



## OPEN Precise disease heterogeneity and progression quantification in MSA and Parkinson's disease using machine learning

Robel K. Gebre<sup>1</sup>✉, Sheelakumari Raghavan<sup>1</sup>, Mari E. Johnson De Tora<sup>1</sup>, Angela J. Fought<sup>2</sup>, Robert I. Reid<sup>3</sup>, Phillip A. Low<sup>4</sup>, Wolfgang Singer<sup>4</sup>✉ & Prashanthi Vemuri<sup>1</sup>✉

Disease progression in multiple system atrophy (MSA) and Parkinson's disease (PD) shows marked patient-to-patient heterogeneity. We hypothesize that machine learning methods applied to multimodal MRI data would aid in optimally identifying critical brain regions impacted in each patient, improve disease differentiation and longitudinal tracking. Using structural and diffusion MRI of MSA (cerebellar and parkinsonian subtypes), PD, and normal participants, we trained binary classifiers and utilized Shapley Additive exPlanations (SHAP) to quantify feature contributions to derive heterogeneity scores (HET). HET outperformed commonly available imaging tools when differentiating between MSA and PD, strongly correlated with clinical markers, and sensitively tracked longitudinal disease progression. HET correctly identified olivopontocerebellar atrophy and striatonigral degeneration as important for disease identification, shed light on the spatio-temporal disease progression, and identified widespread white matter involvement in MSA. Our machine learning approach quantifies MSA and PD heterogeneity and provides a patient-specific measure for precise disease quantification and longitudinal tracking.

Multiple system atrophy (MSA) and Parkinson's disease (PD) are distinct  $\alpha$ -synucleinopathies, each characterized by unique patterns of neurodegeneration and clinical trajectories<sup>1</sup>. Idiopathic PD is characterized by the pathological loss of dopaminergic neurons within the substantia nigra, in the presence of intraneuronal Lewy body inclusions composed of  $\alpha$ -synuclein<sup>2</sup>. MSA involves degeneration of striatonigral (SN) and olivopontocerebellar structures accompanied by widespread oligodendroglial cytoplasmic inclusions of  $\alpha$ -synuclein<sup>3</sup>. Although clinical and pathologic manifestations often overlap, MSA can be differentiated into two main phenotypes: the cerebellar subtype (MSA-C) and the parkinsonian subtype (MSA-P)<sup>4</sup>. Along with cerebellar and brainstem atrophy, MSA-C is characterized by prominent olivopontocerebellar atrophy (OPCA) which manifests as gait ataxia, dysarthria, and the classic "hot cross bun" sign on T2-weighted magnetic resonance imaging (MRI) images<sup>5</sup>. On the other hand, MSA-P is dominated by SN degeneration, resulting in parkinsonian features, putaminal atrophy, and a lateral putaminal hyperintense rim<sup>5</sup>.

Imaging studies using diffusion MRI have demonstrated higher mean diffusivity (MD) in the middle cerebellar peduncles and cerebellum in MSA-C<sup>6</sup>, consistent with cerebellar region degeneration<sup>5</sup>, and elevated MD in the putamina of MSA-P patients even in early disease stages<sup>7</sup>. While diffusion changes in PD are generally less pronounced, reductions in fractional anisotropy (FA) in the substantia nigra and other regions have been reported when compared to healthy controls<sup>8</sup>. Distinguishing MSA-P from PD can also be challenging<sup>9</sup>, especially in the early stages of the disease resulting in high rates of MSA misdiagnosis<sup>10</sup>. Hence, there is a need for advanced analysis techniques such as machine learning (ML) methods that make use of multimodal imaging to better discriminate MSA and PD by considering heterogeneity of the diseases and identify their unique neurodegenerative patterns. In recent years, deep learning (DL) has been applied on MRI images to distinguish between PD and MSA<sup>11</sup>. While such efforts have proven useful to identify predictive features, there has been limited work on imaging markers that quantify longitudinal changes that drive MSA and PD<sup>12</sup>, and even less work to construct a summary measure that correlates well with clinical outcomes and captures disease heterogeneity.

<sup>1</sup>Department of Radiology, Mayo Clinic, Rochester, MN, USA. <sup>2</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. <sup>3</sup>Department of Information Technology, Mayo Clinic, Rochester, MN, USA. <sup>4</sup>Department of Neurology, Mayo Clinic, Rochester, MN, USA. ✉email: gebre.robelkebede@mayo.edu; Singer.Wolfgang@mayo.edu; Vemuri.Prashanthi@mayo.edu

One notable effort is the manually constructed MSA-atrophy index (AI) which is derived by averaging the z-scores of lentiform nucleus (putamen and globus pallidus) and the olivopontocerebellar (cerebellum and brainstem) regions<sup>13</sup>. However, given disease progression differences across individuals, we hypothesize that an ML-based summary metric would optimally identify critical regions specific to individual patients to help advance our understanding of the neurodegenerative mechanisms of MSA-C, MSA-P, and PD. We set out with two goals: (1) determining whether a ML model can improve the differentiation of the diseases and provide a heuristic accounting of disease heterogeneity, and (2) formulating a summary score derived from the ML models that can correlate with clinical measures better than cerebellar changes and existing markers at baseline and follow-ups.

## Results

### Study cohort

This longitudinal study consisted of 17 controls, 15 MSA-C, 12 MSA-P, and 15 PD participants at baseline with a one-year period between follow-up visits totaling 174 observations. Demographics and descriptive statistics are given in Table 1. Clinical assessment of MSA severity was done using the UMSARS (Unified MSA Rating Scale) total, TST (Thermoregulatory Sweat Test), CASS (Composite Autonomic Severity Score) total, and COMPASS (Composite Autonomic Symptoms Scale)-select. Both MSA subtypes showed significant mean differences from controls and from PD across all clinical measures ( $p < 0.05$ ). The PD group showed significant differences only for UMSARS total and COMPASS-select ( $p < 0.05$ ).

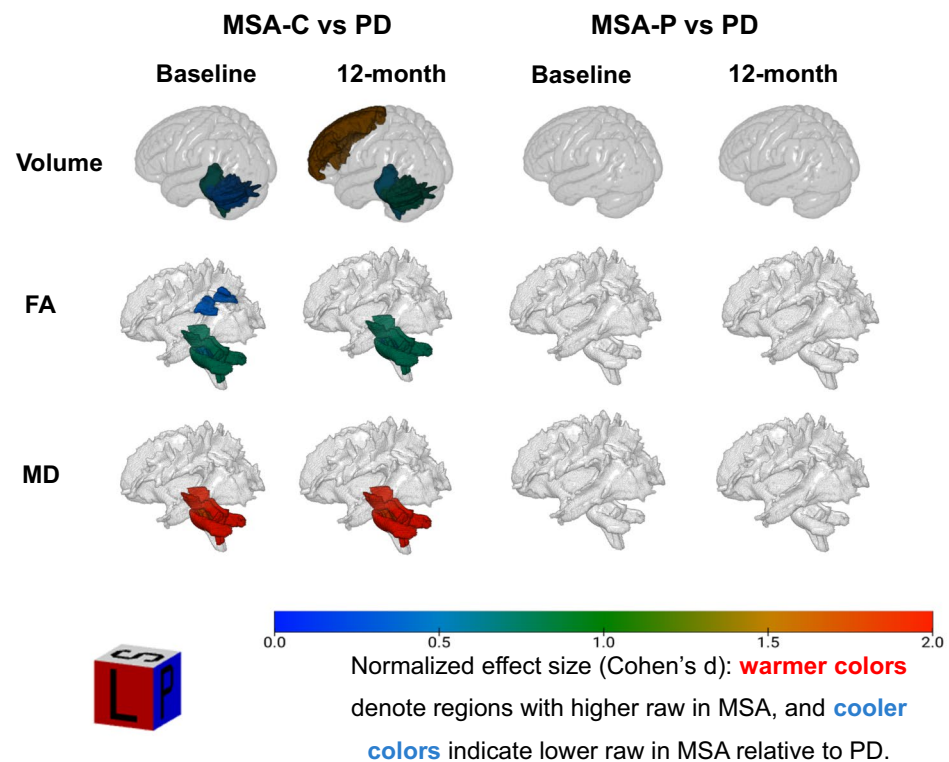
To assess whether the raw structural (T1) and diffusion MRI (dMRI) imaging features differed between the subtypes and PD, regional comparisons were performed at baseline and at the 12-month follow-up (Fig. 1). No significant volumetric or microstructural differences were found between MSA-P and PD at either timepoint ( $p > 0.05$ ). In contrast, MSA-C showed marked OPCA and significantly abnormal fractional anisotropy (FA) and mean diffusivity (MD) within the same infratentorial regions at both timepoints (Fig. 1). At follow-up, PD participants exhibited significant frontal pole atrophy, while MSA-C participants showed abnormal FA in the posterior corona radiata (PCR) at baseline.

### Performance of ML models and ML-derived heterogeneity (HET) scores

Three classifiers were trained for a binary task of discriminating MSA from PD using regional measurements of volume ( $n = 49$ ), FA ( $n = 43$ ) and MD ( $n = 49$ ) as inputs (Supplementary Fig. 1). Five ML models (XGBoost, Random Forest, LightGBM, CatBoost, and AutoGluon) were evaluated to reduce model selection bias. To prevent overfitting and data leakage, we used k-fold cross-validation with ten random seeds, with folds split by Subject ID. The best model was selected based on test-fold performance using the F1 score as the evaluation metric. The hyperparameters used for training are shown in Supplementary Table 1, and the performance plots comparing all the models can be found in Supplementary Figs. 2 & 3. After this search, the final models for volume, FA, and MD were CatBoost, Random Forest, and AutoGluon's Weighted Ensemble respectively, yielding F1-scores and balanced accuracy of 1.00 (SD 0.01), 1.00 (0.01), and 0.98 (0.06) respectively (Table 2). SHAP was then used on these models to compute the feature contributions. Feature importance plots can be found in Supplementary Fig. 4.

	Controls	MSA-C	MSA-P	PD
Count (n)	17	15	12	15
Male (n)	9	10	8	14
Baseline age (y)	63.00 (6.63)	59.71 (6.87) <sup>b</sup>	63.07 (6.33)	66.15 (6.17)
Follow-up (y)	1.00 (0.07)	1.02 (0.12)	1.01 (0.12)	1.07 (0.27)
Visits (n)				
1	17	15	12	15
2	14	13	10	11
3	14	8	6	8
4	13	6	4	8
Disease duration (y)	-	3.82 (4.28)	4.28 (2.19)	8.38 (4.84)
UMSARS total	1.00 (1.06)	27.67 (8.09) <sup>a, b</sup>	29.04 (5.61) <sup>a, b</sup>	21.37 (8.54) <sup>a</sup>
TST %	8.8 (15.27)	51.27 (44.57) <sup>a, b</sup>	59.08 (28.30) <sup>a, b</sup>	7.81 (20.17)
CASS total	0.50 (0.94)	3.67 (1.29) <sup>a, b</sup>	3.92 (2.35) <sup>a, b</sup>	1.33 (1.45)
COMPASS-select	2.57 (2.27)	21.49 (12.02) <sup>a, b</sup>	35.90 (17.65) <sup>a, b</sup>	10.77 (9.32) <sup>a</sup>

**Table 1.** Baseline descriptive statistics of participants included in the study. Count, n, and the mean (standard deviation) are shown. Pair-wise mean value comparisons were conducted between controls and MSA-C, MSA-P and PD, as well as between MSA-C and MSA-P to PD. Either a two-sided independent samples t-test or Mann-Whitney U-test was conducted after checking normality. y = years. <sup>a</sup>Significant when compared to controls,  $p < 0.05$ . <sup>b</sup>Significant when compared to PD,  $p < 0.05$ . UMSARS = Unified MSA Rating Scale. TST % = Thermoregulatory Sweat Test. CASS = Composite Autonomic Severity Score. COMPASS = Composite Autonomic Symptoms Scale



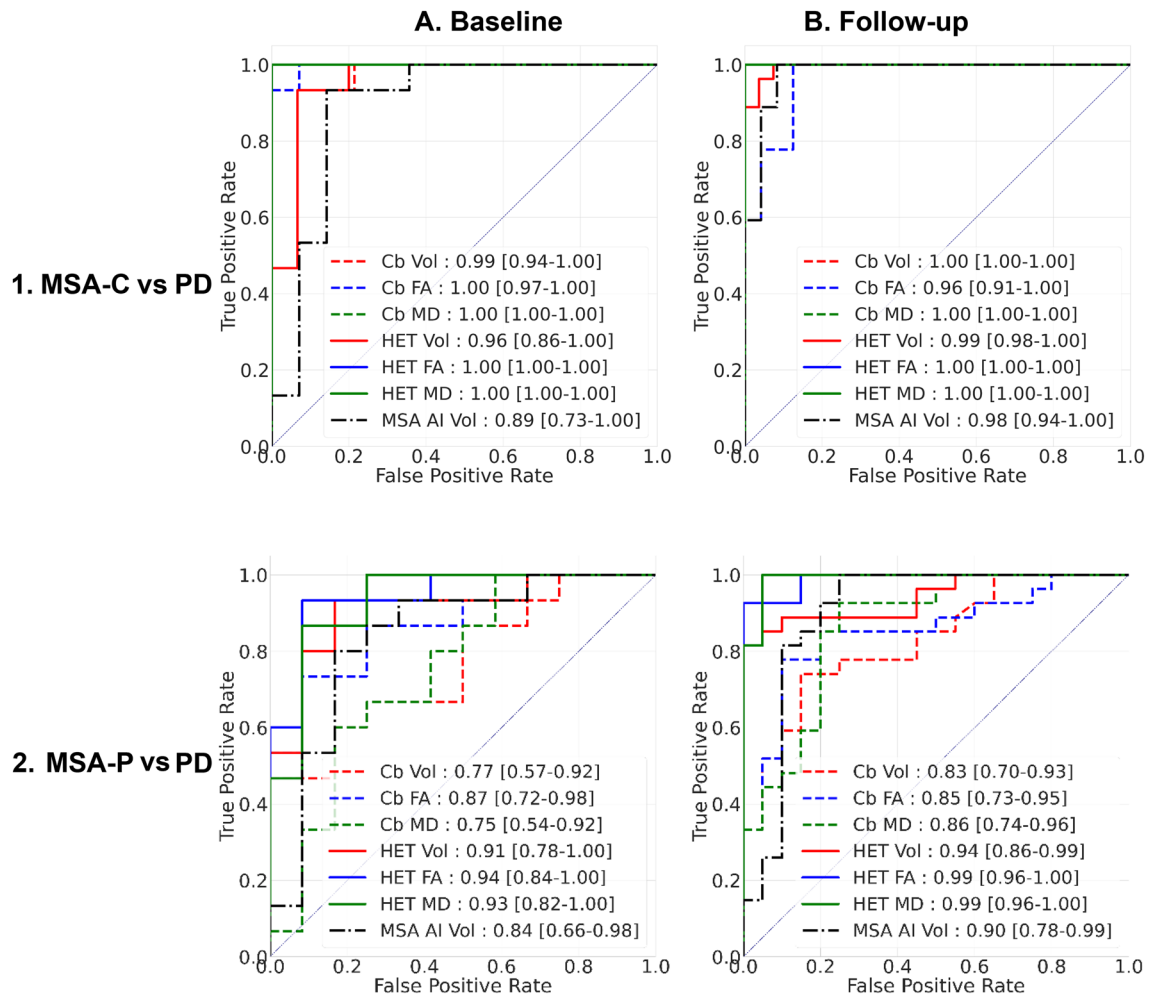
**Fig. 1.** Regional effect size maps for raw volume, fractional anisotropy (FA), and mean diffusivity (MD) measures comparing MSA subtypes and PD at baseline and 12-month follow-up. The colors indicate Cohen's  $d$  for significant regions after multiple comparison correction. Warm colors indicate higher raw regional values in MSA than PD, and cool colors indicate lower raw regional values in MSA than PD. Uncolored regions are not significant. Renderings without coloring show non-significant findings. The color bar reflects Cohen's  $d$  and not raw values.

	Best model	F1-score		Balanced accuracy	
		Train	Test	Train	Test
Volume	CatBoost	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)
FA	Random Forest	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)
MD	AutoGluon	1.00 (0.00)	0.98 (0.06)	1.00 (0.00)	0.98 (0.06)

**Table 2.** Performance summary of the best classifier for input features of volume, fractional anisotropy (FA), and mean diffusivity (MD). The train and test F1-score and balanced accuracy, shown as mean (standard deviation).

Next, the HET scores were computed using SHAP feature contributions. Refer to the Methods section for more details. Briefly, SHAP values for each input type (volume, FA, and MD) were used as weights to obtain the weighted regional measures of heterogeneity (Eq. 2). Subject-level HET scores were then calculated by averaging the weighted regional values across all regions (Eq. 3).

To evaluate performance of the subject-level HET scores to classify MSA subtypes from PD, their Area Under the Curve (AUC) of their Receiver Operator Characteristic (ROC) curves was compared to the cerebellum WM and MSA-AI. The volume HET was able to classify MSA-C from PD at AUCs of 0.96 [95% CI: 0.86–1.00] at baseline and 0.99 [0.98–1.00] at follow-up comparable to the cerebellum WM (0.99 [0.94–1.00] and 1.00 [1.00–1.00]) (Fig. 2). Comparably the MSA-AI performed with AUCs of 0.89 [0.73–1.00] at baseline and 0.98 [0.94–1.00] at follow-up. Similarly, when classifying MSA-P from PD, volume HET consistently outperformed both cerebellum WM and MSA-AI at AUCs of 0.91 [0.78–1.00] at baseline and 0.94 [0.86–0.99] at follow-up. The FA and MD HETs were equally high performing compared to the cerebellum WM in discriminating against the MSA subtypes from PD (AUCs > 0.93) (Fig. 2).



**Fig. 2.** Classification performance shown as receiver operator characteristic plot (ROC) with the area under the curve (AUC) of the heterogeneity (HET) score derived from machine learning models trained on volume, fractional anisotropy (FA), and mean diffusivity (MD). Each HET score is compared to the cerebellum (Cb) white matter (WM) volume (vol), fractional anisotropy (FA), and mean diffusivity (MD), and the MSA-atrophy index (AI) at baseline and follow-up visits.

## HET tracks disease progression

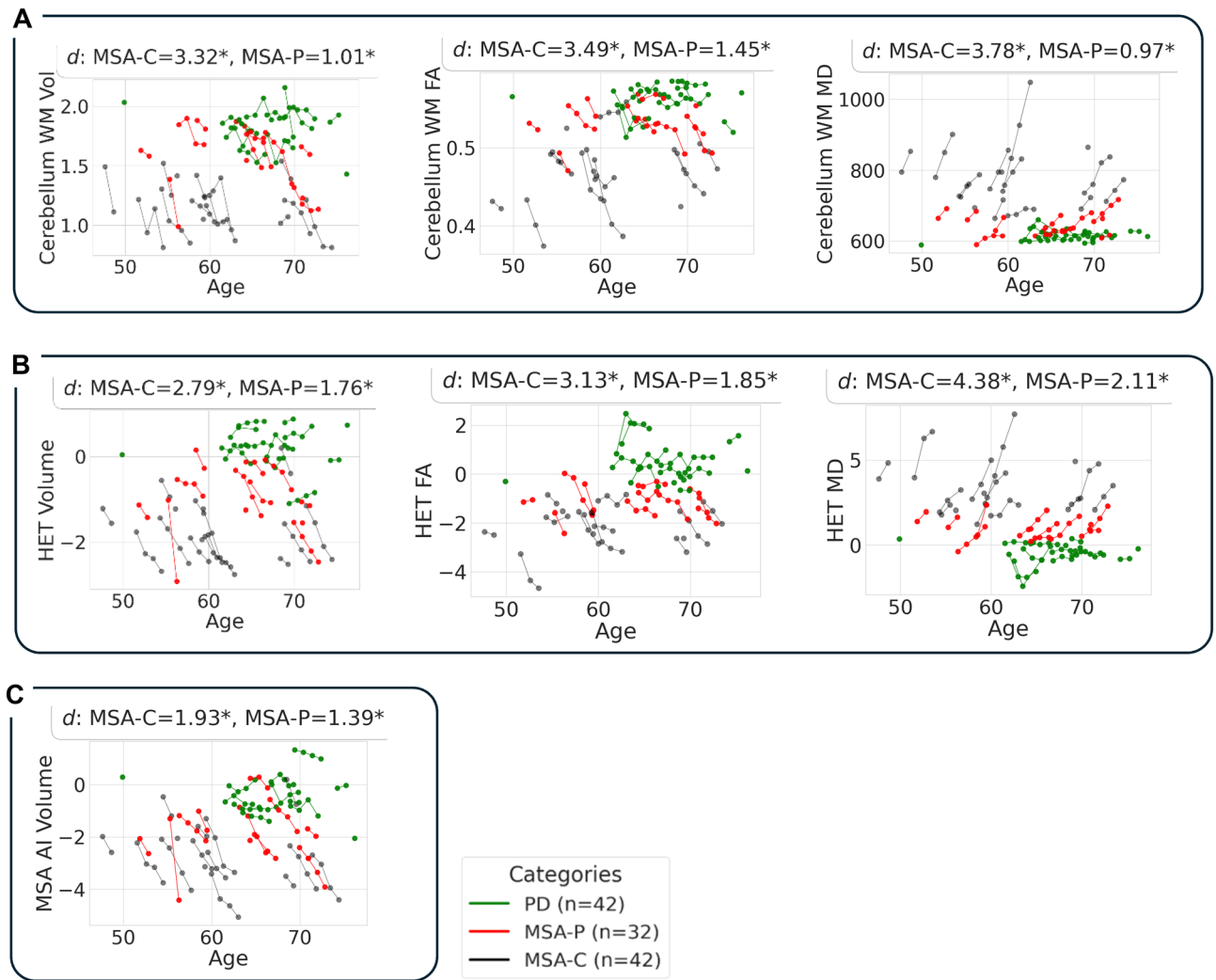
### Longitudinal validation

To evaluate the longitudinal ability of HET to track separation between MSA subtypes and PD, we computed effect size between MSA-C and PD, and MSA-P and PD (Fig. 3). Separation between MSA-C and PD at baseline was comparable for cerebellar WM volume and volume HET (Cohen's  $d = 3.32$  vs.  $2.79$ ), whereas HET improved the separation between MSA-P and PD ( $1.01$  vs.  $1.76$ ). Comparably the MSA-AI showed slightly lower effect size between MSA-C and PD (volume:  $d = 1.93$ ) but performed better than the cerebellum WM when separating MSA-P from PD ( $d = 1.39$ ) (Fig. 3).

The FA and MD HETs showed a similar pattern of group separation at baseline compared to cerebellar WM. For FA, the MSA-C to PD effect size slightly decreased with HET ( $3.49$  vs.  $3.13$ ), whereas the MSA-P to PD effect size slightly improved ( $1.45$  vs.  $1.85$ ). Compared to the cerebellum WM, the MD HET improved the effect sizes across groups ( $3.78$  vs.  $4.38$ , and  $0.97$  vs.  $2.11$ ) (Fig. 3).

### Clinical validation

Change over the first 12-month follow-up period was used to assess sensitivity of HET to track clinically relevant disease progression. Changes in volume, FA and MD HET were significantly correlated to the change in UMSARS total ( $\rho = -0.60$ ,  $p < 0.05$ ,  $\rho = -0.51$ ,  $p < 0.05$ , and  $\rho = 0.37$ ,  $p < 0.05$ ) (Fig. 4). Only cerebellum WM MD changes showed significant correlation to UMSARS total ( $\rho = 0.40$ ,  $p < 0.05$ ) while the cerebellum WM volume ( $\rho = -0.27$ ) and FA ( $\rho = 0.01$ ) changes did not show a significant correlation to UMSARS total change over the 12 month period. Changes in MSA-AI over the 12-month period were significantly correlated with UMSARS total over the same period ( $\rho = -0.54$ ,  $p < 0.05$ ) (Fig. 4).



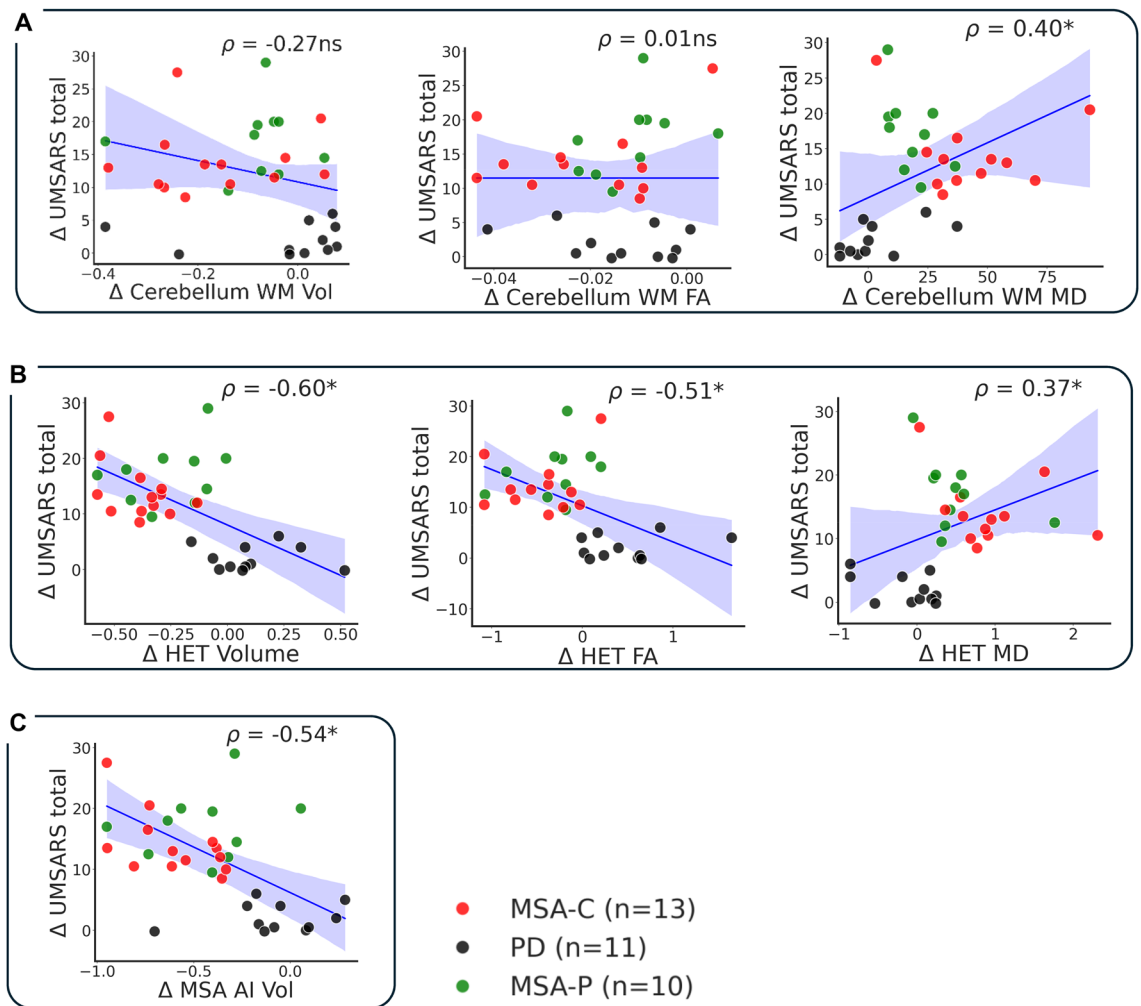
**Fig. 3.** Baseline effect sizes (Cohen's  $d$ ) comparing MSA-C vs. PD and MSA-P vs. PD, listed in that order and separated by commas. Spaghetti plots illustrate longitudinal progression for all participants. Panel **A** shows cerebellar white matter (WM), fractional anisotropy (FA), and mean diffusivity (MD); Panel **B** shows the heterogeneity (HET) scores; and Panel **C** shows the MSA atrophy index (AI). Lines represent individual trajectories for PD (green), MSA-P (red), and MSA-C (black) participants. In the legend the total number of observations ( $n$ ) is reported. The reported  $p$ -values correspond to hypothesis tests of group mean differences, \*  $p < 0.05$ , ns  $p \geq 0.05$ .

### HET captures imaging heterogeneity unique to MSA relative to PD

To quantify whether the regional HET scores can capture known and possibly unique MSA regions, we conducted region-wise mean value comparisons (MSA-C vs. PD and MSA-P vs. PD) at baseline and at the 12-month follow-up. Figure 5 shows the significant effect size differences between the two groups after multiple comparison correction.

We found that at baseline the cerebellum WM as measured by volume HETs showed significant atrophy in both MSA subtypes compared to PD (Fig. 5). In addition, the rostral anterior cingulum cortex (rACC) volume HET value was significantly lower in PD compared to the subtypes. This is in line with previous findings of loss of WM integrity in anterior cingulum in PD<sup>14</sup>. Furthermore, putaminal atrophy across both time points, and the frontal pole and accumbens atrophy at follow-up were significant in MSA-P compared to PD (Fig. 5).

For the dMRI derived HET, the FA HET at baseline showed WM abnormalities in infratentorial regions, including the cerebellum WM, pons, and pontine crossing tract (PCT), in both MSA subtypes compared to PD. Common supratentorial abnormalities were present in the precentral WM (PRCWM) and rectus WM. By the 12-month follow-up, these common FA abnormalities had expanded to include the postcentral WM (POCWM), lateral orbitofrontal WM, the fornix, and fornix-stria in both subtypes and across both timepoints (Fig. 5). Furthermore, several regions also showed subtype-specific FA abnormalities. In MSA-C, cingulum of the hippocampal region (CGH), inferior frontal WM, and posterior limb of the internal capsule (PLIC) were abnormal at both time points. In MSA-P, the cingulate gyrus cingulum (CGC), and superior fronto-



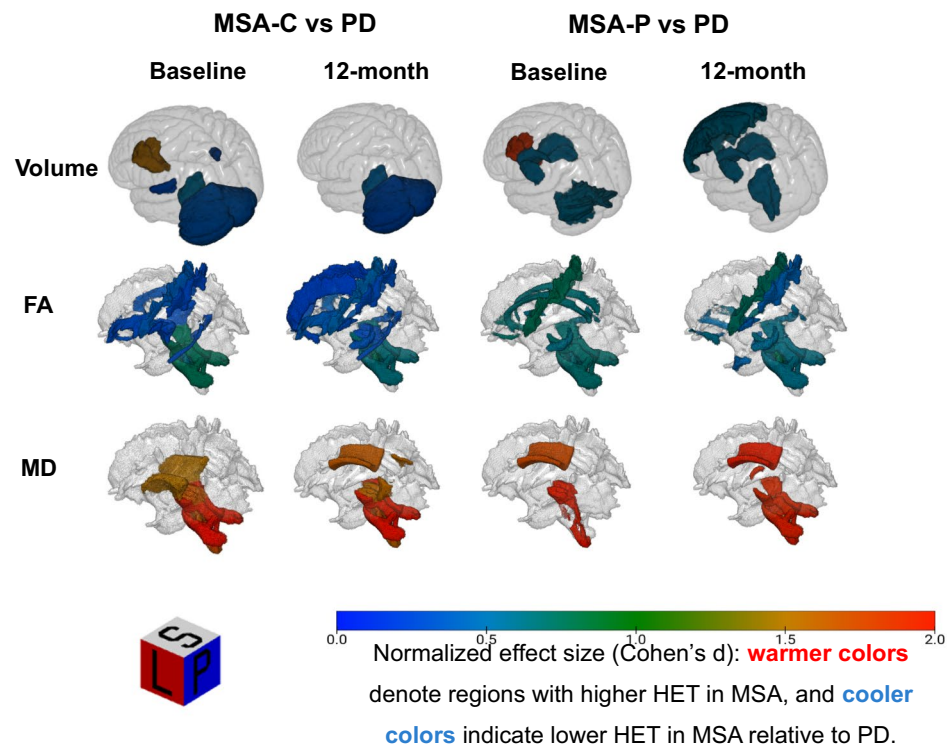
**Fig. 4.** Comparison of (Panel A) cerebellum volume (T1), fractional anisotropy (FA), and mean diffusivity (MD), (Panel B) heterogeneity (HET) scores (volume, FA, and MD), and (Panel C) MSA-atrophy index (AI) versus UMSARS (Unified MSA Rating Scale) total change ( $\Delta$ ) over a 12-month period from baseline are shown for MSA and PD participants. Associations between imaging marker change and UMSARS total change were assessed using Spearman's rho ( $\rho$ ), and p-values correspond to the correlation test. In the legend, the total number of observations (n) is reported, \*  $p < 0.05$ , ns  $p \geq 0.05$ .

orbital fasciculus (SFOF) at baseline, and the entorhinal WM and lateral orbitofrontal WM at follow-up were significantly abnormal (Fig. 5).

MD HET in MSA-C showed significantly elevated values in the cerebellum WM, brainstem, pons, medulla, PCT, anterior limb of the internal capsule (ALIC) and PLIC at baseline. At follow-up, WM MD abnormalities extended to the body of the corpus callosum (BCC) and tapetum (TAP). In MSA-P, MD abnormalities at baseline were restricted to the brainstem, medulla, and BCC. By follow-up, additional abnormalities appeared in the cerebellum WM and the fornix (Fig. 5).

## Discussion

In this study, we demonstrated an ML approach to discriminate MSA from PD using the regional measurements of structural and diffusion MRI. We also implemented a subject-specific score, HET, to serve as a summary measure and as a measure of regional heterogeneity. The HET framework was assessed for its ability to capture clinical and longitudinal disease characteristics, subtype-specific structural and microstructural damage, and its performance compared to cerebellar WM and the MSA-atrophy index (AI). A strength of our study is that by using ML on measurements of the whole brain, we avoided prior assumptions of regional importance attributable to MSA subtypes, hence, allowing a completely data-driven heuristic approach to quantifying the macro and microstructural changes necessary to discriminate between the subtypes and PD. The main findings of our study are: (i) HET performed comparably to cerebellar WM and the MSA-AI for distinguishing MSA from PD, and that (ii) it showed significant associations with clinical progression using UMSARS total over a 12-month period and provided better longitudinal separation of MSA-P from PD than cerebellar WM and MSA-AI; and (iii) the regional HET scores were sensitive to both typical and atypical MSA patterns in both structural and diffusion



**Fig. 5.** Spatial patterns captured by HET. Regional HET effect size maps comparing MSA subtypes and PD at baseline and 12-month follow-up are shown. The colors indicate effect size for significant regions after multiple comparison correction. Warm colors indicate higher regional HET in MSA than PD, and cool colors indicate lower regional HET in MSA than PD. Uncolored regions are not significant. The color bar reflects Cohen's  $d$  and not HET values.

MRI findings; and (iv) most importantly the diffusion derived HETs captured widespread and subtype-specific WM network involvement that aligned well with known MSA pathology.

There is a need for tools capable of detecting MSA in the early disease stages. Established clinical measures such as UMSARS can be insensitive to disease severity<sup>15</sup>. While MRI has been crucial in this endeavor<sup>6,12</sup> advanced modeling is still needed to improve its sensitivity to distinguish between MSA subtypes and PD. DL and ML across various fields of medical image analysis have shown great progress when modeling complex diseases<sup>16–18</sup>. Although ML models are often perceived as “black boxes,” substantial progress over the past decade has produced reliable and reproducible explainable artificial intelligence methods that address this limitation, with ongoing work to refine these approaches<sup>19</sup>. One such method is SHAP<sup>20</sup> which we have exploited in this study. The method relies on cooperative game theory where the goal is to equitably distribute prize to a winning team's players. We applied a similar analogy, asking: for a given classification outcome (MSA vs. PD), how much did each regional MRI measurement contribute to the model's decision? By decomposing predictions into feature-level attributions, we attempted to quantify disease heterogeneity. As shown in Eqs. (1–3) (Methods), we averaged the baseline SHAP contributions across individuals and used these as weights to scale the corresponding raw regional values such that the regions where HET is lower in MSA than PD correspond to not just the raw measurement characteristics but also to the model-identified diagnostic importance attributable to each region. The interpretation of HET values is hence straight forward, for example, for the volumetric measures, lower volume HET in MSA corresponds to more atrophy in MSA relative to PD.

The longitudinal separation between MSA-P and PD obtained using volume HET was better compared to both the cerebellum WM atrophy and MSA-AI (Fig. 3). While the raw cerebellar FA and MD values in older PD participants overlapped with those of MSA-P, the FA and MD HETs were able to separate the two groups. Similarly, the HET scores showed significant correlations with clinical progression, as measured by changes in the UMSARS total, over the 12-month follow-up period. These results highlight HET score's clinical utility potential not only to quantify baseline heterogeneity across disease subtypes, but also to sensitively track longitudinal disease progression. Nonetheless, while HET provided better longitudinal separation, there were few trajectories that did not follow the expected path, which may either reflect genuine subject-level heterogeneity or noise. Furthermore, while the longitudinal and clinical assessments were compared with the cerebellum and MSA-AI, it should be noted, however, that MSA-AI as described in<sup>13</sup>, was calibrated using Human Connectome Project controls rather than controls from the same cohort. These differences could explain its reduced performance in our analysis. In addition, MSA-AI is a singular atrophy marker, whereas the HET framework reflects both structural and microstructural heterogeneity across the entire brain. Thus, the direct comparisons conducted in this study should be interpreted within the proper context and limitations.

Comparing the raw values between MSA subtypes and PD was insensitive to regional differences especially for MSA-P. Repeating the analysis after z-scoring by healthy controls produced the same findings as shown in Fig. 1. On the other hand, the regional volume HET patterns were consistent with the well-established structural degenerations in MSA (Fig. 5). In MSA-C, atrophy was observed in the cerebellar cortex and WM, pons, medulla, and brainstem, matching the OPCA pattern that is characteristic of this subtype. The transverse temporal cortex also showed lower HET values at baseline. Prior studies have reported temporal lobe degeneration in atypical MSA<sup>21–24</sup>, and PD imaging studies have reported reduced volume in the transverse temporal gyrus<sup>25,26</sup>. Its involvement may reflect HET's sensitivity to cortical network changes that occur across synucleinopathies and in atypical MSA presentations. In MSA-P, putaminal and striatal atrophy consistent with the known pattern of SN degeneration were observed. Accumbens and frontal pole also showed lower HET values in MSA-P at follow-up. Accumbens involvement is biologically plausible given its role within the dopaminergic and ventral striatal systems; its atrophy in PD, described as Mavridis' atrophy<sup>27</sup>, has been linked to degeneration of reward-related circuits affected in both MSA and PD<sup>28,29</sup>. Frontal pole atrophy has also been reported in autopsy-confirmed MSA<sup>30</sup> and may reflect later-stage frontal involvement captured by HET.

The FA and MD HET regional patterns suggest spatially and temporally staged WM degeneration in MSA with subtype specific limbic and frontal WM involvement. In MSA-C, the FA HET results showed widespread WM injury at baseline that involved cerebellar and pontine regions, motor projection fibers such as the PLIC and postcentral WM, and limbic and frontal regions including CGH and rectus WM. At follow-up, these abnormalities persisted and further extended into additional frontal regions, including superior frontal and lateral orbitofrontal WM, as well as ALIC and midbrain. In MSA-P, FA HET showed a slightly different trajectory. The CGC and SFOF were significant at baseline but not at follow-up, whereas fornix-stria, postcentral and lateral orbitofrontal WM, and entorhinal WM became abnormal at follow-up which suggests spatial and temporal progressive WM damage. The widespread involvement of motor, limbic and frontal association networks in addition to the expected infratentorial abnormalities are consistent with the widespread WM involvement in MSA which had been reported by del Campo et al.<sup>31</sup>. The MD HET results tell a complementary story that is focused on interhemispheric WM abnormalities. Across both subtypes, MD HET identified abnormalities in the corpus callosum, particularly the BCC and TAP. In MSA-C, the ALIC and PLIC were significantly abnormal at baseline but were no longer significant at follow-up, at which point BCC and TAP emerged as key discriminators. In MSA-P, the BCC was a significant discriminator across both time points, with additional cerebellar and fornix involvement at follow-up. Together, these patterns indicate FA HET captures dynamic, subtype-specific contributions of widespread networks and MD HET capture interhemispheric involvement. Our results are consistent with several prior dMRI studies reporting extensive corticospinal, callosal and limbic WM damage in MSA as well frontal and limbic network involvement<sup>7,32–35</sup>.

The main limitation of this study is the relatively small data size used in model development. Small sample sizes are a common challenge in MSA research due to the rarity of the disease. However, it is worth noting that Mayo Clinic's MONITOR study has one of the world's largest collections of movement disorder patients, making it suitable for an ML application. Nevertheless, to account for potential pitfalls, we implemented repeated cross-validation with multiple random seeds and evaluated models across hundreds of iterations. This rigorous modeling approach provided a broad search space for stable model selection with as little overfitting and data splitting bias as possible. Another limitation was the grouping of MSA-C and MSA-P into one category to avoid further dividing the data into smaller portions. However, this was less of an issue since the longitudinal and clinical correlation results clearly showed the models were able to separate the subtype specific disease characteristics. The resulting HET patterns also aligned well with established pathological and imaging findings which provided validation to our modeling approaches. Nonetheless, further validation is still needed to confirm the results.

In conclusion, our findings demonstrate that heterogeneity scores derived using machine learning can reliably capture the structural and microstructural imaging differences between MSA and PD. The volume, FA, and MD HET measures revealed subtype-specific spatial patterns that closely aligned with established neuropathological hallmarks. These multimodal MRI markers provide a more comprehensive representation of disease burden by improving characterization of MSA heterogeneity. In other words, HET offers an alternative to traditional OPCA and SN markers and to other pre-defined atrophy related indices due to its heuristic approach to regional importance which can be more sensitive to atypical presentations and changes in earlier disease stages. Overall, our findings support the potential of HET as an imaging biomarker framework for tracking disease progression, increasing our mechanistic understanding across atypical parkinsonian syndromes, and ultimately reducing MSA misdiagnosis.

## Methods

### Study participants

Participants enrolled in the Mayo Longitudinal Synucleinopathy Biomarker Study (MONITOR I and II), a prospective and longitudinal study, were included. They were diagnosed with MSA and PD and had obtained standardized quantitative MRI scans at all time points. Patients with MSA-C, MSA-P, and PD were diagnosed by a Mayo Clinic movement disorder specialist based on established criteria<sup>36</sup>. All patients participated in autonomic function testing during their diagnostic assessment. Patients with MSA had to fulfill the consensus criteria for possible or probable MSA and achieve a score of less than 17 (excluding the erectile dysfunction score) on part I of the Unified MSA Rating Scale (UMSARS) to qualify for enrollment, thereby ensuring participation at an early disease stage and aligning with the inclusion criteria for trials of disease-modifying therapies<sup>37–39</sup>. Healthy controls were participants matched for age and sex, showing no signs of neurological disorders or autonomic dysfunction. Participants were generally excluded if they were pregnant or breastfeeding, scored 24 points or lower on the Mini-Mental Status Examination, had a clinically significant or unstable medical or surgical

condition that could hinder safe study completion or influence study results, or had utilized any investigational products within 60 days preceding the baseline assessment.

### Ethics statement and approval

This study was approved by the Mayo Clinic Institutional Review Board (IRB number: 15-005964). The patients were given adequate time to ask questions and think about study participation. Risks, benefits, and alternatives in pursuing this research trial were discussed in detail with the patients. The patients understood the information discussed and agreed to participate in this clinical research study. All questions were answered. Written informed consent was obtained from all participants according to the Declaration of Helsinki. Patients signed the informed consent document prior to any study procedures being performed.

### Clinical assessments

A detailed medical and neurological history was obtained from all participants, followed by a full general and neurological examination. Medications with the potential to influence test results were withheld for five half-lives before neurological assessments, autonomic testing, and MRI acquisition. Neurological impairment in individuals with MSA was rated using the Unified MSA Rating Scale (UMSARS), which includes part I for symptoms and functional status and part II for examination findings<sup>37</sup>. All participants completed standardized autonomic evaluations, including the autonomic reflex screen and the thermoregulatory sweat test. Autonomic deficits were quantified using the Composite Autonomic Severity Score (CASS), a validated measure summarizing the severity and pattern of autonomic dysfunction from these tests<sup>40</sup>. Autonomic symptoms were measured using the Composite Autonomic Symptom Score (COMPASS)<sup>41</sup>.

### Imaging acquisition and processing

MRI data were acquired on a 3-T Siemens Prisma whole body scanner (Siemens Medical Systems, Erlangen, Germany) using a 32-channel head coil.

#### Structural MRI

High-resolution T1-weighted (T1) 3D structural images were acquired using an MPRAGE sequence with 3D distortion correction. Imaging parameters were repetition time (TR) 2300 ms, echo time (TE) 2.95 ms, flip angle 9°, voxel dimensions 1.05 × 1.05 × 1.20 mm, acquisition matrix 256 × 240, and a total scan duration of 312 s across 176 sagittal slices. Then, trained image analysts reviewed all the data. Shading artifacts in the T1 scans were corrected using SPM12 segmentation combined with N3. Regional MRI morphometry was then derived with FreeSurfer v6.0 using the Desikan–Killiany atlas<sup>42</sup>. Middle cerebellar peduncle atrophy, which is common in MSA, is captured within the cerebellar region in FreeSurfer. Regional volumes were expressed as fractions of total intracranial volume (TIV), with TIV estimated in house<sup>43</sup>, and these normalized measures were used as morphometric features for the analyses.

#### Diffusion MRI

The diffusion MRI (dMRI) scans were acquired using a multiband (3 x slice acceleration) single-shot spin-echo axial EPI sequence with the following settings: TR 3400 ms, TE 71 ms, flip angle 90°, acquisition matrix 116 × 116, 2.0-mm isotropic voxels, and NEX 1. Three diffusion weightings were collected: 16 volumes at  $b = 0$ , 48 volumes at  $b = 1000$ , and 64 volumes at  $b = 2000$  s/mm<sup>2</sup>. Gradient directions were uniformly distributed across the sphere for all diffusion shells<sup>44</sup>. Then, to process the dMRI images an intracranial mask was first created for each scan<sup>45</sup>. Noise in the raw diffusion data was estimated and removed, motion and eddy current distortions were corrected, Gibbs ringing was eliminated, and Rician bias was adjusted<sup>42</sup>. Diffusion tensors for the multi shell dataset were then estimated using the nonlinear least squares algorithm implemented in *dipy*<sup>46</sup>, including all  $b$  values in the tensor calculation to maximize SNR. From these tensors, Fractional Anisotropy (FA) and Mean Diffusivity (MD) were computed. Each subject's FA image was nonlinearly aligned to an in-house modified JHU “Eve” white matter atlas using ANTS<sup>47</sup>, enabling extraction of regional median FA and MD. Voxels with MD values greater than  $2 \times 10^{-3}$  or less than  $7 \times 10^{-5}$  mm<sup>2</sup>/s were removed as likely CSF or air. ROIs containing fewer than seven diffusion voxels in subject space were excluded due to unreliable registration. MD values were multiplied by  $10^6$  to simplify interpretation.

### Computation of the heterogeneity (HET) score

#### Training the ML model

Three separate classifiers were run for volume, fractional anisotropy (FA), and mean diffusivity (MD) regional values as inputs and a binary target of 0 for MSA and 1 for PD. The regional input feature sets comprised 49 volume, 43 FA, and 49 MD features. To ensure reliable biological signal, FA was excluded for some regions that are predominantly gray matter<sup>48,49</sup>. The complete atlas segmentation and region list are provided in Supplementary Fig. 1.

The MSA subtypes were grouped as one label since a three-way classifier was not possible with the limited data size. Age was included in all models as a covariate. Sex was not included as a covariate due to the small number of female participants (MSA-C  $n = 5$ , MSA-P  $n = 4$ , PD  $n = 1$ )<sup>50</sup>. Before model training, the regional volume measurements were normalized by total intracranial volume to account for head size differences. In addition, all inputs were z-scored against controls so that each feature reflected deviations from a healthy population.

There are numerous types of ML models with varying characteristics and hyperparameter requirements. To minimize the risk of selection bias from any single model, we trained 4 individual models and an additional AutoML framework and chose the best performer. The 4 models were XGBoost<sup>51</sup>, Random forest<sup>52</sup>, LightGBM<sup>53</sup>, and CatBoost<sup>54</sup> and the AutoML framework was AutoGluon (v1.1.1)<sup>55</sup>. Hyperparameters used in training are

given in Supplementary Table 1. Because of the limited data size, cross-validation was preferred over a single train-test split. The data was divided into  $k$  folds and repeatedly trained and tested on randomly shuffled partitions. The folds were grouped by subject IDs to prevent participants in the training set from appearing in the test set, thereby avoiding data leakage. The train and test partitions within the folds were stratified based on the binary target so that equal proportions of MSA and PD samples were maintained. We used 3 folds corresponding to 67% training and 33% testing split. The number of folds was chosen as 3 so as not to compromise the proportions of binary targets in the splits, i.e., higher  $k$  folds result in fewer number of the MSA and PD in the splits.

To further reduce potential bias during fold splitting, all models were run using 10 seeds, so that each fold split was as random as possible for each seed run. Each individual models were optimized using a randomized grid search with 10 repetitions, resulting in 30 fits across the 3 folds. AutoGluon was trained on the same folds for each seed run, using its built-in optimization techniques such as repetitions and ensembling to identify the ideal configuration. Within each seed, every model's F1 score, and balanced accuracy were computed for both the training and test splits. The model (either an individual classifier or AutoGluon) with the highest mean F1 score and the lowest standard deviation across the three test folds was identified as the winner for that seed. Among all seed-level winners, the one with the highest F1 score was selected as the final best model. This selection procedure was repeated independently for the volume, FA, and MD analyses. Lastly, the final model was explained using SHAP (SHapley Additive exPlanations) (v0.44.1) to assess the contribution of each feature to the model predictions<sup>20</sup>. Extensive literature exists on SHAP's methodology and biomedical applications, refer for example<sup>56</sup>. For added stability and reproducibility, we implemented a bootstrapping ( $n = 200$ ) technique and then took their average for the final SHAP values. All codes are available in our online repository (<https://github.com/RobelGebre/HET>).

#### Heterogeneity (HET) score

To capture spatial heterogeneity in the brain macrostructure and microstructure, we created three HET scores: volume, FA, and MD, each derived from independently trained models. We have previously implemented a similar application for deriving a heterogeneity score using SHAP values for quantifying abnormal tau protein deposition in Alzheimer's disease<sup>57</sup>.

We first compute feature attributions for each diagnostic group,  $dx$ , using SHAP. Baseline SHAP values were then used as regional weights to quantify the heterogeneous contribution of each brain region. The SHAP framework is cross-sectional and does not directly model temporal dynamics, hence only the baseline explanations were used as weights.

Let  $\phi_{i,j,dx}^{t=0}$  denote the regional SHAP value of the ROI features  $j$  for subject  $i$  ( $i = 1, \dots, M; j = 1, \dots, N$ ); then the corresponding baseline feature weights  $\phi_{j,dx}^{t=0}$  were defined as in Equation (Eq. 1). In practice, because SHAP values sum up to the model's predicted probability (ranging from 0 to 1), we can optionally multiply the weights in Eq. 1 by a large constant factor (e.g., 100) to improve numerical stability without altering relative importance.

Next, we defined regional HET by applying the SHAP weights to the corresponding regional measurements  $x_{i,j}^{t \geq 0}$ , producing the regional measures of heterogeneity (Eq. 2). Finally, the subject-level HET score for each subject was computed by averaging the weighted regional values across all ROI (Eq. 3). Throughout the manuscript, unless explicitly stated as regional, "HET" refers to the subject-level HET score.

$$\phi_{j,dx}^{t=0} = \frac{1}{M} \sum_{i=1}^M \phi_{i,j,dx}^{t=0} \quad (1)$$

$$\tilde{x}_{i,j} = \phi_{j,dx}^{t=0} x_{i,j}^{t \geq 0} \quad (2)$$

$$HET_i = \frac{1}{N} \sum_{j=1}^N \phi_{j,dx}^{t=0} x_{i,j}^{t \geq 0} \quad (3)$$

#### Statistical analysis

The MSA-AI, as described in<sup>13</sup>, was computed from volumetric measures of three brain structures: lentiform nucleus consisting of putamen and pallidum, brainstem, and cerebellum. Z-scores for each region were derived by subtracting the predicted mean and dividing by the standard deviation from the control population. The mean were estimated via linear regression adjusted for age and sex<sup>13</sup>. The final index was calculated as the average of the three regional z-scores. It should be noted that MSA-AI is a volumetric calculation of atrophy and not intended for WM microstructure quantification, hence we compared its performance to only the volume derived HET scores.

The ML model performances were evaluated using the area under the curve (AUC) from the receiver operating characteristic curves (ROC) and F1 scores. The ROC AUC measures a model's ability to distinguish between classes with values closer to 1.0 showing better performance. F1 score indicates classification accuracy by balancing false positives and false negatives.

Clinical validation of HET was performed by relating 12-month change from baseline in HET to the corresponding 12-month change from baseline in UMSARS total ( $\Delta = x_{t=12 \text{ months}} - x_{t=0}$ ). The same change-to-change analyses were performed for the MSA-AI and cerebellar WM for comparison. The goal of these analyses was to assess whether earliest visit point changes in HET track longitudinal clinical change in a manner comparable in direction to established imaging markers. For visualization only, scatter plots were Winsorized using percentile clipping to reduce the influence of extreme values on axis scaling<sup>58</sup>.

Mean differences at the region and group-level between MSA subtypes and PD was analyzed using independent samples t-test or Mann-Whitney U-test followed by multiple comparison correction using false discovery rate (FDR)<sup>59</sup>. The appropriate test was determined after checking for normality using Shapiro-Wilk. Effect sizes were evaluated using Cohen's *d* which is defined as the difference in the group means divided by the pooled standard deviation. Cohen's *d* was used to demonstrate separation only after the appropriate test and multiple comparison corrections were conducted. The spearman's rho ( $\rho$ ) was used to assess correlations between clinical scores and the cerebellar WM, MSA-AI, and HET; corresponding two-sided p-values are reported.

## Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request. All the codes are publicly available at <https://github.com/RobelGebre/HET>.

Received: 9 December 2025; Accepted: 23 March 2026

Published online: 30 March 2026

## References

1. Yamasaki, T. R. et al. Parkinson's disease and multiple system atrophy have distinct  $\alpha$ -synuclein seed characteristics. *J. Biol. Chem.* **294**, 1045–1058 (2019).
2. Antonina, K., Kelli, M. & Wei-Li, K. T. Parkinson's disease: etiology, neuropathology, and pathogenesis. In *Parkinson's Disease: Pathogenesis and Clinical Aspects* 3–26. <https://doi.org/10.15586/codonpublications.parkinsonsdisease.2018.ch1> (Codon Publications, 2018).
3. Jellinger, K. A. Multiple System Atrophy: An Oligodendroglioneuronal Synucleinopathy. *J. Alzheimer's Dis.* **62**, 1141–1179 (2018).
4. Fanciulli, A. et al. Elsevier, Multiple system atrophy. In *International Review of Neurobiology* **149** 137–192 (2019).
5. Chelban, V. et al. An update on advances in magnetic resonance imaging of multiple system atrophy. *J. Neurol.* **266**, 1036–1045 (2019).
6. Raghavan, S. et al. White Matter Abnormalities Track Disease Progression in Multiple System Atrophy. *Mov. Disord. Clin. Pract.* **11**, 1085–1094 (2024).
7. Ogawa, T. et al. White matter and nigral alterations in multiple system atrophy-parkinsonian type. *Npj Park Dis.* **7**, 96 (2021).
8. Pasquini, J., Firbank, M. J., Ceravolo, R., Silani, V. & Pavese, N. Diffusion Magnetic Resonance Imaging Microstructural Abnormalities in Multiple System Atrophy: A Comprehensive Review. *Mov. Disord.* **37**, 1963–1984 (2022).
9. Kim, H. J., Stamelou, M. & Jeon, B. Multiple system atrophy-mimicking conditions: Diagnostic challenges. *Parkinsonism Relat. Disord.* **22**, S12–S15 (2016).
10. Litvan, I. What Is the Accuracy of the Clinical Diagnosis of Multiple System Atrophy? A Clinicopathologic Study. *Arch. Neurol.* **54**, 937 (1997).
11. Kiryu, S. et al. Deep learning to differentiate parkinsonian disorders separately using single midsagittal MR imaging: a proof of concept study. *Eur. Radiol.* **29**, 6891–6899 (2019).
12. Vemuri, P. et al. Imaging biomarkers for early multiple system atrophy. *Parkinsonism Relat. Disord.* **103**, 60–68 (2022).
13. Trujillo, P. et al. The MSA Atrophy Index (MSA-AI): An Imaging Marker for Diagnosis and Clinical Progression in Multiple System Atrophy. *Ann. Clin. Transl. Neurol.* **12**, 1823–1833 (2025).
14. De Schipper, L. J., Van Der Grond, J., Marinus, J., Henselmans, J. M. L. & Van Hilten, J. J. Loss of integrity and atrophy in cingulate structural covariance networks in Parkinson's disease. *NeuroImage Clin.* **15**, 587–593 (2017).
15. Palma, J. A. et al. Limitations of the Unified Multiple System Atrophy Rating Scale as outcome measure for clinical trials and a roadmap for improvement. *Clin. Auton. Res.* **31**, 157–164 (2021).
16. Zuo, S., Li, Y., Qi, Y. & Liu, A. Multilevel correlation-aware and modal-aware graph convolutional network for diagnosing neurodevelopmental disorders. *IEEE Trans. Biomed. Eng.* 1–14. <https://doi.org/10.1109/TBME.2025.3617348> (2025).
17. Wang, Y. et al. Integrating Clinical Knowledge Graphs and Gradient-Based Neural Systems for Enhanced Melanoma Diagnosis via the Seven-Point Checklist. *IEEE Trans. Neural Netw. Learn. Syst.* **37**, 37–51 (2026).
18. Dorfner, F. J., Patel, J. B., Kalpathy-Cramer, J., Gerstner, E. R. & Bridge C. P. A review of deep learning for brain tumor analysis in MRI. *Npj Precis Oncol.* **9**, 2 (2025).
19. Saeed, W., Omlin, C. & Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowl.-Based Syst.* **263**, 110273 (2023).
20. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions.
21. Aoki, N. Atypical multiple system atrophy is a new subtype of frontotemporal lobar degeneration: frontotemporal lobar degeneration associated with  $\alpha$ -synuclein.
22. Piao, Y. S. et al. Co-localization of  $\alpha$ -synuclein and phosphorylated tau in neuronal and glial cytoplasmic inclusions in a patient with multiple system atrophy of long duration. *Acta Neuropathol. (Berl.)* **101**, 285–293 (2001).
23. Shibuya, K. et al. Asymmetrical temporal lobe atrophy with massive neuronal inclusions in multiple system atrophy. *J. Neurol. Sci.* **179**, 50–58 (2000).
24. Jellinger, K. A. Heterogeneity of Multiple System Atrophy: An Update. *Biomedicines* **10**, 599 (2022).
25. Çavuşoğlu, B. et al. Cortical Thickness Alterations in Parkinson's Disease with Mild Cognitive Impairment. *Turk. J. Neurol.* **29**, 126–133 (2023).
26. Yuan, J. et al. Alterations in cortical volume and complexity in Parkinson's disease with depression. *CNS Neurosci. Ther.* **30**, e14582 (2024).
27. Mavridis, I. N. & Pyrgelis, E. S. Nucleus accumbens atrophy in Parkinson's disease (Mavridis' atrophy): 10 years later.
28. Abos, A. et al. Differentiation of multiple system atrophy from Parkinson's disease by structural connectivity derived from probabilistic tractography. *Sci. Rep.* **9**, 16488 (2019).
29. Jellinger, K. A. The Pathobiology of Behavioral Changes in Multiple System Atrophy: An Update. *Int. J. Mol. Sci.* **25**, 7464 (2024).
30. Konagaya, M., Sakai, M., Matsuoka, Y., Konagaya, Y. & Hashizume, Y. Multiple system atrophy with remarkable frontal lobe atrophy. *Acta Neuropathol. (Berl.)* **97**, 423–428 (1999).
31. Del Campo, N. et al. Broad white matter impairment in multiple system atrophy. *Hum. Brain Mapp.* **42**, 357–366 (2021).
32. Hara, K. et al. Corpus callosal involvement is correlated with cognitive impairment in multiple system atrophy. *J. Neurol.* **265**, 2079–2087 (2018).
33. Ji, L., Wang, Y., Zhu, D., Liu, W. & Shi, J. White matter differences between multiple system atrophy (parkinsonian type) and Parkinson's disease: A diffusion tensor image study. *Neuroscience* **305**, 109–116 (2015).
34. Worker, A. et al. Diffusion Tensor Imaging of Parkinson's Disease, Multiple System Atrophy and Progressive Supranuclear Palsy: A Tract-Based Spatial Statistics Study. *PLoS ONE* **9**, e112638 (2014).
35. Minnerop, M. et al. Callosal tissue loss in multiple system atrophy—A one-year follow-up study. *Mov. Disord.* **25**, 2613–2620 (2010).

36. Gilman, S. & Wenning, G. K. Second consensus statement on the diagnosis of multiple system atrophy.
37. Wenning, G. K. et al. Development and validation of the Unified Multiple System Atrophy Rating Scale (UMSARS). *Mov. Disord.* **19**, 1391–1402 (2004).
38. Levin, J. et al. Safety and efficacy of epigallocatechin gallate in multiple system atrophy (PROMESA): a randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* **18**, 724–735 (2019).
39. Low, P. A. et al. Efficacy and safety of rifampicin for multiple system atrophy: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* **13**, 268–275 (2014).
40. Low, P. A. Composite Autonomic Scoring Scale for Laboratory Quantification of Generalized Autonomic Failure. *Mayo Clin. Proc.* **68**, 748–752 (1993).
41. Lipp, A. et al. Prospective differentiation of multiple system atrophy from Parkinson disease, with and without autonomic failure. *Arch Neurol.* **66**, (2009).
42. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
43. Schwarz, C. G. et al. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *NeuroImage Clin.* **11**, 802–812 (2016).
44. Caruyer, E., Lenglet, C., Sapiro, G. & Deriche, R. Design of multishell sampling schemes with uniform coverage in diffusion MRI. *Magn. Reson. Med.* **69**, 1534–1540 (2013).
45. Reid, R. I., Nedelska, Z., Schwarz, C. G., Ward, C. & Jack, C. R. Diffusion specific segmentation: skull stripping with diffusion MRI data alone. In *Computational Diffusion MRI* (eds Kaden, E., Grussu, F., Ning, L., Tax, C. M. W. & Veraart, J.) 67–80. (Springer International Publishing, 2018).
46. Garyfallidis, E. et al. Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinformatics* **8**, (2014).
47. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* **54**, 2033–2044 (2011).
48. Jones, D. K. & Cercignani, M. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR Biomed.* **23**, 803–820 (2010).
49. Seo, Y., Rollins, N. K. & Wang, Z. J. Reduction of bias in the evaluation of fractional anisotropy and mean diffusivity in magnetic resonance diffusion tensor imaging using region-of-interest methodology. *Sci. Rep.* **9**, 13095 (2019).
50. Kaplan, S. Prevalence of multiple system atrophy: A literature review. *Rev. Neurol. (Paris)*. **180**, 438–450 (2024).
51. Chen, T., Guestrin, C. & XGBoost: a scalable tree boosting system. 785–794 <https://doi.org/10.1145/2939672.2939785> (2016).
52. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
53. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree.
54. Dorogush, A. V., Ershov, V. & Gulin, A. CatBoost: gradient boosting with categorical features support. <https://doi.org/10.48550/arXiv.1810.11363> (2018).
55. Erickson, N. et al. AutoGluon-Tabular: robust and accurate AutoML for structured data. <http://arxiv.org/abs/2003.06505> (2020).
56. Gramegna, A. & Giudici, P. S. H. A. P. An Evaluation of Discriminative Power in Credit Risk. *Front. Artif. Intell.* **4**, 752558 (2021).
57. Gebre, R. K. et al. Advancing Tau PET quantification in Alzheimer disease with machine learning: introducing THETA, a novel Tau summary measure. *J. Nucl. Med.* <https://doi.org/10.2967/jnumed.123.267273> (2024).
58. Wilcox, R. R. & Keselman, H. J. Modern Regression Methods that can Substantially Increase Power and Provide a more Accurate Understanding of Associations. *Eur. J. Personal.* **26**, 165–174 (2012).
59. Noble, W. S. How does multiple testing correction work? *Nat. Biotechnol.* **27**, 1135–1137 (2009).

## Author contributions

R.K.G., W.S., and P.V., contributed toward idea, conception, and design of the study. R.K.G. conducted all analyses, results, and writing of the manuscript. S.R. contributed to data analysis and interpretations. M.E.J.T. performed visual quality checks and post-processing on the images used in the study. A.J.F. contributed to the statistical analysis. R.R. analyzed and processed the diffusion images. P.A.L. contributed to the interpretation and manuscript critique. All authors contributed to the review and critique of the final manuscript.

## Funding

This study was supported by NIH (R01NS092625, R01 NS097495, U19 AG71754, UL1 TR000135), FDA (R01 FD07290), grants from the Michael J. Fox Foundation for Parkinson's disease, Sturm Foundation, Bishop Dr. Karl Golser Foundation, Mayo Center of Regenerative Medicine, and Mayo Funds.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-45949-5>.

**Correspondence** and requests for materials should be addressed to R.K.G., W.S. or P.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026