et al. Morphometric MRI profiles of multiple system atrophy variants and implications for differential diagnosis. Mov Disord 2019;34:1041-8.
21. Lee EA, Cho HI, Kim SS, et al. Comparison of magnetic resonance imaging in subtypes of multiple system atrophy. Parkinsonism Relat Disord 2004;10:363-8.
22. Schulz JB, Skalej M, Wedekind D, et al. Magnetic resonance imaging–based volumetry differentiates idiopathic Parkinson's syndrome from multiple system atrophy and progressive supranuclear palsy. Ann Neurol 1999;45:65-74.
23. Abos A, Baggio HC, Segura B, et al. Differentiation of multiple system atrophy from Parkinson's disease by structural connectivity derived from probabilistic tractography. Sci Rep 2019;9:16488.
24. Péran P, Barbagallo G, Nemmi F, et al. MRI supervised and unsupervised classification of Parkinson's disease and multiple system atrophy. Mov Disord 201833:600-8
25. Scherfler C, Göbel G, Müller C, et al. Diagnostic potential of automated subcortical volume segmentation in atypical parkinsonism. Neurology 2016;86:1242-9.
26. Lezak MD. *Neuropsychological Assessment*. Oxford: Oxford University Press; 2012.
27. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The neuropsychiatric inventory: comprehensive assessment of psychopathology in dementia. Neurology 1994;44:2308-14.
28. Beck A, Steer RBG. *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation; 1996.
29. Starkstein SE, Mayberg HS, Preziosi TJ, et al. Reliability, validity, and clinical correlates of apathy in Parkinson's disease. J Neuropsychiatry Clin Neurosci 1992;4:134-9.
30. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17:87-97.
31. Ségonne F, Pacheco J, Fischl B. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans Med Imaging 2007;26:518-29
32. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA 2000;97:11050-5.
33. Gómez-Verdejo V, Parrado-Hernández E, Tohka J. Sign-consistency based variable importance for machine learning in brain imaging. Neuroinformatics 2019;17:593-609.
34. Koutsouleris N, Dwyer DB, Degenhardt F, et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. JAMA Psychiatry 2021;78:195.
35. Chelban V, Bocchetta M, Hassanein S, et al. An update on advances in magnetic resonance imaging of multiple system atrophy. J Neurol 2019;266:1036-45.
36. Kawai Y, Suenaga M, Takeda A, et al. Cognitive impairments in multiple system atrophy: MSA-C vs MSA-P. Neurology 2008;70:1390-6.