



## OPEN ACCESS

## EDITED BY

Chuanming Li,  
Chongqing University Central Hospital, China

## REVIEWED BY

Yang Xiang,  
University of Electronic Science and  
Technology of China, China  
Zhaohui Yao,  
Renmin Hospital of Wuhan University, China

## \*CORRESPONDENCE

Heng Yang  
✉ 283810363@qq.com  
Yili Deng  
✉ 405502269@qq.com

<sup>†</sup>These authors have contributed equally to  
this work and share first authorship

RECEIVED 19 June 2025

ACCEPTED 18 August 2025

PUBLISHED 08 September 2025

## CITATION

Li Z, Zhang W, Yang R, Chen D, Li X, Wang K,  
Cheng L, Yang H and Deng Y (2025)  
Development of a radiomics-based model for  
diagnosis of multiple system atrophy using  
multimodal MRI. *Front. Neurol.* 16:1650350.  
doi: 10.3389/fneur.2025.1650350

## COPYRIGHT

© 2025 Li, Zhang, Yang, Chen, Li, Wang,  
Cheng, Yang and Deng. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Development of a radiomics-based model for diagnosis of multiple system atrophy using multimodal MRI

Zhichao Li<sup>1†</sup>, Wei Zhang<sup>1†</sup>, Ran Yang<sup>1</sup>, Dong Chen<sup>2</sup>, Xin Li<sup>2</sup>,  
Kun Wang<sup>2</sup>, Lei Cheng<sup>3</sup>, Heng Yang<sup>3\*</sup> and Yili Deng<sup>3\*</sup>

<sup>1</sup>Department of Radiology, Chongqing Western Hospital, Chongqing, China, <sup>2</sup>Department of  
Radiology, Second People's Hospital of Jiulongpo District, Chongqing, China, <sup>3</sup>Department of Internal  
Medicine, Second People's Hospital of Jiulongpo District, Chongqing, China

**Introduction:** Multiple system atrophy (MSA) is a rapidly progressive neuro-degenerative disorder characterized by autonomic dysfunction, levodopa-unresponsive parkinsonism, cerebellar ataxia, and corticospinal tract involvement. Early diagnosis remains challenging due to overlapping clinical manifestations and the absence of reliable biomarkers. This study aimed to develop a radiomics-based diagnostic model using multimodal MRI to improve MSA detection.

**Methods:** A retrospective cohort of 62 clinically probable MSA patients (per the 2022 Movement Disorder Society criteria), and 73 matched healthy controls underwent 3.0-T MRI (T1WI, T2WI, FLAIR, DWI). Seven brain regions (bilateral cerebellar hemispheres, middle cerebellar peduncles, putamen, and pons) were manually segmented. A total of 1,502 radiomics features were extracted per region, using PyRadiomics (IBSI-compliant). Features with an intraclass correlation coefficient (ICC)  $\geq 0.75$  were retained, and the least absolute shrinkage and selection operator (LASSO) regression identified the top discriminative features to construct region-specific radiomics scores (Rad-scores). A logistic regression (LR) model integrated Rad-scores from all regions. Model performance was evaluated via precision, recall, and F1-score in training, testing, and validation cohorts (split ratio 6:2:2), and compared with visual assessments by two radiologists.

**Results:** The LR model achieved high performance: accuracy was 0.98 in the training cohort, 0.97 in the testing cohort, and 0.95 in the validation cohort. Notably, classification precision for MSA reached 1.0 (indicating no false positives) across all cohorts. SHapley Additive exPlanations (SHAP) analysis revealed that the left putamen Rad-score as the most influential predictor. The model significantly outperformed radiologists' visual assessments (radiologist AUCs: 0.559 and 0.535;  $P < 0.001$ ). Asymmetry was observed, with left-hemisphere structures (putamen/cerebellar) exhibiting greater diagnostic contributions.

**Conclusion:** Multimodal MRI radiomics accurately differentiates MSA from healthy controls, even in the absence of conventional MRI markers. The Rad-score model demonstrates high sensitivity (89% recall in the validation cohort) and perfect specificity (100% precision), providing a clinically actionable tool for early MSA diagnosis.

## KEYWORDS

radiomics, magnetic resonance imaging (MRI), diagnostic model, machine learning, multiple system atrophy, neurodegenerative disorders

## Introduction

Multiple system atrophy (MSA) is a neurodegenerative disorder of unknown etiology and insidious onset, characterized primarily by autonomic dysfunction, poorly levodopa-responsive parkinsonism, cerebellar ataxia, and corticospinal tract dysfunction (1). MSA diagnosis remains challenging due to overlapping clinical manifestations with other neurodegenerative diseases and the lack of reliable biomarkers (2, 3). Epidemiological studies indicate that MSA progresses rapidly with shortened survival, underscoring the critical importance of early diagnosis for symptom management, prognosis evaluation, precision therapy development, and drug discovery (4).

Historically, MSA diagnosis relied on clinical symptoms, signs, and neuroimaging findings (5, 6). Although neuropathological examination remains the gold standard, biopsy-associated risks and patient reluctance limit its utility. Clinical diagnosis alone faces limitations due to phenotypic heterogeneity and symptom overlap across neurodegenerative disorders. Consequently, neuroimaging has been incorporated as supportive evidence in diagnostic criteria (7, 8). Previous studies identified key MRI features: in the MSA-P subtype: Hypointensity in the putamen on T2-weighted imaging (T2WI) and susceptibility-weighted imaging (SWI), with hyperintensity on T2\* sequences (9); in the MSA-C subtype: the “hot cross bun sign” (pontine cruciform hyperintensity on T2WI/FLAIR) and middle cerebellar peduncle (MCP) hyperintensity. The “hot cross bun sign” exhibits 99% specificity and 45% sensitivity in differentiating MSA-C from spinocerebellar ataxias, while MCP hyperintensity shows 99% specificity and 68% sensitivity (10). The grading of pontine “hot cross bun sign” (11) correlates positively with cerebellar ataxia severity in MSA-C. These characteristic MRI markers aid in distinguishing MSA from Parkinson’s disease (PD), progressive supranuclear palsy (PSP), and sporadic late-onset ataxia, though sensitivity in early-stage disease remains suboptimal (12). While PET-CT and SPECT offer diagnostic value, high cost and radiation exposure hinder widespread clinical adoption (13, 14). Transcranial sonography further suffers from limited sensitivity and specificity (15).

In 2022, the International Movement Disorder Society updated diagnostic criteria, stratifying MSA into four tiers: Neuropathologically established, Clinically established, Clinically probable, Possible prodromal MSA (6). The same year, China released its expert consensus, aligning with international standards while incorporating regional evidence (16). This consensus explicitly mandates multimodal MRI—including T1 (axial/sagittal), T2, ADC, SWI, and T2 FLAIR sequences—as essential for diagnosis, differential evaluation, and disease monitoring. It emphasizes that precise diagnosis requires integrating clinical, imaging, and laboratory data, highlighting the need for novel methods to enhance diagnostic accuracy (12).

Despite these advances, there remains a pressing need for more sensitive and objective imaging diagnostic model. This study aims to develop optimal diagnostic model for MSA based on radiomics features derived from multimodal MRI, providing a novel and precise diagnostic tool for clinical practice.

## Materials and methods

### Subjects

This retrospective study analyzed image data from 69 patients with multiple system atrophy (MSA) admitted to the Second People’s Hospital of JiuLongPo district between October 2022 and June 2024. All patients underwent brain MRI prior to admission. Patients were included if they met the following criteria: (1) Diagnosis of clinically probable MSA according to the 2022 International Movement Disorder Society (MDS) diagnostic criteria (6); (2) Completion of standardized brain MRI protocols, including T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), fluid-attenuated inversion recovery (FLAIR), and diffusion-weighted imaging (DWI); (3) No treatments potentially affecting MRI findings within 3 months before enrollment. Patients were excluded for: (1) Comorbid neurological disorders (e.g., stroke, other neurodegenerative diseases); (2) Use of neuroactive medications within 3 months; (3) History of neurosurgery altering brain structure; (4) Incomplete MRI sequences (missing T1WI, T2WI, or T2-FLAIR) or significant artifacts compromising image quality. Based on these criteria, 7 patients were excluded (1 with a history of cerebral hemorrhage, 4 with cerebral infarction lesions, 2 with severe MRI artifacts). Ultimately, 62 patients with clinically probable MSA were included. Healthy normal controls ( $n = 73$ ) were selected from individuals undergoing routine brain MRI at the hospital’s health examination center during the same period. Controls were matched to patients for age, sex, and educational level. Exclusion criteria for controls: Family history of neurological disorders; Use of centrally acting medications; MRI evidence of asymptomatic cerebral infarction or white matter hyperintensities.

This retrospective study was approved by the Ethics Committee of the Second People’s Hospital of JiuLongPo district. Written informed consent was waived in accordance with national ethical guidelines due to the retrospective nature of the research (17).

### MRI acquisition protocol

All participants underwent brain MRI in the supine position using a 3.0-T scanner (Siemens VIDA, Siemens Healthineers, Erlangen, Germany). Imaging was performed with body coil transmission and 20-channel phased-array head/neck coil for signal reception. The standardized protocols included:

1. T1-weighted Imaging (T1WI): Sequence: fast low-angle shot (FLASH), Orientation: Axial, Parameters: TR = 236 ms, TE = 2.46 ms, Slice thickness = 5 mm, FOV =  $220 \times 220 \text{ mm}^2$ , Matrix =  $202 \times 288$ , Averages = 1.
2. T2-weighted Imaging (T2WI): Sequence: turbo spin echo (TSE), Orientation: Axial, Parameters: TR = 1,500 ms, TE = 80 ms, Echo train length = 198, Slice thickness = 5 mm, FOV =  $220 \times 220 \text{ mm}^2$ , Matrix =  $256 \times 320$ , Averages = 1.
3. T2-fluid-attenuated inversion recovery (FLAIR): Sequence: turbo inversion recovery spin echo, Orientation: Axial, Parameters: TR = 9,000 ms, TE = 84 ms, Inversion time (TI)

= 2,500 ms, Slice thickness = 5 mm, FOV =  $220 \times 220$  mm<sup>2</sup>, Matrix =  $192 \times 256$ , Parallel imaging acceleration factor = 1.

4. Diffusion-weighted Imaging (DWI): Sequence: single-shot echo planar imaging (SS-EPI), Orientation: Axial, Parameters: TR = 4,200 ms, TE = 68 ms, *b*-values = 0 and 1,000 s/mm<sup>2</sup>, Slice thickness = 5 mm, FOV =  $220 \times 220$  mm<sup>2</sup>, Matrix =  $116 \times 120$ , Number of diffusion directions = 3.

Imaging coverage extended from the vertex to the foramen magnum, encompassing the entire cerebrum, brainstem, and cerebellum.

## Image processing and feature extraction

All imaging data were exported from the scanner in DICOM format and converted to NIfTI format using MRICroGL software (v2.1.60; Chris Rorden, University of South Carolina, USA). The resulting NIfTI files were imported into the open-source medical imaging platform 3D Slicer (18) (v5.7.0; Slicer Community, <http://www.slicer.org>) for subsequent processing.

Segmentation of seven brain regions was independently performed by two certified radiologists (each with more than 10 years of specialized experience): left cerebellar, left middle cerebellar peduncle (MCP), left putamen, pons, right cerebellar, right MCP, and right putamen (6).

The segmentation workflow included: the ROIs of T1WI, T2WI, FLAIR and ADC sequences were manually delineated along the boundaries of the above brain regions, and the volume of interest (VOI) of each brain region was constructed by ROI interpolation (19, 20).

Standardized radiomics feature extraction was performed through a three-stage protocol: (1) Segmented images underwent isotropic resampling to a uniform voxel resolution of 1 mm<sup>3</sup> using third-order B-spline interpolation to minimize interpolation artifacts; (2) Feature calculation was executed via the open-source Python library PyRadiomics (v3.1.0a2) (21), with all parameters strictly compliant with the Image Biomarker Standardization Initiative (IBSI) guidelines (21) to ensure reproducibility; (3) Four feature classes were extracted, including morphological features from original images to quantify volumetric and shape characteristics (e.g., sphericity, surface area), texture features from original images capturing spatial intensity heterogeneity (e.g., gray-level co-occurrence matrix metrics), frequency-domain features derived from wavelet-transformed images for multiscale frequency component analysis (e.g., Haar wavelet decompositions), and edge-enhanced features generated via Laplacian of Gaussian (LoG) filtering ( $\sigma = 1.0$ – $7.0$  mm) to accentuate microstructural boundaries and high-frequency details.

Multimodal radiomics feature integration was achieved by concatenating 1,502 radiomics features extracted from each brain region in each sequence.

Mathematical definitions of texture features followed the PyRadiomics documentation (<https://pyradiomics.readthedocs.io/latest/features.html>).

The complete technical workflow is illustrated in Figure 1.

## Radiomics feature selection

To ensure robustness of radiomics features, 70% of randomly selected samples ( $n = 94/135$ ) were allocated for feature extraction. Regions of interest (ROIs) were independently delineated by two certified radiologists (each with > 10 years of experience) following a standardized workflow described above, and features were extracted uniformly. Inter-observer agreement was evaluated using the intraclass correlation coefficient (ICC). Features demonstrating high reproducibility ( $\text{ICC} \geq 0.75$ ) were retained for subsequent analysis.

Retained features were Z-score normalization to eliminate scale differences, followed by application of the least absolute shrinkage and selection operator (LASSO) algorithm for region-specific feature selection. The optimal penalty coefficient ( $\lambda$ ) was determined via 10-fold cross-validation (22), and the top five features with highest discriminative power per brain region were selected.

## Diagnostic model development and evaluation

The radiomics score (Rad-score) for each brain region was calculated as

$$\text{Rad-score} = \sum (\text{Feature Value} \times \text{Feature Weight}) + b_0, \quad (1)$$

where Feature Weight denotes the coefficient derived from selected features, and  $b_0$  represents the intercept term.

For all 135 samples, region-specific Rad-scores were calculated to generate a multi-regional biomarker matrix comprising seven Rad-scores per subject. The dataset was split into training, testing, and validation cohorts in a 6:2:2 ratio ( $n = 81/27/27$ ). A logistic regression (LR) model integrated the seven regional Rad-scores. In order to improve the stability of model evaluation, hierarchical 10-fold cross-validation is used on the training set (22), and the hyperparameters of each algorithm were optimized by grid search to determine the optimal parameter combination. The training curve was plotted on the training set to assess model performance. Classification reports were computed for both testing and validation cohorts.

The machine learning model was implemented using the scikit-learn Python library (version 1.5.1). The model performance was assessed using the area under the receiver operating characteristic curve (AUC) of the test set and the classification report, and the SHapley Additive exPlanations (SHAP) method was used to analyze the feature contribution and the decision logic of the model (23, 24). Finally, a nomogram was constructed to visualize the prediction results.

## Visual assessment of MRI scans

Two certified radiologists (each with >10 years of experience) independently performed a blinded assessment of 135 samples to evaluate suspicion of multiple system atrophy (MSA) diagnosis. This evaluation was based strictly on the MRI

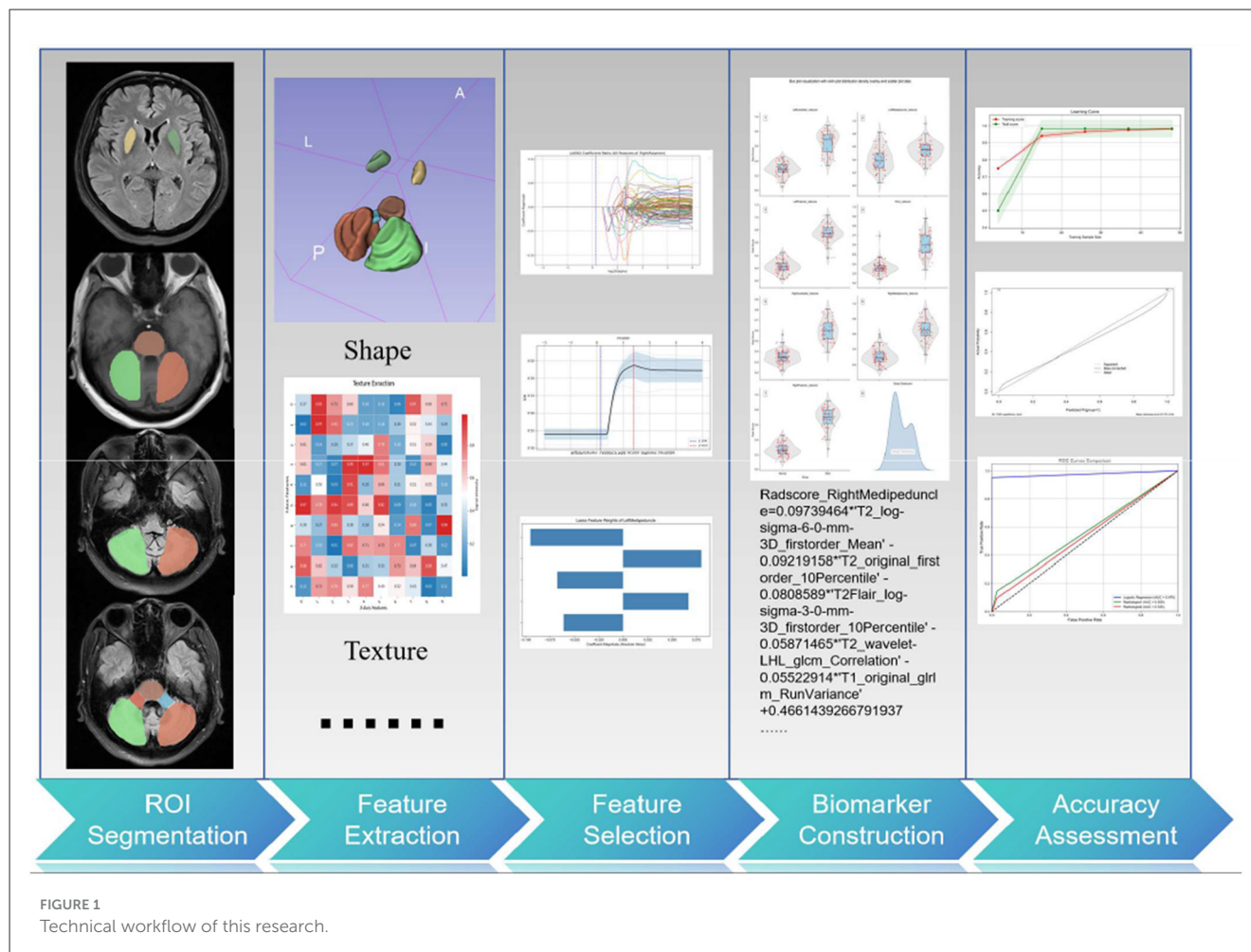


FIGURE 1  
Technical workflow of this research.

markers described in the 2022 International Movement Disorder Society (MDS) diagnostic criteria for MSA (6), without access to clinical information.

## Statistical analysis

Data analyses were performed using R software (v4.4.2) and Python (v3.9). Continuous variables conforming to a normal distribution were expressed as mean  $\pm$  standard deviation (SD) and compared between groups using the independent samples *t*-test. Non-normally distributed data were presented as median (interquartile range) [M (IQR)] and analyzed via the Mann-Whitney *U* test. Categorical variables were reported as frequency (percentage) with intergroup comparisons conducted using Chi-square tests.

Machine learning model performance was evaluated using: AUC, Class-specific accuracy, recall, F1-score. Statistical differences in AUC values between machine learning models and interpretations by two radiologists were assessed using DeLong's test. Model interpretability was analyzed via the SHAP package (v0.43.0) in Python to quantify feature contributions. Inter-observer agreement of the visual judgments of MRI images

between radiologists was evaluated using Cohen's kappa coefficient. A threshold of  $P < 0.05$  was defined for statistical significance.

## Results

### Demographic characteristics

A total of 62 patients with clinically probable MSA (mean age  $66.3 \pm 7.8$ , 35 females), 73 healthy controls (mean age  $67.6 \pm 10.5$ , 31 females), and the same 62 MSA patients were enrolled. Statistical analysis showed that there was no significant difference in age between groups (Mann-Whitney *U* test, two-tailed test,  $P > 0.05$ ), and there was no significant difference in gender distribution between groups (chi-square test, two-tailed test,  $P > 0.05$ ). Detailed data on demographic characteristics are provided in Table 1.

### Feature selection and construction of Rad-score

Robust features demonstrating intraclass correlation coefficients (ICC)  $\geq 0.75$  were selected from multimodal composite features within each brain region. These features were subsequently subjected to least absolute shrinkage and



TABLE 1 Demographic characteristics of study participants.

| Characteristic       | Normal group<br>( <i>n</i> = 73) | MSA group<br>( <i>n</i> = 62) | <i>P</i> -value    |
|----------------------|----------------------------------|-------------------------------|--------------------|
| Gender, <i>n</i> (%) |                                  |                               |                    |
| Female               | 38 (52.1%)                       | 24 (38.7%)                    | 0.168 <sup>a</sup> |
| Male                 | 35 (47.9%)                       | 38 (61.3%)                    |                    |
| Age (years)          |                                  |                               |                    |
| Median [Min, Max]    | 65.0 [38.0, 88.0]                | 67.0 [47.0, 83.0]             | 0.575 <sup>b</sup> |
| IQR [Q1, Q3]         | 23.0 [54.0, 77.0]                | 15.0 [55.0, 70.0]             |                    |

<sup>a</sup>Gender differences were analyzed using the Chi-square test; <sup>b</sup>age data violated the normality assumption (Shapiro–Wilk test: *P* = 0.0291 for Normal, *P* = 0.0304 for MSA group); therefore, intergroup age comparisons were performed with the Mann–Whitney *U* test.

selection operator (LASSO) regression analysis with 10-fold cross-validation. The five features exhibiting the strongest predictive weights (Supplementary Figure 1) were retained to construct the radiomics biomarker (Rad-score) using the formulas in Supplementary Table 2.

The radiomics signature (Rad-score) for each brain region was calculated using the aforementioned formula. Composite distribution plots of Rad-scores were subsequently generated (Figure 2). Intergroup differences were observed between the MSA cohort and healthy controls, indicating distinct distribution patterns.

Composite plot showing Rad-score distributions in specific brain regions. Figures 2.1–2.7 represent the left cerebellar, left medipeduncle, left putamen, pons, right cerebellar, right medipeduncle, and right putamen, respectively. Figure 2.8 displays the overall Rad-score distribution, with bimodal peaks indicating distinct mean values between groups.

### The LR model

The LR model was identified as the optimal predictive model and underwent further evaluation. The logistic equation is as follows:

$$\begin{aligned} \log(P/(1 - P)) = & (1.4005 * \text{LeftCerebellar\_RADscore}) \\ & + (0.5226 * \text{LeftMedipeduncle\_RADscore}) \\ & + (2.0314 * \text{LeftPutamen\_RADscore}) \\ & + (0.7332 * \text{Pons\_RADscore}) \\ & + (0.9171 * \text{RightCerebellar\_RADscore}) \\ & + (1.4922 * \text{RightMedipeduncle\_RADscore}) \\ & + (1.6966 * \text{RightPutamen\_RADscore}) \\ & - 4.2657 \end{aligned}$$

As evidenced by the learning curve derived from the training cohort (Figure 3), both training and test scores of the logistic regression (LR) model converged asymptotically toward 0.98. This convergence indicates the absence of overfitting and confirms robust generalization capabilities.

Performance metrics for the LR model across training, test, and validation sets are summarized in Table 2. The model's

discriminative power and generalization characteristics for the two sample classes were comprehensively evaluated using four core metrics: Precision, Recall, F1-Score, and Support. All datasets exhibited high classification performance (Macro Avg F1 ≥ 0.95), establishing model robustness. The near-identical accuracies of the training set (Accuracy = 0.98) and test set (Accuracy = 0.97) further substantiate the absence of overfitting. In the validation set, moderately reduced recall (0.89) was observed for multiple system atrophy (MSA) samples relative to other datasets. Conversely, normal group samples achieved perfect recall (1.00) universally, demonstrating complete capture of this class. Notably, MSA classification consistently yielded precision of 1.00, indicating zero false positives.

### SHAP-based model interpretability analysis

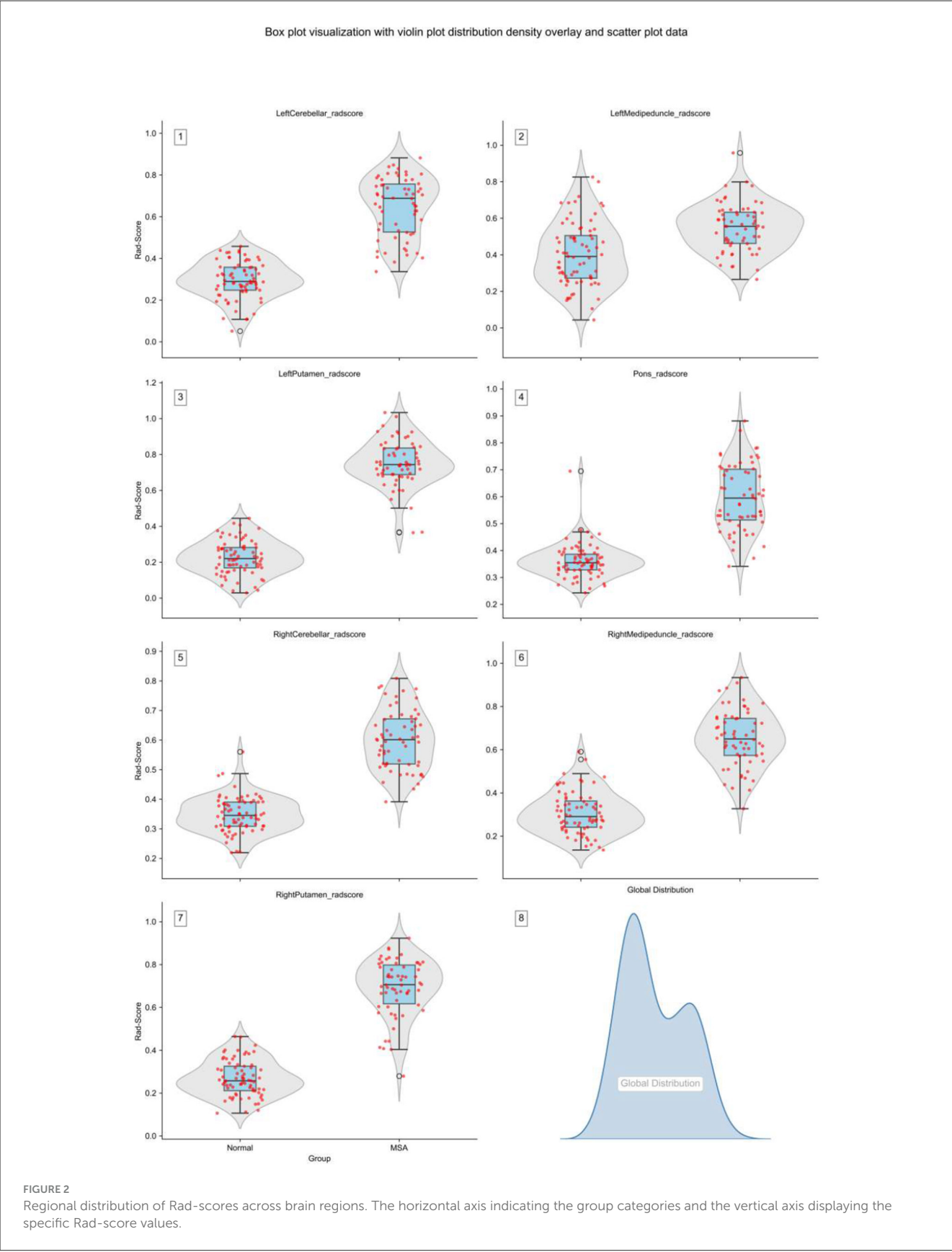
SHAP analysis was performed to interpret the contribution of regional radiomics signatures (RADscore) and the model's decision-making mechanism. Figure 4A illustrates the hierarchical feature importance in the prediction model, where the vertical axis ranks features by descending importance and the horizontal axis denotes the mean absolute SHAP value. The analysis identified the left putamen rad-score as the most influential predictor. Figure 4B provides a detailed summary plot of this ranking; each point represents an individual sample, with a color gradient (blue to red) indicating low-to-high feature magnitudes. The vertical axis sorts features by importance, while the distribution illustrates correlations between feature values and their corresponding SHAP values. SHAP analysis revealed significant lateralized contributions of imaging biomarkers across these brain regions.

### Construction of nomogram

Based on the established logistic regression model, a nomogram (Figure 5) predicting the probability of multiple system atrophy (MSA) was constructed using the following predictors: rad-scores of the left cerebellar hemisphere, left medipeduncle, left putamen, pons, right cerebellar hemisphere, right medipeduncle, and right putamen.

### Visual assessment of radiologists

Two radiologists performed independent assessments on 135 cases blinded to clinical information. Radiologist A classified 127 cases as normal and 8 as multiple system atrophy (MSA), while Radiologist B classified 124 as normal and 11 as MSA. Consensus diagnoses identified 118 normal cases and 2 MSA cases (Figure 6). The Cohen's kappa coefficient for inter-rater agreement was 0.152. Receiver operating characteristic (ROC) curves for the logistic regression (LR) model and both radiologists are shown in Figure 7, with areas under the curve (AUC) of 0.559 (95% CI: 0.48–0.63) for Radiologist A and 0.535 (95% CI: 0.45–0.62) for Radiologist B. The DeLong test comparing diagnostic performance between



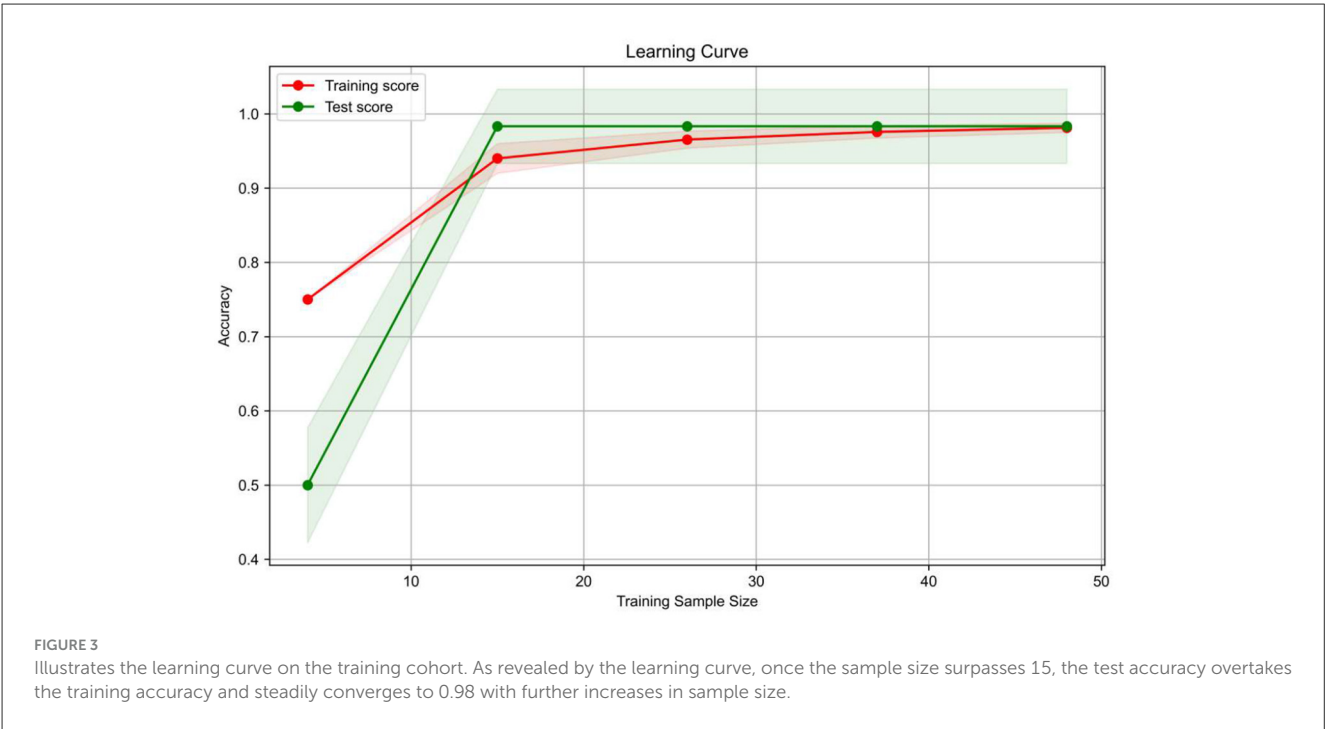


TABLE 2 Classification report of the logistic regression model across training, test, and validation cohorts.

| Dataset    | Class/statistic | Precision | Recall | F1-score | Support | Accuracy |
|------------|-----------------|-----------|--------|----------|---------|----------|
| Training   | Normal          | 0.96      | 1.00   | 0.98     | 27      | –        |
|            | MSA             | 1.00      | 0.96   | 0.98     | 27      | –        |
|            | Macro avg       | 0.98      | 0.98   | 0.98     | –       | –        |
|            | Weighted avg    | 0.98      | 0.98   | 0.98     | 54      | –        |
|            | Overall         | –         | –      | –        | –       | 0.98     |
| Testing    | Normal          | 0.96      | 1.00   | 0.98     | 24      | –        |
|            | MSA             | 1.00      | 0.94   | 0.97     | 16      | –        |
|            | Macro avg       | 0.98      | 0.97   | 0.97     | –       | –        |
|            | weighted avg    | 0.98      | 0.97   | 0.97     | 40      | –        |
|            | Overall         | –         | –      | –        | –       | 0.97     |
| Validation | Normal          | 0.92      | 1.00   | 0.96     | 22      | –        |
|            | MSA             | 1.00      | 0.89   | 0.94     | 19      | –        |
|            | Macro avg       | 0.96      | 0.95   | 0.95     | –       | –        |
|            | Weighted avg    | 0.96      | 0.95   | 0.95     | 41      | –        |
|            | Overall         | –         | –      | –        | –       | 0.95     |

Radiologist A and Radiologist B yielded no significant difference ( $Z = 0.803$ ,  $P = 0.422$ ).

### Discussion

Multiple system atrophy (MSA) is a rare disease; because samples are hard to collect, studies are often characterized by small sample sizes and high-dimensional feature spaces, which can easily

lead to overfitting if not handled properly. In this study, 70% of the samples were randomly selected for feature extraction of brain regions, aiming to reduce the excessive dependence of the model on the training data, thereby reducing the risk of overfitting. The core logic of this approach, a subsampling approach, is to use a random masking mechanism similar to Dropout to form regularization to improve generalization through data-level randomness (25–27). Previous studies have shown the utility of feature screening in small, high-dimensional data such as the one used in our study (28).

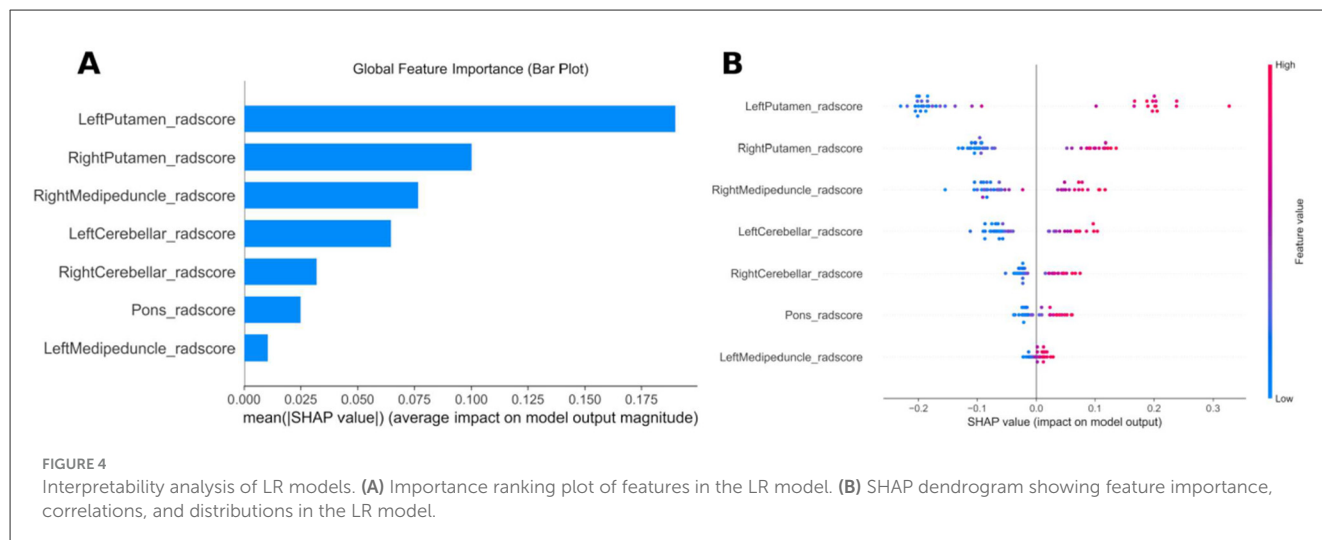


FIGURE 4 Interpretability analysis of LR models. **(A)** Importance ranking plot of features in the LR model. **(B)** SHAP dendrogram showing feature importance, correlations, and distributions in the LR model.

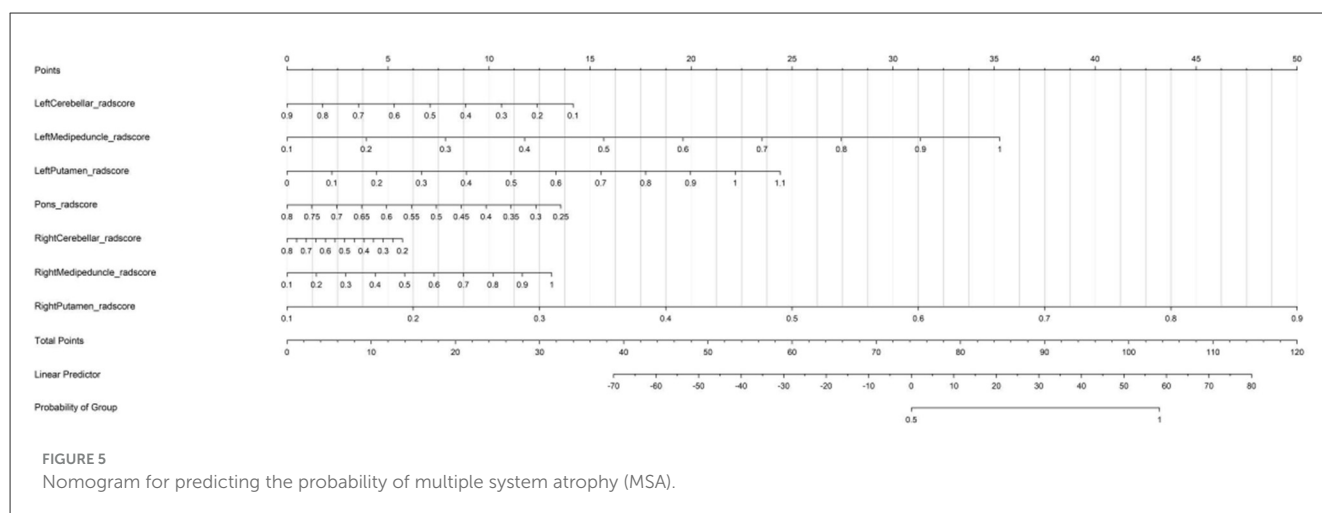


FIGURE 5 Nomogram for predicting the probability of multiple system atrophy (MSA).

To efficiently identify the most discriminative features from extensive feature pools, we performed Z-score normalization on intraclass correlation coefficient (ICC)-validated features per brain region (29), then integrated four sequences (T1, T2, T2 FLAIR, ADC) into multimodal representations (30). Subsequently, LASSO regression was applied to extract key features from each region's multimodal set (31). Least absolute shrinkage and selection operator (LASSO) regression is a linear regression method combining feature selection and regularization. The core of LASSO regression is to realize sparse modeling by introducing L1 regularization term. Lasso regression can solve the problem of high-dimensional data redundancy by compressing the coefficients of unimportant features to zero and automatically screening key variables. Only a few nonzero coefficients are retained in the generated model, which improves the interpretation of the model. L1 regularization can deal with multicollinearity problems more effectively than ridge regression (L2 regularization) (32). Due to these characteristics, Lasso regression has been widely used in radiomics. Radiomics features are often in thousands of dimensions, and Lasso can simplify model parameters by filtering out 99% redundant features from the original features

(33). Features were selected using internal 10-fold cross validation in the training set by the minimum mean squared error (MSE) (22). Among the non-zero weight features obtained from lasso regression, the five features with the greatest weight influence were selected as the variables to calculate the Rad-score. Based on the weight of feature variables and regression intercept construct of LASSO regression, Rad-score construction formulas (see [Supplementary material](#)) for seven brain regions were established as biomarker (34). The combined plot shows that the rad-score of each brain region has some discrimination power.

In order to maximize the model performance, a total of seven RAD-scores from seven brain regions in each sample were combined into a new research sample for logistic regression modeling. In order to improve the generalization ability, the total 135 samples were randomly divided into training group, test group and validation group according to 6:2:2 (35). The training group was used for modeling, and the learning curve within that group was plotted. The learning curve began to converge when the training sample reached 15, and the test score and training score increased with the training sample, and tended to converge to a curve pattern with the same value, indicating that the model did

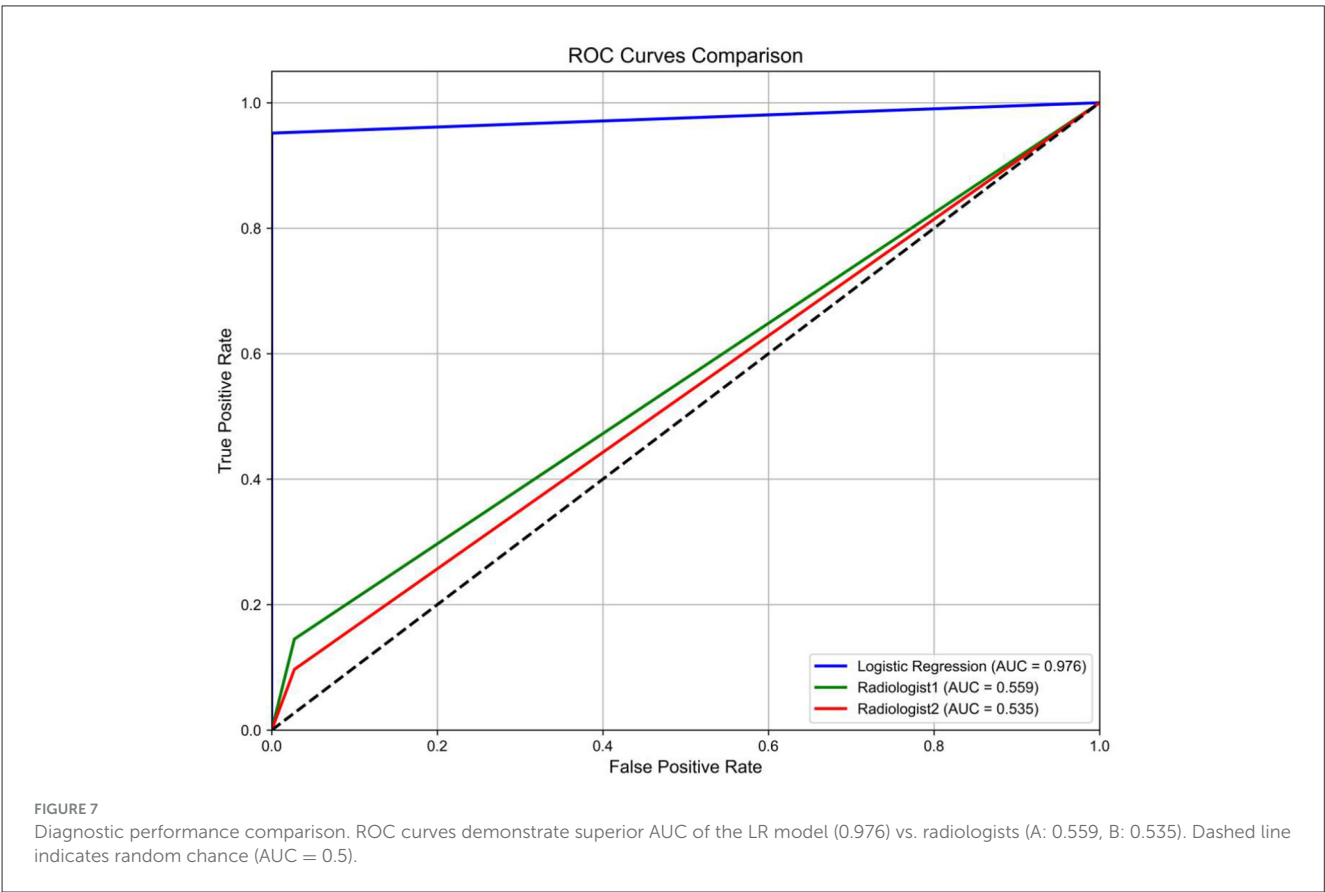
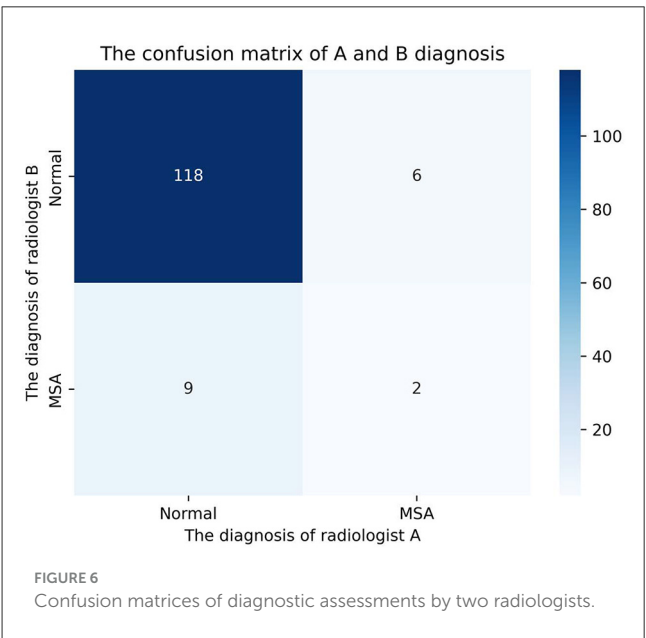


not overfit, showing that the model had good generalization ability (36). Among the key indicators, only MSA in the validation group achieved a recall rate of 0.89, while the others achieved an accuracy rate and recall rate of more than 0.9, showing excellent classification functions on the validation set and the test cohorts (37). This

0.89 sensitivity highlights the model’s potential as a screening tool for MSA.

In this study, we demonstrated that the radiomics-based Rad-score exhibited greater sensitivity than conventional MRI imaging markers currently incorporated in diagnostic criteria for multiple system atrophy (MSA). The seven brain regions delineated in this study are the MRI imaging biomarkers mentioned in the diagnostic criteria of MSA, and can serve as on basis for clinical diagnosis of MSA (6). Although MRI abnormalities in MSA patients have high specificity, their sensitivity is usually low. Moreover, the clinical utility of these MRI findings in improving diagnostic accuracy remains to be fully elucidated (38). All cases included in this study were diagnosed as clinically probable multiple system atrophy (MSA) due to the absence of characteristic MRI findings. For further validation, 135 samples were evaluated by two radiologists with more than 10 years of experience. The AUCs for Radiologist A and Radiologist B were 0.559 and 0.535, and the kappa coefficient of agreement between them was 0.152. These results demonstrate that macroscopic MRI features alone were insufficient for accurate diagnosis in this study. In contrast, the diagnostic model based on the Rad-score derived from seven brain regions showed excellent classification performance, supporting its practical utility.

The weights of the RADscores of these seven brain regions in the model found in this study are also consistent with the laterality characteristics of MSA found by other research methods. The SNAP diagram of the Logistic regression model shows that among the seven brain regions, the influence weights are ranked as follows: left putamen > right putamen > right medipeduncle



> left cerebellar > right cerebellar > pons > left medipeduncle. The results showed that the influence of the left putamen was greater than that of the right medipeduncle, and the influence of the left cerebellar hemisphere was greater than that of the right cerebellar hemisphere. These findings may be related to the pathogenesis of MSA. There is also an important tendency of hemisphere lateralization in the process of PD. Therefore, PD is considered an inherently asymmetric disease in clinical practice. This clinical asymmetry is associated with more severe contralateral nigrostriatal degeneration (39). Some studies have shown a “left hemisphere susceptibility” in this condition, as the left nigrostriatal pathway is more affected than the right (40). Previous PET imaging studies based on altered  $^{18}\text{F}$ -DOPA uptake have confirmed that the loss of  $^{18}\text{F}$ -DOPA uptake rate in the nigrostriatal system in selected populations of drug-naïve Parkinson’s disease cohorts is predominantly on the most affected side, so that the left hemisphere image depicts the more affected side. While the less affected side (LAS) corresponds to the right hemisphere, the reduced topography was mainly in the putamen of the left hemisphere with maximum uptake loss in the anterior-posterior axis and dorsoventral axis, respectively (41). In the study by Van Laere and colleagues, left putamen uptake was observed in 24 of 38 patients (63.1%) with right-sided predominant disease ( $P < 0.001$ ), indicating that this laterality is also present in IPD such as MSA (42). The dopaminergic system is thought to be primarily responsible for this lateralization due to its critical role in motor control. Inherent interhemispheric imbalances in nigrostriatal dopamine (DA) levels in humans and animals have been shown to be associated with lateralization of motor behavior (43). This change can cause the corresponding changes in the images of the putamen. Although such changes cannot be detected in the macroscopic image features, the RADscore constructed by radiomics can accurately detect the changes in the left and right putamen. In the MSA group in the present study, the changes of the left putamen were greater than those of the right putamen, which is consistent with previous studies.

Minori Furuta et al. found that MSA patient exhibited laterality changes in the middle cerebellar peduncle on SPECT (44). While conventional MRI failed to reveal these alterations, radiomics captured them and confirmed laterality patterns reported previously. Similarly, Francesca Caso observed atrophy of the left cerebellar hemisphere but not the right cerebellar hemisphere in patients with MSA-P by 1.5T magnetic resonance imaging, suggesting that atrophy of the left cerebellar hemisphere may be more easily observed at the macroscopic level than that of the right (45). This study found that the effect of the left cerebellar hemisphere is also called right hemisphere enlargement, which is consistent with this. These results are consistent with the laterality of previous studies, and further support the proposed RADscore as a biomarker to not only preferentially screen out highly suspected MSA cases. In addition, Eun Hye Jeong et al. found through  $^{123}\text{I}$ -FP-CIT SPECT study that the asymmetry of putamen was more obvious in the early stage of the disease, and this asymmetry decreased with the extension of follow-up time (46). The patients in this study belonged to the early stage of the disease when none of the macroscopic imaging markers required by the guidelines were found, so the RADscore difference of the putamen was more significant. The Rad-score may serve as a potential biomarker

for the early diagnosis of multiple system atrophy (MSA). The diagnostic model based on the Rad-score demonstrates promising diagnostic performance in identifying MSA cases.

## Conclusion

In conclusion, for patients with clinically suspected multiple system atrophy (MSA) but lacking definitive MRI markers, the radiomics-based RAD score offers a sensitive imaging biomarker that enables the construction of a diagnostic model capable of distinguishing MSA from healthy controls and improving overall diagnostic accuracy.

## Limitations

This study has several limitations. First, it was a single-center retrospective analysis, which may limit the generalizability of the findings to broader or more diverse populations. Second, although we included patients with clinically probable MSA and healthy controls, the diagnosis was primarily based on clinical criteria, which may introduce selection bias. Third, the radiomics model was built using manually delineated regions of interest (ROIs), and thus may be subject to inter- and intra-observer variability; future studies incorporating automated segmentation techniques are warranted. Finally, external validation using an independent cohort is needed to further confirm the robustness and clinical applicability of the RAD score as a diagnostic biomarker.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding authors.

## Ethics statement

The studies involving humans were approved by the Institutional Review Board of the Second People’s Hospital of Jiulongpo District. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants’ legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

ZL: Funding acquisition, Conceptualization, Software, Investigation, Visualization, Writing – review & editing, Resources, Writing – original draft, Project administration, Validation, Supervision, Formal analysis, Data curation, Methodology. WZ: Project administration, Writing – original draft, Software, Resources, Visualization, Data curation, Methodology, Investigation, Writing – review & editing, Conceptualization,

Funding acquisition, Validation, Supervision, Formal analysis. RY: Investigation, Writing – review & editing, Data curation. DC: Data curation, Investigation, Writing – review & editing. XL: Writing – review & editing, Investigation, Data curation. KW: Investigation, Data curation, Writing – review & editing. LC: Investigation, Writing – review & editing, Data curation. HY: Conceptualization, Funding acquisition, Supervision, Resources, Writing – review & editing. YD: Writing – review & editing, Funding acquisition, Resources.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by grants from the Science and Health Joint Medicine Research Project of Chongqing (including traditional Chinese medicine) (grant number 2024MSXM141) and the Medical Research Project of the Chongqing Municipal Health Commission (Project No. 2024WSJK103).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Goh YY, Saunders E, Pavey S, Rushton E, Quinn N, Houlden H, et al. Multiple system atrophy. *Pract Neurol*. (2023) 23:208–21. doi: 10.1136/pn-2020-002797
- Stankovic I, Fanciulli A, Sidoroff V, Wenning GK. A review on the clinical diagnosis of multiple system atrophy. *Cerebellum*. (2023) 22:825–39. doi: 10.1007/s12311-022-01453-w
- Krismer F, Fanciulli A, Meissner WG, Coon EA, Wenning GK. Multiple system atrophy: advances in pathophysiology, diagnosis, and treatment. *Lancet Neurol*. (2024) 23:1252–66. doi: 10.1016/S1474-4422(24)00396-X
- Zhang L, Hou Y, Cao B, Wei Q, Ou R, Liu K, et al. Longitudinal evolution of motor and non-motor symptoms in early-stage multiple system atrophy: a 2-year prospective cohort study. *BMC Med*. (2022) 20:446. doi: 10.1186/s12916-022-02645-1
- Watanabe H, Nagao R, Mizutani Y, Ito M. [limitations of the second consensus statement on the diagnosis of multiple system atrophy]. *Brain Nerve Shinkei Kenkyu No Shinpo*. (2023) 75:101–8. doi: 10.11477/mf.1416202289
- Wenning GK, Stankovic I, Vignatelli L, Fanciulli A, Calandra-Buonaura G, Seppi K, et al. The movement disorder society criteria for the diagnosis of multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. (2022) 37:1131–48. doi: 10.1002/mds.29005
- Aludin S, Schmill LPA. MRI signs of parkinson's disease and atypical parkinsonism. *ROFO Fortschr Geb Rontgenstr Nuklearmed*. (2021) 193:1403–10. doi: 10.1055/a-1460-8795
- van Eimeren T. Central autonomic dysfunction in multiple system atrophy: can we measure it with MRI? *Clin Auton Res Off J Clin Auton Res Soc*. (2020) 30:185–7. doi: 10.1007/s10286-020-00695-0
- Focke NK, Helms G, Pantel PM, Scheewe S, Knauth M, Bachmann CG, et al. Differentiation of typical and atypical parkinson syndromes by quantitative MR imaging. *AJNR Am J Neuroradiol*. (2011) 32:2087–92. doi: 10.3174/ajnr.A2865
- Kim M, Ahn JH, Cho Y, Kim JS, Youn J, Cho JW. Differential value of brain magnetic resonance imaging in multiple system atrophy cerebellar phenotype and spinocerebellar ataxias. *Sci Rep*. (2019) 9:17329. doi: 10.1038/s41598-019-53980-y
- Zhu S, Deng B, Huang Z, Chang Z, Li H, Liu H, et al. “Hot cross bun” is a potential magnetic marker for the severity of cerebellar ataxia in MSA-C. *NPJ Park Dis*. (2021) 7:15. doi: 10.1038/s41531-021-00159-w

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2025.1650350/full#supplementary-material>

- Pellecchia MT, Stankovic I, Fanciulli A, Krismer F, Meissner WG, Palma JA, et al. Can autonomic testing and imaging contribute to the early diagnosis of multiple system atrophy? A systematic review and recommendations by the movement disorder society multiple system atrophy study group. *Mov Disord Clin Pract*. (2020) 7:750–62. doi: 10.1002/mdc3.13052
- Zhao Y, Wu P, Wu J, Brendel M, Lu J, Ge J, et al. Decoding the dopamine transporter imaging for the differential diagnosis of parkinsonism using deep learning. *Eur J Nucl Med Mol Imaging*. (2022) 49:2798–811. doi: 10.1007/s00259-022-05804-x
- Villena-Salinas J, Ortega-Lozano SJ, Amrani-Raissouni T, Agüera E, Caballero-Villarraso J. Follow-up findings in multiple system atrophy from [<sup>123</sup>I]ioflupane single-photon emission computed tomography (SPECT): a prospective study. *Biomedicines*. (2023) 11:2893. doi: 10.3390/biomedicines11112893
- Mei YL, Yang J, Wu ZR, Yang Y, Xu YM. Transcranial sonography of the substantia nigra for the differential diagnosis of parkinson's disease and other movement disorders: a meta-analysis. *Park Dis*. (2021) 2021:8891874. doi: 10.1155/2021/8891874
- Parkinson's Disease and Movement Disorders Group, Neurology Branch of Chinese Medical Association, Parkinson's Disease and Movement Disorders Group. Expert consensus on diagnostic criteria for multiple system atrophy in China (2022). *Chin J Neurol*. (2023) 56:15–29.
- National Health Commission of the People's Republic of China, Ministry of Education of the People's Republic of China, Ministry of Science and Technology of the People's Republic of China, State Administration of Traditional Chinese Medicine. *Circular on the Issuance of Ethical Review Measures for Life Science and Medical Research Involving Human Beings*. (2023). Available online at: [https://www.gov.cn/zhengce/zhengceku/2023-02/28/content\\_5743658.htm](https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm) (Accessed August 3, 2025).
- Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*. (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001
- Li C, Wang H, Chen Y, Fang M, Zhu C, Gao Y, et al. A nomogram combining MRI multisequence radiomics and clinical factors for predicting recurrence of high-grade serous ovarian carcinoma. *J Oncol*. (2022) 2022:1716268. doi: 10.1155/2022/1716268
- Huo X, Wang Y, Ma S, Zhu S, Wang K, Ji Q, et al. Multimodal MRI-based radiomic nomogram for predicting telomerase reverse transcriptase promoter

- mutation in IDH-wildtype histological lower-grade gliomas. *Medicine*. (2023) 102:11. doi: 10.1097/MD.00000000000036581
21. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Lck S. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2020) 295:191145. doi: 10.1148/radiol.2020191145
  22. Yossofzai O, Fallah A, Maniquis C, Wang S, Ragheb J, Weil AG, et al. Development and validation of machine learning models for prediction of seizure outcome after pediatric epilepsy surgery. *Epilepsia*. (2022) 63:1956–69. doi: 10.1111/epi.17320
  23. Bernard D, Doumard E, Ader I, Kemoun P, Pagès JC, Galinier A, et al. Explainable machine learning framework to predict personalized physiological aging. *Aging Cell*. (2023) 22:e13872. doi: 10.1111/accel.13872
  24. Li J, Liu S, Hu Y, Zhu L, Mao Y, Liu J. Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: retrospective cohort study. *J Med Internet Res*. (2022) 24:e38082. doi: 10.2196/38082
  25. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. (2014) 15:1929–58.
  26. Kim BJ, Kim SW. *Stochastic Subsampling with Average Pooling*. (2024). Available online at: <https://xueshu.baidu.com/usercenter/paper/show?paperid=1y3h0gg0v9010jb078380e00be037530> (Accessed May 27, 2025).
  27. Zhang Z, Xu ZQJ. Implicit regularization of dropout. *IEEE Trans Pattern Anal Mach Intell*. (2024) 46:4206–17. doi: 10.1109/TPAMI.2024.3357172
  28. Haftorn KL, Romanowska J, Lee Y, Page CM, Magnus PM, Håberg SE, et al. Stability selection enhances feature selection and enables accurate prediction of gestational age using only five DNA methylation sites. *Clin Epigenetics*. (2023) 15:114. doi: 10.1186/s13148-023-01528-3
  29. *Standardize Data Using Z-Score/Standard Scalar | Python*. Available online at: <https://www.hackersrealm.net/post/standardize-data-using-standard-scalar> (Accessed May 27, 2025).
  30. Zhang YF, Zhou C, Guo S, Wang C, Yang J, Yang ZJ, et al. Deep learning algorithm-based multimodal MRI radiomics and pathomics data improve prediction of bone metastases in primary prostate cancer. *J Cancer Res Clin Oncol*. (2024) 150:78. doi: 10.1007/s00432-023-05574-5
  31. Xi LJ, Guo ZY, Yang XK, Ping ZG. [Application of LASSO and its extended method in variable selection of regression analysis]. *Zhonghua Yu Fang Yi Xue Za Zhi*. (2023) 57:107–11. doi: 10.3760/cma.j.cn112150-20220117-00063
  32. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Ser B Stat Methodol*. (2011) 73:267–88. doi: 10.1111/j.1467-9868.2011.00771.x
  33. Du P, Liu X, Wu X, Chen J, Cao A, Geng D. Predicting histopathological grading of adult gliomas based on preoperative conventional multimodal MRI radiomics: a machine learning model. *Brain Sci*. (2023) 13:912. doi: 10.3390/brainsci13060912
  34. Du L, Yuan Q, Han Q. A new biomarker combining multimodal MRI radiomics and clinical indicators for differentiating inverted papilloma from nasal polyp invaded the olfactory nerve possibly. *Front Neurol*. (2023) 14:1151455. doi: 10.3389/fneur.2023.1151455
  35. Feng Z, Li H, Liu Q, Duan J, Zhou W, Yu X, et al. CT radiomics to predict macrotrabecular-massive subtype and immune status in hepatocellular carcinoma. *Radiology*. (2023) 307:e221291. doi: 10.1148/radiol.221291
  36. *A Deep Dive into Learning Curves in Machine Learning | ML-Articles – Weights & Biases*. Available online at: <https://wandb.ai/mostafaibrahim17/ml-articles/reports/A-Deep-Dive-Into-Learning-Curves-in-Machine-Learning--Vmlldzo0NjA1ODY0> (Accessed May 27, 2025).
  37. Rainio O, Teuvo J, Klén R. Evaluation metrics and statistical tests for machine learning. *Sci Rep*. 14:1–14. doi: 10.1038/s41598-024-56706-x
  38. Kim HJ, Jeon B, Fung VSC. Role of magnetic resonance imaging in the diagnosis of multiple system atrophy. *Mov Disord Clin Pract*. (2017) 4:12–20. doi: 10.1002/mdc3.12404
  39. Holmes AA, Matarazzo M, Mondesire-Crump I, Katz E, Mahajan R, Arroyo-Gallego T. Exploring asymmetric fine motor impairment trends in early parkinson's disease via keystroke typing. *Mov Disord Clin Pract*. (2023) 10:1530–5. doi: 10.1002/mdc3.13864
  40. Ortelli P, Ferrazzoli D, Zarucchi M, Maestri R, Frazzitta G. Asymmetric dopaminergic degeneration and attentional resources in parkinson's disease. *Front Neurosci*. (2018) 12:972. doi: 10.3389/fnins.2018.00972
  41. Pineda-Pardo JA, Sánchez-Ferro Á, Monje MHG, Pavese N, Obeso JA. Onset pattern of nigrostriatal denervation in early parkinson's disease. *Brain J Neurol*. (2022) 145:1018–28. doi: 10.1093/brain/awab378
  42. Kathuria H, Mehta S, Ahuja CK, Chakravarty K, Ray S, Mittal BR, et al. Utility of imaging of nigrosome-1 on 3T MRI and its comparison with 18F-DOPA PET in the diagnosis of idiopathic parkinson disease and atypical parkinsonism. *Mov Disord Clin Pract*. (2020) 8:224–30. doi: 10.1002/mdc3.13091
  43. *Hemispheric Differences in the Mesostriatal Dopaminergic System - PubMed*. Available online at: <https://pubmed.ncbi.nlm.nih.gov/24966817/> (Accessed May 22, 2025).
  44. Furuta M, Sato M, Tsukagoshi S, Tsushima Y, Ikeda Y. Criteria-unfulfilled multiple system atrophy at an initial stage exhibits laterality of middle cerebellar peduncles. *J Neurol Sci*. (2022) 438:120281. doi: 10.1016/j.jns.2022.120281
  45. Caso F, Canu E, Lukic MJ, Petrovic IN, Fontana A, Nikolic I, et al. Cognitive impairment and structural brain damage in multiple system atrophy-parkinsonian variant. *J Neurol*. (2020) 267:87–94. doi: 10.1007/s00415-019-09555-y
  46. Jeong EH, Sunwoo MK, Lee JY, Han SK, Hyung SW, Song YS. Serial changes of I-123 FP-CIT SPECT binding asymmetry in parkinson's disease: analysis of the PPMI data. *Front Neurol*. (2022) 13:976101. doi: 10.3389/fneur.2022.976101