Tract-based Quantitative MRI for Resolving the Clinico-Radiological Paradox in Multiple Sclerosis

Osama Abdullah[1,2], Puti Wen*[2,3], Abdelmalek W. Abdelrazeq*[,2], Mariya Matrosova*[5], Lev Brylev[4], Amar Ahmad[6], Vasiliy Bryukhov[5], David Melcher[2,3], Bas Rokers[2,3]*

Author affiliations:

1 Core Technology Platforms, New York University Abu Dhabi, Abu Dhabi, UAE

2 Center for Brain and Health, New York University Abu Dhabi, Abu Dhabi, UAE

3 Department of Psychology, New York University Abu Dhabi, Abu Dhabi, UAE

4 Yas Clinic Khalifa City, Abu Dhabi, UAE

5 Russian Center of Neurology and Neurosciences, Moscow, Russia

6 Public Health Research Center, New York University Abu Dhabi, Abu Dhabi, UAE

* denotes equal contribution

Abstract

The clinico-radiological paradox in multiple sclerosis (MS) arises because conventional MRI measures, particularly total lesion volume, fail to fully capture the true burden of disability. These broad volumetric measures overlook the dual influence of where lesions occur and what is their microstructural composition. This study aimed to improve the clinical relevance of lesion analysis in MS by refining both how and where lesions are measured. While traditional approaches emphasize anatomical lesion burden, studies of structural connectivity in neurological disease have shown that functional deficits often arise from tract-specific interruptions within distributed brain networks. We hypothesized that filtering lesions through clusters of functionally meaningful white matter tracts, rather than broad anatomical compartments, would enable more accurate identification of clinically relevant damage. To test our hypothesis, we studied 132 participants, including 89 patients with MS (49 relapsing–remitting, 17 primary progressive, 23 secondary progressive; 62 women and 27 men; mean age 41 years, range 18–68) and 43 healthy controls (28 women and 15 men; mean age 36 years, range 23–58). All underwent standardized 3 Tesla MRI

including fluid-attenuated inversion recovery (FLAIR), T1 mapping, magnetization transfer ratio (MTR), and diffusion tensor imaging (DTI) with fractional anisotropy (FA) and mean diffusivity (MD).

We evaluated associations between imaging measures and disability, and assessed predictive performance with ridge-penalized regression across Expanded Disability Status Scale (EDSS), and motor scores Timed 25-Foot Walk (T25FW), Nine-Hole Peg Test (9HPT), as well as progression risk defined by the MSPro scale. Filtering lesions through functionally critical tracts revealed associations obscured by conventional approaches. Tract-based lesion T1 values correlated more strongly with disability than classical regional metrics (e.g., EDSS $r = 0.83$ in occipitoparietal tracts vs $r = 0.47$ in periventricular regions). Multimodal tract-specific metrics including T1, MTR, FA, and MD significantly predicted disability. In multivariate models, tract-based frameworks consistently outperformed classical region-based ones, achieving higher discrimination and better model fit for binarized EDSS and MSPro (AUC = 0.86–0.96 for tract-based vs 0.57–0.86 for classical anatomical regions). For continuous outcomes (T25FW, 9HPT), tract-based models explained substantially more variance ($R^2 = 0.19–0.43$ for tract-based vs 0.01–0.34 for classical regions). Including demographic covariates improved model fit in both approaches, yet tract-based quantitative metrics remained superior to classical region-based measures.

By integrating lesion location and microstructural composition, tract-based quantitative MRI enhances disability prediction and provides interpretable biomarkers to support individualized monitoring and therapeutic decision-making.

1. Introduction

Multiple sclerosis (MS) is a chronic, immune-mediated disease of the central nervous system characterized by inflammatory demyelination and neurodegeneration, most prominently in white matter, with post-mortem studies showing that lesions vary widely in demyelination, axonal loss, gliosis, and microglial activity depending on stage, location, and patient-specific factors[1–3]. This heterogeneity is mirrored clinically in the diverse disability profiles seen in patients, ranging from gait and motor impairment to visual dysfunction, cognitive decline, and sensory symptoms, raising the challenge of how lesion characteristics can be meaningfully linked to functional outcome[4–8]. Conventional MRI, particularly T2 and FLAIR, guides diagnosis and monitoring but depicts lesions as uniform bright spots in broad anatomical regions, so lesion counts and volumes often miss true disease burden. This shortcoming gives rise to the clinico-radiological paradox, where clinical disability and disease course often do not align with visible lesion load on routine anatomical MRI, underscoring the need for more advanced imaging approaches that can bridge this disconnect[9–11].

Quantitative MRI (qMRI) offers promising complementary measures that augment conventional MRI by probing tissue microstructure beyond visible lesions. Metrics such as magnetization transfer ratio (MTR) and T1 relaxation times are sensitive to myelin and axonal content[12,13]. MTR and T1 metrics have been used to quantify demyelination within lesions, successfully stratifying them by disease severity and stage[14,15]. A particularly compelling finding came from a longitudinal study showing that reductions in MTR could be detected in normally-appearing white matter (NAWM) up to four months before new lesions became visible on conventional MRI[16]. Diffusion tensor imaging (DTI) probes tissue microstructure by characterizing water diffusion. Scalar metrics such as reduced fractional anisotropy (FA) and increased mean (MD) or radial diffusivity (RD) indicate axonal disorganization and demyelination, even in normal-appearing white matter[17]. By reconstructing white matter tracts from diffusion imaging, studies have moved beyond gross anatomical compartments, providing a more precise mapping of disease impact and revealing stronger correlations with clinical disability than region-based approaches[18–27]. Together, these findings highlight the potential of qMRI and tract-based approaches to capture clinically meaningful pathology that conventional lesion measures overlook.

Parallel to imaging advances, clinical outcomes are now assessed with multiple domain-specific scales: the Expanded Disability Status Scale (EDSS, for motor deficit)[28], Symbol Digit Modalities Test (SDMT, for processing speed)[29], Timed 25-Foot Walk (T25FW, for gait)[30], Nine-Hole Peg Test (9HPT, for manual dexterity)[31], and MSPro scale (for progression risk)[32,33]. While these developments enrich both imaging and disability assessment, most previous studies relied on simple univariate correlations, often using a single clinical score. Moreover, previous studies frequently used gross anatomical compartments to explain disability (for example, correlating EDSS with MTR in periventricular lesions; $r = 0.316$, $p = 0.007$)[34], an approach that, although informative, lacks functional specificity and reinforces oversimplified interpretations of "global disability." This reliance on univariate and regionally crude analyses introduces confounds, observer bias, and noise, thereby perpetuating the paradox.

In this study, we propose a tract-aware, multimodal framework that integrates qMRI metrics with the functional anatomy of white matter tracts. By quantifying both lesion location (*where*) and microstructural composition (*what*), and correlating these measures with multiple disability scores across progressively structured statistical models, we aim to establish stronger and more specific structure-function relationships. We hypothesize that this comprehensive approach will better account for lesion heterogeneity and the multidimensional nature of disability, enabling improved patient stratification and more accurate distinction of MS phenotypes compared to conventional lesion volume measures.

2. Materials and Methods

2.1 Subject Selection, Clinical Assessment, and Data Management

A total of 132 participants (89 MS patients and 43 healthy controls) were recruited between June 2022 and June 2024 at the Research Center of Neurology in Moscow, Russia. Consent was obtained according to the Declaration of Helsinki and that it has been approved by the ethical committee.

MS patients were clinically classified to relapsing remitting (RRMS), primary progressive (PPMS), or secondary progressive (SPMS) subtypes based on established diagnostic criteria, with inclusion limited to individuals aged 18–70. Control subjects, aged 18–75, were included if

they had no cognitive or neurological disorders, no severe somatic illness, and no structural brain abnormalities on MRI beyond age-related nonspecific white matter hyperintensities. General exclusion criteria for all participants included MRI contraindications (e.g., pacemaker), pregnancy, recent corticosteroid therapy (within 30 days for MS patients), structural brain pathology, or corrupted MRI data. All subjects provided written informed consent prior to participation. Clinical disability was assessed using the EDSS, SDMT, T25FW, 9HPT (dominant and non-dominant hand), and MSPro. We focused our analysis on EDSS, MSPro, and 9HPT, which primarily capture motor and physical disability domains with overlapping functional substrates, whereas SDMT (reflecting cognitive processing speed and frontoparietal network integrity) was reserved for a separate investigation outside the scope of this work. Tables 1–2, Fig 1, and Fig 2 summarize the full demographic and clinical profiles.

Table 1 *Summary of demographic and clinical metrics across MS subtypes and healthy controls, reporting group-wise means and standard deviations. The table highlights differences in age, disease duration, disability scores, and cognitive performance between subtypes. DMT: Disease modifying treatment*

Parameter	Healthy Controls	RRMS	PPMS	SPMS
N (% Female)	43 (65%)	49 (76%)	17 (59%)	23 (65%)
Age	35.60 ± 10.83	35.65 ± 10.25	47.88 ± 7.01	48.87 ± 12.50
Disease duration (years)	-	9.17 ± 7.48	7.53 ± 5.26	17.22 ± 8.37
EDSS	-	1.81 ± 0.95	3.94 ± 1.47	5.17 ± 1.14
DMT: Yes, n (%)	-	5 (12%)	9 (52,9%)	6 (26,1%)
Therapy: 1st line	-	29 (65.9%)	2 (22.2%)	11 (64.7%)
Therapy: 2nd line	-	15 (34.1%)	7 (77.8%)	6 (35.3%)
T25FW (s)	5.63 ± 0.52	6.23 ± 1.01	9.76 ± 3.57	13.95 ± 9.94
9HPT-D (s)	18.47 ± 2.31	19.81 ± 2.65	24.12 ± 4.51	28.61 ± 9.37
9HPT-ND (s)	20.00 ± 2.87	21.56 ± 3.89	27.39 ± 5.58	35.39 ± 14.16
SDMT Total	61.88 ± 10.24	56.49 ± 11.03	46.88 ± 8.78	38.43 ± 8.07
SDMT Mean	61.11 ± 10.37	55.88 ± 11.34	46.29 ± 9.02	37.50 ± 8.14
MSPro	-	1.22 ± 0.59	2.53 ± 0.62	2.55 ± 0.74

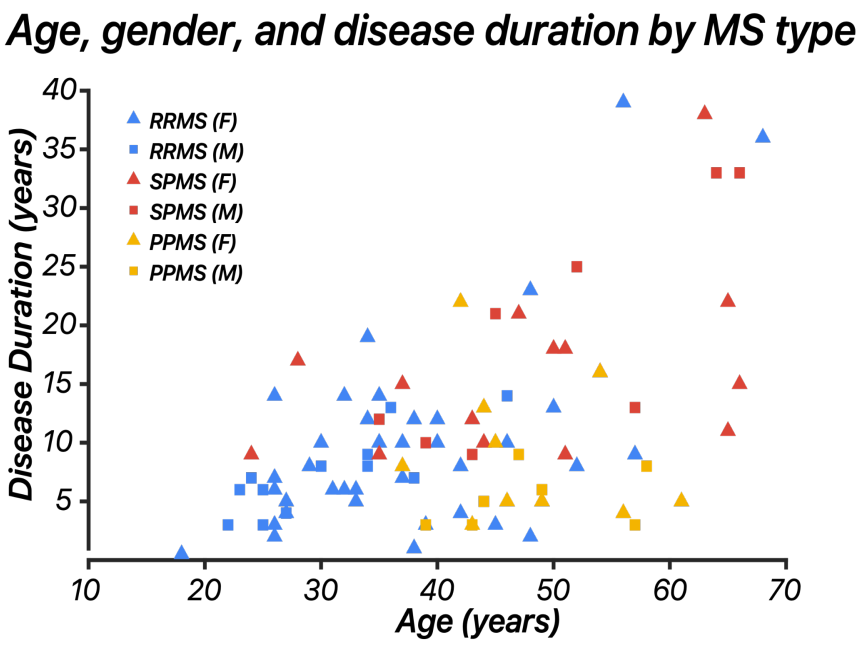

Figure 1 *Summarizes demographics (age, gender, and disease duration) across MS subtypes. For each subject age (x-axis), gender (male: square, female: triangle), disease duration in years (y-axis), and MS subtype (RRMS: blue, SPMS: red, PPMS: yellow) are represented.*

Table 2 *Comparative overview of the six disability assessment tools employed in this study: EDSS, SDMT, T25FW, 9HPT (dominant and non-dominant hands), and MSPro. Each scale is characterized along three key dimensions: (1) the functional domains it is designed to assess, (2) a concise description of the task(s) involved in administering the test, and (3) the specific scoring system used to quantify patient performance or impairment.*

Scale	Expanded Disability Status Scale (EDSS)	Symbol Digit Modalities Test (SDMT)	Timed 25-Foot Walk (T25FW)	Nine-Hole Peg Test (9HPT)	MSPRO
Functional Domains	**Pyramidal** (muscle weakness or difficulty moving limbs) **Cerebellar** (coordination) **Brainstem** (speech and eye movement) **Sensory** (numbness or loss of sensations) **Bowel** and bladder function **Visual** function **Cerebral** or mental function **Ambulation** (walking and use of aids)	**Visual scanning** and processing **Attention** **Working memory** **Psychomotor speed** **Visual-motor** coordination **Executive** functioning	**Ambulation** **Lower limb** strength **Balance** and motor control	**Fine motor** coordination **Finger dexterity** **Visual-motor** integration **Bilateral upper limb** function (dominant and non-dominant hands) **Independence** in activities of daily living (ADLs)	**Disease activity** (relapse frequency) **Symptoms**: Motor, visual, coordination/balance, speech, sensory, bladder/bowel **Cognitive** and fatigue-related functions **Mobility**, self-care, occupational functioning, hobbies, and social engagement
Type of Tasks	**Motor tasks**: Evaluating muscle strength and coordination **Sensory tests**: Assessing response to touch, pain, and vibration **Visual assessments**: Testing visual acuity and field **Cognitive evaluations**: Observing mental status and mood **Gait analysis**: Measuring walking distance and stability	In **writing** (written format), or **Orally** (oral format), depending on administration choice. The written form primarily assesses motor and visual processing, while the oral version helps distinguish whether deficits are motor-based or cognitive.	The individual is instructed to walk 25 feet as quickly and safely as possible. The time taken to complete the walk is recorded.	Pick up nine pegs one at a time and place them into nine holes on a pegboard as quickly as possible. Remove them one at a time and return them to the container. Each hand is tested separately. The task is completed under standardized conditions.	Symptom History (last 6 months) Impact Assessment impact severity on areas of functioning.
Scoring System	0.0–3.0: No to moderate disability but is fully ambulatory 3.5–5.5: Increasing difficulty walking but still ambulatory without aid for 300+ meters 6.0–7.5: Ambulation severely restricted; requires assistance (e.g., cane, walker, or wheelchair) 8.0–9.5: Essentially restricted to bed or chair; retains some self-care functions 10.0: Death due to MS	The raw score is the number of correct substitutions completed within 90 seconds. Standardized scores are not consistently available in the test manual; instead, performance is typically interpreted based on standard deviations (SD) from the mean	Healthy Controls: Median T25FW time is approximately 3.7 seconds (range: 2.8–5.2 seconds) 20% or greater increase in T25FW time is considered a clinically meaningful change, indicating potential disease progression	Normative Data: Norms vary based on age and sex. Cut-off values: >18 seconds is generally considered abnormal for the dominant hand. <0.5 pegs/s may indicate high risk for activity limitations. 33.3 seconds (corresponding to ~0.27 pegs/s) separates mild from more severe impairment	Section A: Relapse and Recovery (max 14 points) Section B: Symptoms (max 91 points) Section C: Impacts (max 14 points) Total Score = Sum of all above sections Interpretation Output: 1: (<46.3): Unlikely progression 2: (46.3–57.8): Possible progression 3: (>57.8): Likely progression

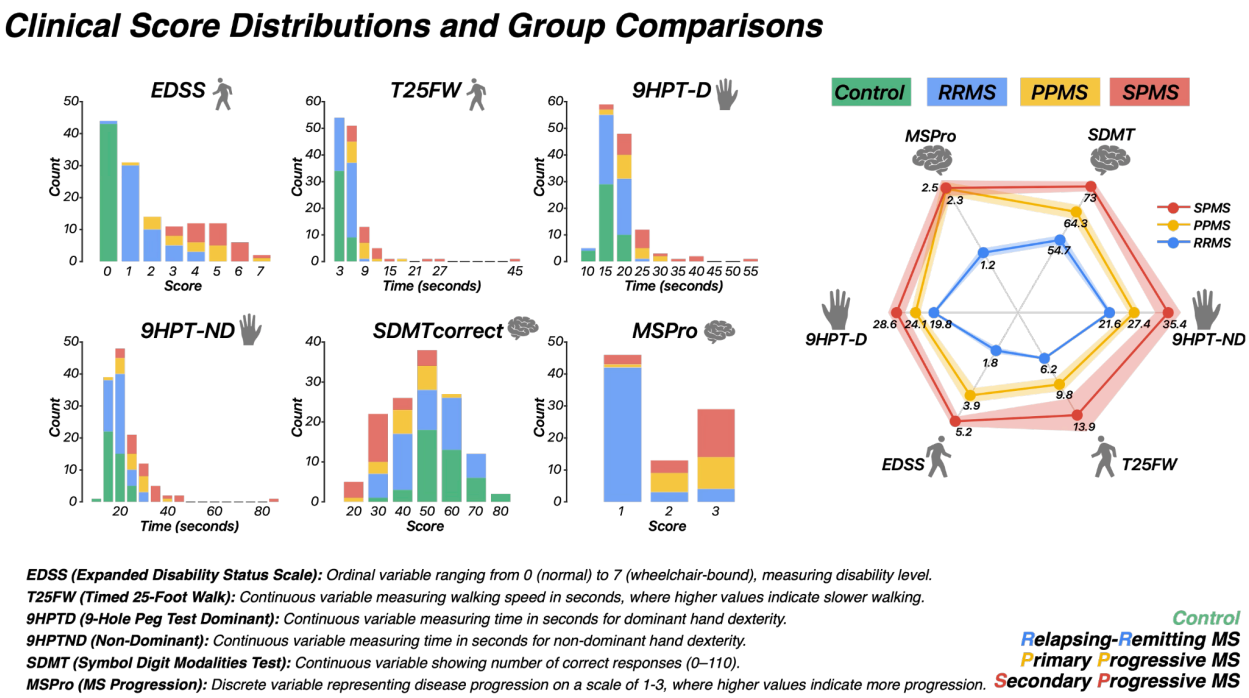

Figure 2 Represents the cohort's clinical score distribution subtypes comparisons. *The radar plot (right) displays the mean and standard error of the mean (SEM) for key clinical scores (EDSS, SDMT, 9HPTD, 9HPTND, T25FW, and MSPRO) across RRMS, SPMS, and PPMS groups. Clear separation is observed between subtypes, particularly between RRMS and SPMS, reflecting increasing disability with disease severity. Bar plots (left) represent each individual clinical disability score and their accompanying numerical scales. Controls were included when appropriate.*

2.2 MRI Acquisition

A multimodal 3T MRI protocol was acquired on a Siemens Magnetom Prisma (Syngo MR VE11) at the Research Center of Neurology in Moscow, using a 64-channel Head/Neck coil. FLAIR imaging was acquired using a 3D T2-weighted SPACE sequence with the following parameters: TR = 7000 ms, TE = 390 ms, TI = 2200 ms, flip angle = 120°, voxel size = 0.47 × 0.47 × 0.60 mm^3, matrix size = 512 × 512, FOV = 240 × 240 mm^2. T1-mapping was performed using a 3D MP2RAGE sequence with TR = 5000 ms, TE = 2.98 ms, TI1 = 700 ms, TI2 = 2500 ms, flip angles = 4° and 5° for the first and second inversions respectively, and isotropic voxel size = 1.00 × 1.00 × 1.00 mm^3, matrix size = 256 × 240, FOV = 256 × 240 mm^2. Magnetization transfer imaging was acquired using a 3D gradient-echo sequence with TR = 34 ms, TE = 6 ms, flip angle = 10°, voxel size = 1.00 × 1.00 × 3.00 mm^3, matrix size = 256 × 192, FOV = 256 × 192 mm^2. Both MT-on and MT-off images were collected for MTR calculation. Diffusion-weighted imaging was performed

using a single-shot echo-planar imaging sequence with TR = 5600 ms, TE = 82 ms, flip angle = 90°, matrix size = 122 × 122, multiband acceleration factor = 2. A multi-shell diffusion scheme was employed with b-values of 0, 1000, and 2500 s/mm², using 64 diffusion directions per shell. The total acquisition time was approximately one hour.

2.3 Image Preprocessing

Anatomical segmentation utilized the UNI image from the MP2RAGE acquisition after noise removal with the PreSurfer toolkit (https://github.com/srikash/presurfer). Cortical and subcortical structures were reconstructed using FreeSurfer (v7.4.1, recon-all)[35], and FLAIR images were registered to the structural space. Lesions were segmented with a custom-trained U-Net implemented in MONAI[36] and manually corrected in FSLeyes[37] by a physician (AWA), with verification by a trained neurologist (LB). Quantitative parameter mapping was then performed using qMRLab[38], T1 relaxation maps were estimated from the MP2RAGE sequence with the mp2rage model (TI1/TI2 = 700/2500 ms, flip angles = 4°/5°, TR = 5000 ms). Magnetization transfer ratio (MTR) maps were computed from MT-on and MT-off images using the formula: $MTR = [(S_0 - S_{mt})/S_0] \times 100\%$, where S_0 and S_{mt} represent signal without and with MT saturation, respectively. Diffusion preprocessing was carried out with TractoFlow (Nextflow pipeline) in the Scilus 1.5.0 Singularity container[39] including susceptibility distortion correction (TOPUP) and eddy current correction (EDDY) in FSL. DTI reconstruction yielded FA and MD maps. Local deterministic tractography was performed in DSI Studio with FA-based seeding (threshold = 0.1). Preprocessed DWI, b-values, and b-vectors were converted into diffusion source (.src) files, followed by generalized Q-sampling imaging (GQI; sampling length ratio = 1.25) aligned to the T1 space. Using the resulting fiber orientation distribution (.fib) files, 27 canonical white matter tracts were automatically reconstructed with DSI Studio's ICBM-152 template-based deterministic tracking[40]. All imaging metrics (FLAIR, T1 map, MTR, FA, MD) and tract masks were co-registered to the individual T1 structural space for consistency in subsequent analyses (Fig. 3).

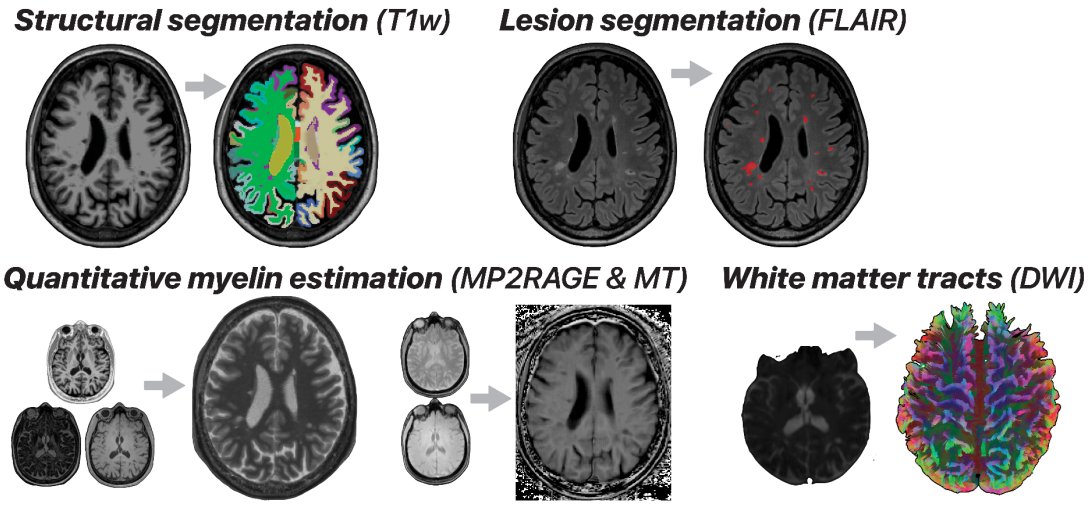

Figure 3 MRI acquisition and preprocessing. *UNI images from MP2RAGE were used for anatomical segmentation in FreeSurfer. FLAIR scans provided lesion masks generated by a U-Net model with manual correction. Quantitative maps included T1 relaxation and magnetization transfer ratio, and diffusion processing enabled tractography of major white-matter pathways.*

2.4 Lesion Load and Quantitative Metrics

2.4.1 Lesion location metrics

Lesion quantification was performed using two complementary frameworks: a classical anatomical definition, familiar to neurologists (periventricular, juxtacortical, infratentorial, and deep white matter), and a tract-based definition grounded in functional white matter pathways. In the classical framework, lesion masks were first intersected with the four canonical anatomical compartments, and lesion load was recorded as (1) lesion number (LN, count of discrete lesions), (2) lesion volume (LV, cumulative lesion volume in ml), and (3) normalized lesion volume (Lnorm, LV expressed as a percentage of total intracranial volume). In the tract-based framework, 27 subject-specific white matter fiber pathways which were collapsed into four functional clusters (association, projection/brainstem, occipitoparietal, and cerebellar pathways). For each cluster, lesion load was likewise defined as LN, LV, and Lnorm (LV expressed as a percentage of tract volume). Specific tract assignments are listed in Table 3 with representative renderings in Fig. 4. This dual-scale (classical-based vs tract-based) framework enabled direct testing of whether pathway-specific lesion mapping offers greater clinical insight than conventional anatomical compartmentalization. Note that although MS lesions are often prominent in periventricular

regions, callosal fibers (except Forceps Major) were excluded to avoid double-counting lesions in overlapping tracts.

2.4.2 Lesion composition metrics

For each participant, mean T1, MTR, FA, and MD values were calculated within lesions falling inside (1) the four classical compartments and (2) the four tract-based clusters. Because tract clusters form clearly defined fiber bundles, the same metrics were also sampled from NAWM by excluding lesion voxels from tract masks, yielding an internal reference free of visible pathology on 3D FLAIR. This generated, for each subject, a matrix of eight lesion-level measurements (four metrics × two location schemes) and a set of tract-specific NAWM values. NAWM sampling was not attempted for the classical compartments, whose diffuse anatomical borders preclude a reliable lesion-free reference. The combined lesion-based and tract-NAWM metrics constituted the lesion-composition feature set supplied to the statistical modeling stage.

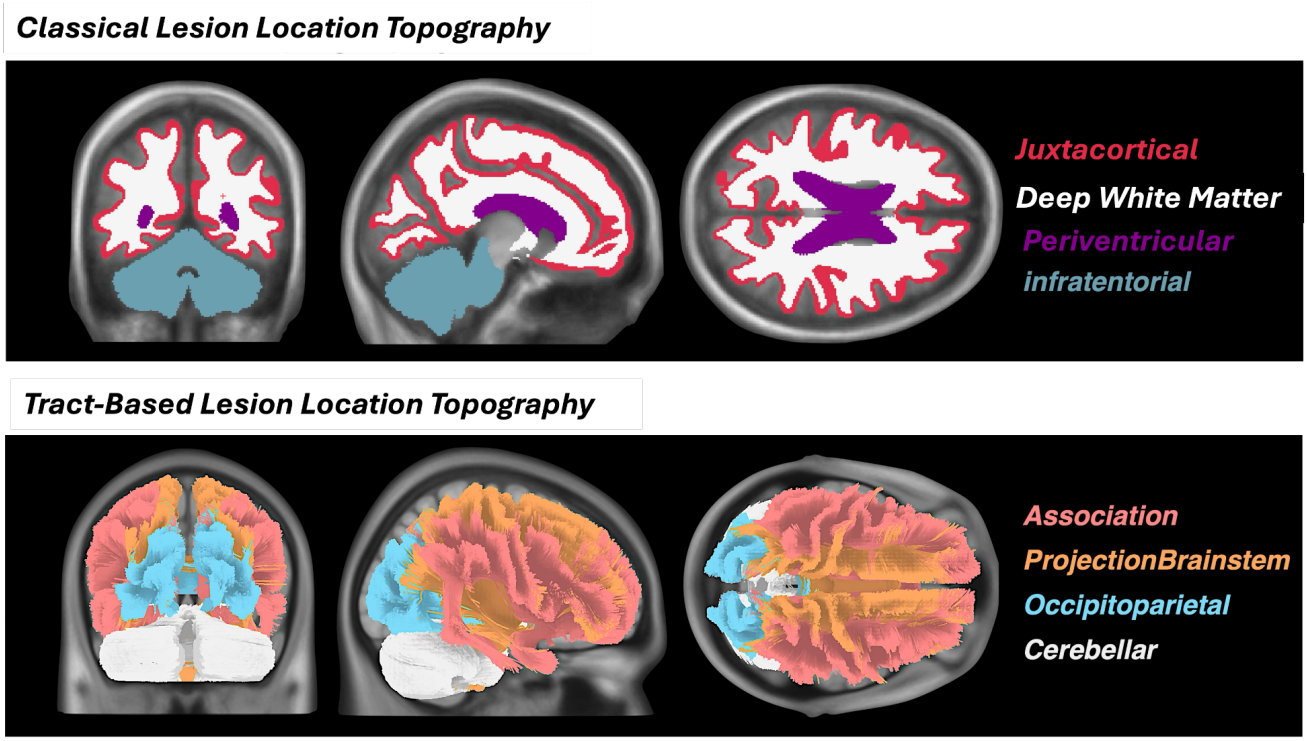

Figure 4 Defining Lesion Locations. *Lesions were classified based on either their classical anatomical regions (top panel) or their positions along white matter tracts (bottom panel). For clarity and comparability, four representative topographies were selected for each approach.*

Table 3 *Categorizes major white matter tracts into functional groups, including occipitoparietal, association, projection, brainstem, cerebellar, and commissural systems. Each group lists specific tracts (e.g., corticospinal tract under projection, corpus callosum under occipitoparietal) that were likely analyzed for lesion burden or structural integrity in the study.*

Tracts			
Occipitoparietal	**Association**	**Projection Brainstem**	**Cerebellar**
Projection Basal Ganglia: Optic Radiation	Arcuate Fasciculus	Corticospinal Tract	Inferior Cerebellar Peduncle
Vertical Occipital Fasciculus	Cingulum	Corticobulbar Tract	Vermis
Corpus Callosum: Forceps Major	Extreme Capsule	Corticopontine Tract	Middle Cerebellar Peduncle
	Frontal Aslant Tract	Medial Lemniscus	Superior Cerebellar Peduncle
	Hippocampus Alveus	Dentatorubro-thalmic Tract	
	Inferior Fronto Occipital Fasciculus	Non-Decussating Dentatorubro-thalamic Tract	
	Inferior Longitudinal Fasciculus	Reticular Tract	
	Middle Longitudinal Fasciculus	Medial Forebrain Bundle	
	Parietal Aslant Tract		
	Superior Longitudinal Fasciculus		
	Uncinate Fasciculus		
	Vertical Occipital Fasciculus		

2.5 Statistical Analysis

2.5.1 Descriptive Group Summaries

The objective was to explore subtype separation patterns by visualizing group-wise mean values and standard errors of the mean (SEM) across MS subtypes for both lesion location and lesion composition (including NAWM) using polar (radar) plots.

2.5.2 Univariate Correlation Analysis with Clinical Outcomes

Correlations were obtained between the various lesion metrics and clinical disability scores. For ordinal clinical variables (EDSS, MSPro), Spearman's rank correlation was used, while Pearson correlation was applied to continuous variables (T25FW, 9HPTD, 9HPTND). To control for potential confounding effects, partial correlations were calculated using patient sex, age, and disease duration as covariates. For each lesion metric–clinical score pair, the correlation coefficient (r) is reported, and statistically significant correlations ($p < 0.05$) are indicated with an asterisk (*).

2.5.3 Multivariate Predictive Modeling

Associations between MRI-derived metrics and clinical disability were modeled using ridge-penalized regression, comparing a Tract-Based model and a Classical-Based model. Binary outcomes representing clinically meaningful disability, EDSS $\geq$ 3 (indicating moderate impairment in at least one functional system) and MSPro $\geq$ 2 (indicating progressive disease state), were analyzed using ridge-penalized logistic regression. Continuous outcomes (T25FW, 9HPT-D, 9HPT-ND) were analyzed with ridge-penalized linear regression. The regularization parameter (λ) was optimized using 10-fold cross-validation. Model performance was summarized using AUC with bootstrapped 95% CIs and Akaike Information Criterion (AIC) for binary outcomes, and R^2 (reported with and without healthy controls) for continuous outcomes. Covariates varied by the clinical outcome: age, sex, and disease duration for binary outcomes; age and sex for continuous outcomes (disease duration is excluded when healthy controls are included). Each analysis was performed both with and without demographic covariates. All analyses were conducted in MATLAB R2024b[41].

For the Tract-Based model, predictors were defined as tail percentiles from tract voxel distributions (10th percentile for MTR/FA; 90th percentile for T1/MD). These cut-offs reflect that reduced FA/MTR and elevated T1/MD signify demyelination and axonal disorganization. This approach avoided missing values in tracts without visible lesions on FALIR scans, allowed inclusion of healthy controls, and captured the most abnormal voxels within each tract.

The Classical model used mean lesion values within the four predefined classical anatomical regions (see Fig 4, top panel). Because lesion presence was required in all four regions, patients

without complete lesion coverage were excluded. The resulting cohort size was reported for each analysis. Tail percentiles were not applied in the classical model, as lesion locations are undefined outside lesion masks.

3. Results

3.1 Classical versus tract-based lesion load: Group-wise Comparison of Lesion Frameworks

3.1.1 Lesion Location

The overall lesion load in the Classical method represents the total lesion volume derived from whole-brain 3D FLAIR. Overall lesion load was lowest in RRMS, increased in PPMS, and more than doubled in SPMS as shown in Fig 5a. Periventricular lesions accounted for the greatest share of this volume, followed by juxtacortical and deep-white-matter compartments, with infratentorial regions contributing the least. The tract-based metric, which includes only lesions intersecting predefined white-matter fiber bundles (Table 3), captured 22% of the total lesion volume in RRMS, 37% in PPMS, and 27% in SPMS. Within these restricted tract-based masks, lesion burden rose progressively from RRMS through PPMS to SPMS, reproducing the phenotype-dependent escalation observed with the classical measure.

Tract-based lesion loads, especially the Lnorm, separate the three MS subtypes far more clearly than classical region-based metrics, while simultaneously highlighting the functional pathways most relevant to disease evolution (radar plots in Fig 5b). Across all three tract-level measures, raw LN, LV and Lnorm, the distributions for RRMS, SPMS and PPMS are minimally or non-overlapping, mirroring the known escalation of disease burden. Lnorm, which adjusts for tract size and hemispheric side, produces the widest inter-quartile gaps. On the other hand, classical region-based metrics show inconsistent patterns, for instance the periventricular band, classical LN provides almost no separation between MS subtypes, whereas periventricular LV rises sharply, exaggerating group differences. The opposite pattern emerges in juxtacortical and deep-white-matter regions, where LN inflates differences and LV compresses them.

In the analyzed tracts, lesion load was highest in association fibers, followed by projection/brain-stem tracts, occipito-parietal bundles and cerebellar pathways. Because these tracts underpin vision, cognition, sensorimotor control and balance, their progressive involvement aligns with the stepwise clinical worsening observed across disability scores in Fig 2.

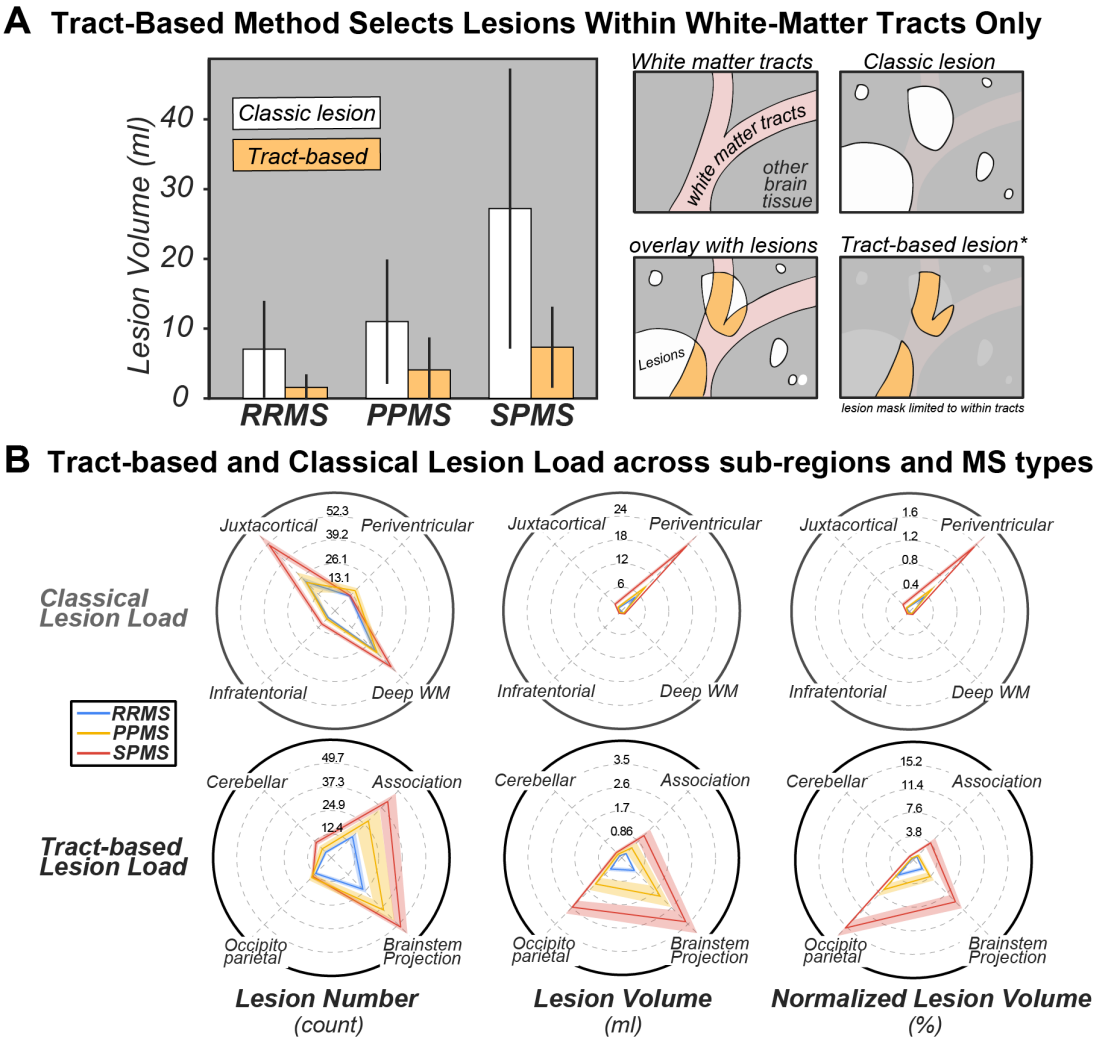

Figure 5 A) *Left panel: Box plots showing absolute whole-brain lesion volumes (mL) across multiple sclerosis (MS) subtypes, using classical (white bars) versus tract-based (yellow bars) lesion definitions. Error bars represent the SEM across subjects within each subtype group. The classical lesion volume represents total lesion burden without anatomical constraints, treating each lesion as a single contiguous mask. In contrast, the tract-based lesion volume includes only lesion voxels intersecting predefined white matter tracts, effectively excluding non-tract regions. Right panel: Schematic illustration of the tract-based filtering mechanism. Only voxels within white matter tracts or tract intersections are retained, whereas the classical approach includes the full lesion extent regardless of anatomical context.*

(B) *Radar plots comparing lesion load across MS subtypes (RRMS: blue; PPMS: yellow; SPMS: red), based on classical anatomical regions (top panel) and tract-based definitions (bottom panel). Lesion load metrics include Lesion Number (count), Lesion Volume (mL), and Normalized Lesion Volume (%).*

3.1.2 Lesion Composition

Quantitative assessment of lesion composition using tract-based regions provided greater separation between MS subtypes than classical anatomical lesion mapping. As shown in Fig 6, T1, MTR, FA, and MD values measured in lesions located within classical regions showed minimal differences between RRMS, PPMS, and SPMS groups. In contrast, tract-based lesion mapping showed subtype-specific trends in several white matter systems. The occipitoparietal tract demonstrated the most consistent separation across all tissue metrics when comparing MS subtypes. From RRMS to SPMS, we observed a 24.2% increase in T1, a 7.1% decrease in MTR, a 17.1% increase in MD, and a 7.5% decrease in FA. These directional changes across all four metrics reflect cumulative microstructural damage and provide a coherent tissue signature of progression within the occipitoparietal system. Other tract clusters demonstrated similar directionality but with varying magnitudes. In the cerebellar tract, FA and MTR values were greater in SPMS than PPMS, deviating from the trend seen in other regions.

Tract-based NAWM analysis showed a monotonic pattern across subtypes for all four metrics. T1 and MD values in occipitoparietal NAWM increased progressively from control to RRMS, PPMS, and SPMS. MTR and FA values showed a corresponding decrease across these groups. These directional changes were observed in most tract clusters, including the projection-brainstem and association tracts. In contrast, cerebellar NAWM again deviated from this pattern in FA, with SPMS showing greater values than controls.

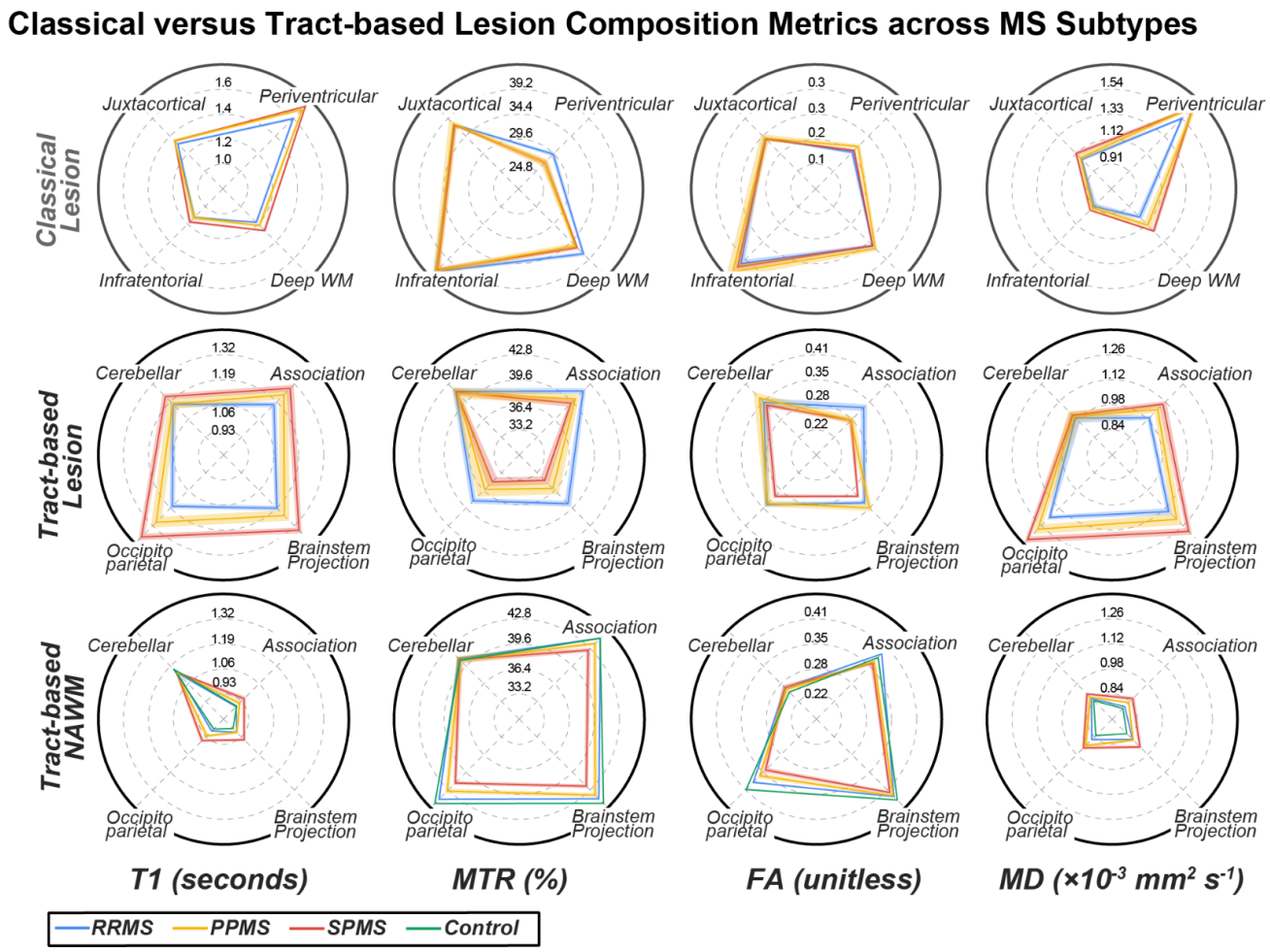

Figure 6 Radar plots illustrating lesion composition metrics across multiple sclerosis (MS) subtypes (RRMS: blue; PPMS: yellow; SPMS: red) and healthy controls (green). *Each quantitative MRI metric is shown for classical lesions (top row), tract-based lesions (middle row), and tract-based NAWM (bottom row). Quantitative parameters include T1 relaxation time (s), MTR (%), FA (unitless), and MD (×10⁻³ mm²/s).*

3.2 Classical versus tract-based: Univariate correlation analysis with clinical outcomes

We next evaluated the correlations between MRI biomarkers and disability measures, comparing tract-based versus anatomical lesion clustering, and testing whether lesion location alone is sufficient to explain clinical outcomes or whether adding lesion composition provides additional predictive value.

3.2.1 Lesion Location and Disability Correlation

Spearman's correlation analysis between lesion number and disability scores showed generally low associations across both classical and tract-based regions (Fig S1). In the classical regions, the highest correlation was observed in the infratentorial region with 9HPTND (r = 0.51). Among tract-based measures, the strongest correlation was observed in cerebellar tracts when correlated with 9HPTND (r = 0.45). Correlation values across other regions, especially periventricular region, and disability scores remained low.

Lesion volume demonstrated low to moderate correlations with disability (Fig S2). In classical regions, the periventricular area showed the highest associations with EDSS (r = 0.51) and 9HPTND (r = 0.57). Tract-based metrics showed the strongest correlations in the occipitoparietal (EDSS: r = 0.49; 9HPTND: r = 0.55) and projection-brainstem tracts (EDSS: r = 0.47; 9HPTND: r = 0.53).

Lnorm showed similar spatial patterns to absolute volume demonstrating moderate to low correlation strength across disability scales (Fig S3). In classical regions, the periventricular correlation with EDSS was r = 0.50 and with 9HPTND r = 0.56. Tract-based measures showed moderate associations, particularly in the occipitoparietal (EDSS: r = 0.50; 9HPTND: r = 0.55) and projection-brainstem tracts (EDSS: r = 0.47; 9HPTND: r = 0.58).

3.2.2 Lesion Composition and Disability Correlation

Across all metrics (Figs S4-S7), tract-based lesion regions consistently showed stronger correlations with EDSS compared to classical lesion regions, with particularly high associations observed in the occipitoparietal and projection-brainstem tracts. NAWM in these tracts also showed moderate associations with EDSS, while correlations with other clinical measures (e.g., T25FW, 9HPT, MSPro) were generally lower and more variable.

T1 values in tract-based lesions correlated strongly with EDSS (Fig. S4), particularly in the occipitoparietal (r = 0.83), projection-brainstem (r = 0.79), association (r = 0.75), and cerebellar (r = 0.68) tracts. In contrast, classical-based regions such as periventricular and deep white matter showed weaker EDSS correlations with T1 (r=0.47 and r=0.44, respectively). T1 values in tract-based NAWM showed low to moderate EDSS associations, with the highest correlations in the occipitoparietal (r = 0.52) and projection-brainstem (r = 0.52) tracts.

MTR values in tract-based lesions showed moderate EDSS correlations (Fig. S5), led by cerebellar (r = 0.59), association (r = 0.58), occipitoparietal (r = 0.54), and projection-brainstem (r = 0.51) tracts. In classical regions, EDSS associations were weaker (periventricular: r = 0.43; juxtacortical: r = 0.24; infratentorial: r=0.29; deep white matter: r=0.40). In tract-based NAWM, MTR–EDSS correlations peaked in the occipitoparietal (r = 0.63) and projection-brainstem (r = 0.56) tracts, while other tracts showed weaker associations.

FA in tract-based lesions correlated moderately with EDSS (Fig. S6), most strongly in cerebellar (r = 0.66) and occipitoparietal (r = 0.61) tracts, followed by association and projection-brainstem (both r = 0.59). Classical lesion regions showed no significant EDSS associations (r range: 0.01–0.14). FA values in tract-based NAWM were moderately associated with EDSS in occipitoparietal (r = 0.65) and projection-brainstem (r = 0.37) tracts, with no significant correlations in other tracts (r = 0.13). Other clinical scores showed modest associations in occipitoparietal region (T25FW, 9HPT: 0.29–0.39), with rest of associations of tract-based regions being weak or insignificant.

MD values in tract-based lesions yielded strong EDSS correlations (Fig. S7): occipitoparietal and projection-brainstem tracts both r = 0.80, followed by association (r = 0.77) and cerebellar (r = 0.69). Other clinical scores showed moderate correlations with tract-based measures (T25FW: 0.38–0.44; 9HPT: 0.37–0.39), while MSPro remained insignificantly associated (0.07–0.12). In classical lesions, MD–EDSS correlations were modest (deep white matter: r = 0.43; others < 0.33), with similarly insignificant or low correlations for other clinical scores. Tract-based NAWM MD also showed moderate EDSS associations (projection-brainstem: r = 0.66; occipitoparietal: r = 0.59), while other clinical correlations remained insignificant or low (T25FW and 9HPTND: 0.11–0.24).

In summary, univariate analyses showed that lesion location alone provided only modest associations with disability, with correlations generally weak across both classical and tract-based regions. Incorporating lesion composition strengthened these associations, particularly for tract-based metrics, where T1 and MD measures in occipitoparietal and projection–brainstem tracts correlated strongly with EDSS. However, correlations with other clinical outcomes such as T25FW, 9HPT, and MSPro remained limited.

3.3 Classical versus tract-based: Predictive Modeling

The goal of the multivariate predictive modeling was to evaluate whether tract-based frameworks outperform classical region-based approaches when predicting clinical disability in MS. Binary disability outcomes (Table 4) showed a clear advantage for the tract-based models. Across all MRI metrics, tract-based models achieved higher discrimination and better model fit than their classical counterparts for both EDSS and MSPro. The tract-based models reached AUCs of 0.86–0.96 compared with 0.57–0.85 for classical models. The AIC values were generally lower (i.e., better) in tract-based vs. classical-based models (e.g., 47.9 vs 79.9 for EDSS-T1; 38.2 vs 64.5 for MSPro-MTR). Continuous outcomes (Fig 7) demonstrated a similar pattern. Across all MRI metrics (T1, MTR, FA, MD) tract-based models explained more variance in clinical disability (T25FW, 9HPT-D, 9HPT-ND) than classical models. The improvement was most pronounced for FA and MTR, where tract-based R^2 values reached 0.19–0.43, compared with 0.01–0.34 for classical models. When demographic covariates were included, model performance improved across both binary and continuous outcomes (see Supplementary Table 1 and Fig. S8).

Table 4. *Summary of multivariate models comparing tract-based and classical approaches (without demographics) for binary disability outcomes (EDSS and MSPro). Model discrimination was measured by* ***AUC*** *(area under the ROC curve; higher values indicate better performance), and model fit was assessed by AIC (Akaike Information Criterion; lower values indicate a better fit).*

Clinical Metric (n=number of subjects)	Tract-based AUC (95% CI)	Classical-based AUC (95% CI)	Tract-based AIC	Classical-based AIC
EDSS-T1 (n=80)	0.962 (0.911-1.000)	0.850 (0.756-0.931)	47.834	79.868
EDSS-MTR (n=80)	0.906 (0.835-0.963)	0.732 (0.613-0.841)	66.309	81.932
EDSS-FA (n=79)	0.897 (0.823-0.959)	0.578 (0.440-0.707)	60.806	95.932
EDSS-MD (n=79)	0.902 (0.826-0.963)	0.861 (0.771-0.936)	67.109	76.189
MSPro-T1 (n=79)	0.893 (0.807-0.963)	0.832 (0.739-0.919)	31.616	65.178
MSPro-MTR (n=79)	0.888 (0.810-0.953)	0.660 (0.538-0.772)	38.196	64.461
MSPro-FA (n=78)	0.861 (0.772-0.931)	0.568 (0.438-0.697)	45.710	65.947
MSPro-MD (n=78)	0.886 (0.803-0.952)	0.709 (0.592-0.816)	47.064	64.616

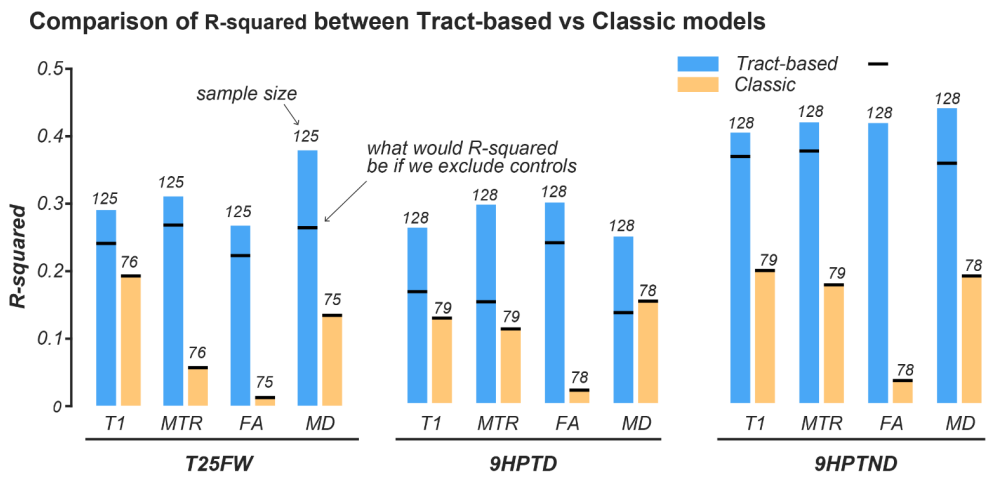

Figure 7. Comparison of explanatory power (R^2) between tract-based and classical models (without demographics) across continuous disability outcomes (T25FW, 9HPTD, and 9HPTND). *Each bar represents the proportion of variance in disability explained by the model for five imaging metrics (T1, MTR, FA, MD). Blue bars indicate tract-based models, and yellow bars represent classical models. Numbers above bars denote sample sizes used in each analysis, and black horizontal lines show R^2 values computed using matched samples (i.e., when controls were excluded to achieve equal n across models).*

4. Discussion

4.1 Study recap

This study aimed to improve the clinical relevance of lesion analysis in MS by refining how and where lesions are measured. While traditional approaches focus on anatomical lesion burden, "disconnection" studies have demonstrated that functional deficits in neurological disease often arise from tract-specific interruptions in distributed brain networks[20,26].

We hypothesized that filtering lesions through clusters of functionally meaningful white matter tracts rather than broad anatomical compartments would allow for more accurate identification of clinically relevant damage. The premise is intuitive: lesions that disrupt specific tracts or their intersections likely carry greater functional weight than those in other regions. Such filtering provides a biologically grounded and reproducible framework for lesion inclusion and exclusion, enabling more meaningful analysis of tissue properties within these tracts.

The clinico-radiological paradox in MS persists because not all lesions are the same. While lesion location has traditionally been used to explain disability[42,43], our findings confirm that lesion position alone, even when filtered through functionally relevant white matter tracts, offers limited explanatory power. This is expected: lesion number, volume, or normalized volume may reflect spatial patterns of disease but cannot differentiate between biologically distinct lesion types.

4.2 Tract-Based Filtering and the Value of Lesion Location

MS lesions are both spatially and biologically heterogeneous, varying across patients and within the same individual. A single patient may harbor lesions at distinct anatomical locations that differ in their functional importance, while at the same time, lesions at different disease stages (e.g., RRMS vs. SPMS) differ in composition, chronicity, and impact on disability[44,45]. This dual heterogeneity, of where lesions occur and what they contain, lies at the core of the paradox originally described by Barkhof, in which conventional MRI lesion load fails to reflect clinical severity[9].

Hence, capturing MS disease severity and progression with quantitative MRI hinges on two complementary dimensions: lesion topography (precise spatial distribution) and lesion microstructure (intrinsic pathology such as demyelination, axonal transection, and gliosis). Nonetheless, anatomically defined ROIs, like the periventricular compartment, treat every lesion that contacts the ependymal surface as equivalent, regardless of its microstructural damage or functional connectivity. Because the periventricular mask can extend far from the ventricular lining whenever plaques merely “touch” the ROI, confluent lesions in SPMS disproportionately inflate total burden and obscure stage-specific changes. Likewise, classical anatomical masks potentially include lesions irrelevant to key neural circuits while excluding strategically disruptive foci.

By restricting analysis to lesions within well-characterized, functionally important white-matter tracts, tract-based filtering reduces this noise and sharpens specificity. As illustrated in Fig. 5a, our approach isolates the spatial zones where lesion microstructure mirrors the connections/disconnections inside an ROI. Moreover, this framework now enables analysis of

NAWM within restricted dimensions defined by each tract, thereby reflecting underlying biological and molecular heterogeneity.

4.3 Integrating Tissue Composition Enhances Clinical Associations

Building on tract-based spatial filtering, we incorporated quantitative MRI modalities (T1, MTR, and diffusion) to probe underlying lesion pathology. While previous studies reported modest correlations between mean T1 values of whole-brain white matter lesions and EDSS[46] or between MTR of cortical lesions and EDSS[47], these approaches treated all lesions equally. In contrast, tract-based analysis minimizes contributions from silent regions and amplifies signals from strategically damaging lesions. T1 values from tract-based lesion masks showed strong associations with EDSS ($r_{occipitoparietal} = 0.82$, $r_{projection\text{-}brainstem} = 0.79$, $r_{association} = 0.77$, $r_{cerebellar} = 0.70$). These substantially outperformed classical regions such as periventricular ($r = 0.59$) and juxtacortical ($r = 0.26$). A similar pattern was observed with MTR that tract-based lesion MTR values yielded stronger correlations compared to the classical compartments. Importantly, these high correlations emerged only when location and composition were considered jointly; either factor alone offered diminished correlations. In NAWM, T1 and MTR values in the occipitoparietal and projection-brainstem tracts correlated with EDSS up to $r = 0.63$, supporting the role of diffuse microstructural injury in disability progression. This layered analysis also revealed that not all tracts contribute equally, while averaged tract clusters captured moderate correlations with other disability scores. Individual tracts, particularly those involved in sensorimotor integration and balance, showed sharper association.

Diffusion-based MRI methods reconstruct white matter architecture in vivo, enabling interrogation of structural integrity along biologically meaningful pathways. While classical lesion compartments (e.g., periventricular or juxtacortical regions) group tissue based on anatomical proximity, their clinical relevance has often been diluted by arbitrary anatomical ROIs that ignore white matter architecture. By anchoring diffusion metrics to specific tracts, we observed markedly stronger and more interpretable correlations with disability. This is not surprising: FA and MD reflect fiber coherence and water diffusivity along directional axes, and thus their biological relevance emerges most clearly when measured along coherent, functionally relevant fiber systems. Diffusion metrics in the occipitoparietal and association bundles consistently showed strong correlations with EDSS across both lesion and NAWM compartments, suggesting that

damage in these pathways integral to visuospatial coordination, attention, and sensorimotor integration[48] may be particularly impactful on global disability. On the other hand, the cerebellar tracts showed weaker associations, likely reflecting well-known challenges in tractography in this region, including signal dropout, complex fiber geometry, and distortion artifacts[49]. This again highlights that the quality of the spatial model directly affects the clinical signal: even biologically important structures may fail to correlate if their spatial definition is noisy.

4.4 Multivariate Comparison of Predictors

Tract-informed modeling substantially enhanced the prediction of disability and progression compared with classical region-based approaches. By integrating lesion location with microstructural composition through tail-based percentile metrics rather than FLAIR lesion masks, tract-based predictors captured the combined anatomical and pathological context underlying clinical impairment. The markedly higher AUC, lower AIC, and greater explained variance observed in tract-based vs classical-based models indicate that tract-based framework more accurately reflects the functional architecture disrupted in MS. In contrast, classical-based region metrics, which rely on FLAIR-derived lesion masks within broad anatomical boundaries, likely showed weaker associations with disability because of averaging over heterogeneous lesion properties. Mapping pathology along tract-defined rather than region-defined substrates improves prediction of clinically relevant damage and offers a more biologically grounded framework for modeling MS-related functional decline.

4.5 Methodological Considerations and Future Directions

While our findings highlight the clinical relevance of tract-informed lesion and NAWM metrics, several methodological limitations merit consideration. From a reproducibility standpoint, both MP2RAGE-based T1 mapping and MTR pose challenges for multi-center implementation. MP2RAGE, though offering superior tissue contrast and quantitative stability, is not yet widely available and requires longer acquisitions. Likewise, MTR measures show significant inter-vendor variability even under standardized conditions, reflecting incomplete harmonization of acquisition and processing workflow.[50] These limitations constrain cross-site comparability of quantitative T1 and MTR metrics and limit their large-scale clinical adoption. Ongoing efforts toward vendor-

neutral sequences and transparent workflows are essential to improve reproducibility and facilitate broader use in MS research.[50] A more practical approach is to use DTI metrics (FA and MD) to quantify lesion composition. DTI is likely the preferred method, as it provides both tract-based lesion localization and compositional information through FA and MD measures.

More broadly, while our tract-based framework enables refined tissue sampling, future work should further tailor tract filtering to specific clinical endpoints rather than applying universal lesion mask across all outcomes. The inclusion of spinal cord atrophy measures and advanced imaging modalities such as quantitative susceptibility mapping (QSM) could improve sensitivity to neurodegeneration, particularly in progressive MS[51,52]. Longitudinal studies will also be critical to validate whether these tract-specific markers predict clinical decline over time. Follow-up imaging will allow us to assess whether early qMRI abnormalities evolve in parallel with symptom progression, helping define actionable thresholds. Finally, methodological advances in cerebellar tract segmentation and motion correction will likely enhance correlation detection in regions currently affected by signal instability or tracking artifacts[49].

5. Conclusion

This study demonstrated that combining lesion location with microstructural composition through tract-based qMRI mapping offers a clinically meaningful advancement in MS progression prediction. The integration of spatial and biological lesion attributes translates into clinically interpretable risk estimation beyond statistical measures. As quantitative MRI becomes standardized and computational modeling more accessible, this framework enables imaging-driven tools for identifying at-risk patients and supporting tailored therapeutic decisions in routine practice. Ultimately, we believe this approach transforms lesion analysis from a descriptive exercise to a functional biomarker, bridging radiological findings with clinical outcomes.

Acknowledgements

This study was supported by the Core Technology Platforms (CTP), Biomedical Imaging Core at New York University Abu Dhabi (NYUAD). We also acknowledge the Computational Research Center at NYUAD for providing supercomputing resources.

Funding

This work was supported by the NYUAD Center for Brain and Health, funded by Tamkeen under the NYU Abu Dhabi Research Institute grant CG012. OA and AWA were supported by a grant from the National Multiple Sclerosis Society UAE, under project LAMINATE. AA is a member of the Public Health Research Center which is funded by Tamkeen through the NYUAD Research Institute (NYUAD-G1206).

Competing interests

The authors report no competing interests.

References

1. Lassmann H. The pathology of multiple sclerosis and its evolution. *Philos Trans R Soc Lond B Biol Sci*. 1999;354(1390):1635-1640.

2. Allen IV, McKeown SR. A histological, histochemical and biochemical study of the macroscopically normal white matter in multiple sclerosis. *J Neurol Sci*. 1979;41(1):81-91.

3. Bitsch A, Schuchardt J, Bunkowski S, Kuhlmann T, Brück W. Acute axonal injury in multiple sclerosis. Correlation with demyelination and inflammation. *Brain*. 2000;123 (Pt 6):1174-1183.

4. Cameron MH, Nilsagard Y. Balance, gait, and falls in multiple sclerosis. *Handb Clin Neurol*. 2018;159:237-250.

5. Younger DS. Multiple sclerosis: Motor dysfunction. *Handb Clin Neurol*. 2023;196:119-147.

6. Benedict RHB, DeLuca J, Enzinger C, Geurts JJG, Krupp LB, Rao SM. Neuropsychology of multiple sclerosis: Looking back and moving forward. *J Int Neuropsychol Soc*. 2017;23(9-10):832-842.

7. Glanz BI, Holland CM, Gauthier SA, et al. Cognitive dysfunction in patients with clinically isolated syndromes or newly diagnosed multiple sclerosis. *Mult Scler*. 2007;13(8):1004-1010.

8. Rae-Grant AD, Eckert NJ, Bartz S, Reed JF. Sensory symptoms of multiple sclerosis: a hidden reservoir of morbidity. *Mult Scler*. 1999;5(3):179-183.

9. Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. *Curr Opin Neurol*. 2002;15(3):239-245.

10. Barkhof F. MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS). *Mult Scler*. 1999;5(4):283-286.

11. Filippi M, Rocca MA, Comi G. The use of quantitative magnetic-resonance- based techniques monitor evolution multiple sclerosis. *Lancet Neurol*. 2003;2:337-346.

12. Moccia M, van de Pavert S, Eshaghi A, et al. Pathologic correlates of the magnetization transfer ratio in multiple sclerosis. *Neurology*. 2020;95(22):e2965-e2976.

13. Rahmanzadeh R, Weigel M, Lu PJ, et al. A comparative assessment of myelin-sensitive measures in multiple sclerosis patients and healthy subjects. *NeuroImage Clin*. 2022;36(103177):103177.

14. Rocca MA, Mastronardo G, Rodegher M, Comi G, Filippi M. Long-term changes of magnetization transfer-derived measures from patients with relapsing-remitting and secondary progressive multiple sclerosis. *AJNR Am J Neuroradiol*. 1999;20(5):821-827.

15. Ravano V, Piredda GF, Krasensky J, et al. Tract-wise microstructural analysis informs on current and future disability in early multiple sclerosis. *J Neurol*. 2024;271(2):631-641.

16. Gloor M, Andelova M, Gaetano L, et al. Longitudinal analysis of new multiple sclerosis lesions with magnetization transfer and diffusion tensor imaging. *Eur Radiol*. 2024;34(3):1680-1691.

17. Zollinger LV, Kim TH, Hill K, Jeong EK, Rose JW. Using diffusion tensor imaging and immunofluorescent assay to evaluate the pathology of multiple sclerosis. *J Magn Reson Imaging*. 2011;33(3):557-564.

18. Rimkus C de M, Junqueira T de F, Callegaro D, Otaduy MCG, Leite C da C. Segmented corpus callosum diffusivity correlates with the Expanded Disability Status Scale score in the early stages of relapsing-remitting multiple sclerosis. *Clinics (Sao Paulo)*. 2013;68(8):1115-1120.

19. Sechi E, Keegan BM, Kaufmann TJ, Kantarci OH, Weinshenker BG, Flanagan EP. Unilateral motor progression in MS: Association with a critical corticospinal tract lesion: Association with a critical corticospinal tract lesion. *Neurology*. 2019;93(7):e628-e634.

20. Chylińska M, Karaszewski B, Komendziński J, et al. The association between white matter tract structural connectivity and information processing speed in relapsing-remitting multiple sclerosis. *Neurol Sci*. 2023;44(9):3221-3232.

21. Madsen MAJ, Wiggermann V, Marques MFM, et al. Linking lesions in sensorimotor cortex to contralateral hand function in multiple sclerosis: a 7 T MRI study. *Brain*. 2022;145(10):3522-3535.

22. Kim M, Seo JW, Kim MS, Lee KH, Kim M. White matter tract density index is associated with disability in multiple sclerosis. *Neurobiol Dis*. 2024;198(106548):106548.

23. Winter M, Tallantyre EC, Brice TAW, Robertson NP, Jones DK, Chamberland M. Tract-specific MRI measures explain learning and recall differences in multiple sclerosis. *Brain Commun*. 2021;3(2):fcab065.

24. Harrison DM, Shiee N, Bazin PL, et al. Tract-specific quantitative MRI better correlates with disability than conventional MRI in multiple sclerosis. *J Neurol*. 2013;260(2):397-406.

25. Storelli L, Pagani E, Meani A, Preziosa P, Filippi M, Rocca MA. Advanced diffusion-weighted imaging models better characterize white matter neurodegeneration and clinical outcomes in multiple sclerosis. *J Neurol*. 2022;269(9):4729-4741.

26. Ravano V, Andelova M, Fartaria MJ, et al. Validating atlas-based lesion disconnectomics in multiple sclerosis: A retrospective multi-centric study. *NeuroImage Clin*. 2021;32(102817):102817.

27. Wilson M, Tench C, Morgan P, Blumhardt L. Pyramidal tract mapping by diffusion tensor magnetic resonance imaging in multiple sclerosis: improving correlations with disability. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2003;74:203-207.

28. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33(11):1444-1452.

29. Smith A. Symbol Digit Modalities Test. *PsycTESTS Dataset*. Published online June 13, 2016. doi:10.1037/t27513-000

30. Motl RW, Cohen JA, Benedict R, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult Scler*. 2017;23(5):704-710.

31. Feys P, Lamers I, Francis G, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler*. 2017;23(5):711-720.

32. Ziemssen T, Vandercappellen J, Jordan Mondragon V, Giovannoni G. MSProDiscuss™ clinical decision support tool for identifying multiple sclerosis progression. *J Clin Med*. 2022;11(15):4401.

33. Ziemssen T, Piani-Meier D, Bennett B, et al. A physician-completed digital tool for evaluating disease progression (multiple sclerosis progression discussion tool): Validation study. *J Med Internet Res*. 2020;22(2):e16932.

34. Pirpamer L, Kincses B, Kincses ZT, et al. Periventricular magnetisation transfer abnormalities in early multiple sclerosis. *NeuroImage Clin*. 2022;34(103012):103012.

35. Fischl B. FreeSurfer. *Neuroimage*. 2012;62(2):774-781.

36. Cardoso MJ, Li W, Brown R, et al. MONAI: An open-source framework for deep learning in healthcare. *arXiv [csLG]*. Published online November 4, 2022. http://arxiv.org/abs/2211.02701

37. Woolrich MW, Jbabdi S, Patenaude B, et al. Bayesian analysis of neuroimaging data in FSL. *Neuroimage*. 2009;45(1 Suppl):S173-86.

38. Karakuzu A, Boudreau M, Duval T, et al. qMRLab: Quantitative MRI analysis, under one umbrella. *J Open Source Softw*. 2020;5(53):2343.

39. Mitra-Behura S, Fiolka RP, Daetwyler S. Singularity containers improve reproducibility and ease of use in computational image analysis workflows. *Front Bioinform*. 2021;1:757291.

40. Hsu YC, Lo YC, Chen YJ, Wedeen VJ, Isaac Tseng WY. NTU-DSI-122: A diffusion spectrum imaging template with high anatomical matching to the ICBM-152 space. *Hum Brain Mapp*. 2015;36(9):3528-3541.

41. MATLAB. Accessed October 9, 2025. https://www.mathworks.com

42. Sanderson-Cimino M, Panizzon MS, Elman JA, et al. Periventricular and deep abnormal white matter differ in associations with cognitive performance at midlife. *Neuropsychology*. 2021;35(3):252-264.

43. Minneboo A, Barkhof F, Polman CH, Uitdehaag BMJ, Knol DL, Castelijns JA. Infratentorial lesions predict long-term disability in patients with initial findings suggestive of multiple sclerosis. *Arch Neurol*. 2004;61(2):217-221.

44. Vellinga MM, Geurts JJG, Rostrup E, et al. Clinical correlations of brain lesion distribution in multiple sclerosis. *J Magn Reson Imaging*. 2009;29(4):768-773.

45. Dutta R, Trapp BD. Relapsing and progressive forms of multiple sclerosis: insights from pathology. *Curr Opin Neurol*. 2014;27(3):271-278.

46. Harper JG, York EN, Meijboom R, et al. Quantitative T1 brain mapping in early relapsing-remitting multiple sclerosis: longitudinal changes, lesion heterogeneity and disability. *Eur Radiol*. 2024;34(6):3826-3839.

47. Amann M, Papadopoulou A, Andelova M, et al. Magnetization transfer ratio in lesions rather than normal-appearing brain relates to disability in patients with multiple sclerosis. *J Neurol*. 2015;262(8):1909-1917.

48. De Benedictis A, Duffau H, Paradiso B, et al. Anatomo-functional study of the temporo-parieto-occipital region: dissection, tractographic and brain mapping evidence from a neurosurgical perspective. *J Anat*. 2014;225(2):132-151.

49. Habas C, Manto M. Probing the neuroanatomy of the cerebellum using tractography. *Handb Clin Neurol*. 2018;154:235-249.

50. Karakuzu A, Biswas L, Cohen-Adad J, Stikov N. Vendor-neutral sequences and fully transparent workflows improve inter-vendor reproducibility of quantitative MRI. *Magn Reson Med*. 2022;88(3):1212-1228.

51. Zivadinov R, Weinstock-Guttman B, Pirko I. Iron deposition and inflammation in multiple sclerosis. Which one comes first? *BMC Neurosci*. 2011;12(1):60.

52. Tranfa M, Pontillo G, Petracca M, et al. Quantitative MRI in multiple sclerosis: From theory to application. *AJNR Am J Neuroradiol*. 2022;43(12):1688-1695.

Supplementary Material

Supplementary Table 1 *Summary of multivariate models comparing tract-based and classical approaches (with demographics) for binary disability outcomes (EDSS and MSPro). Model discrimination was measured by* ***AUC*** *(area under the ROC curve; higher values indicate better performance), and model fit was assessed by AIC (Akaike Information Criterion; lower values indicate a better fit).*

Clinical Metric (n=number of subjects)	Tract-Based AUC (95% CI)	Classical-Based AUC (95% CI)	Tract-Based AIC	Classical-Based AIC
EDSS-T1 (n=80)	0.956 (0.904–0.999)	0.885 (0.803–0.957)	47.834	79.868
EDSS-MTR (n=80)	0.919 (0.853–0.971)	0.871 (0.787–0.941)	66.309	81.932
EDSS-FA (n=79)	0.935 (0.876–0.978)	0.786 (0.679–0.881)	60.806	95.932
EDSS-MD (n=79)	0.923 (0.859–0.973)	0.897 (0.822–0.959)	67.109	76.189
MSPro-T1 (n=79)	0.989 (0.970–1.000)	0.934 (0.877–0.983)	31.616	65.178
MSPro-MTR (n=79)	0.979 (0.950–0.999)	0.929 (0.866–0.978)	38.196	64.461
MSPro-FA (n=78)	0.966 (0.926–0.993)	0.922 (0.852–0.976)	45.710	65.947
MSPro-MD (n=78)	0.965 (0.917–0.996)	0.926 (0.859–0.978)	47.064	64.616

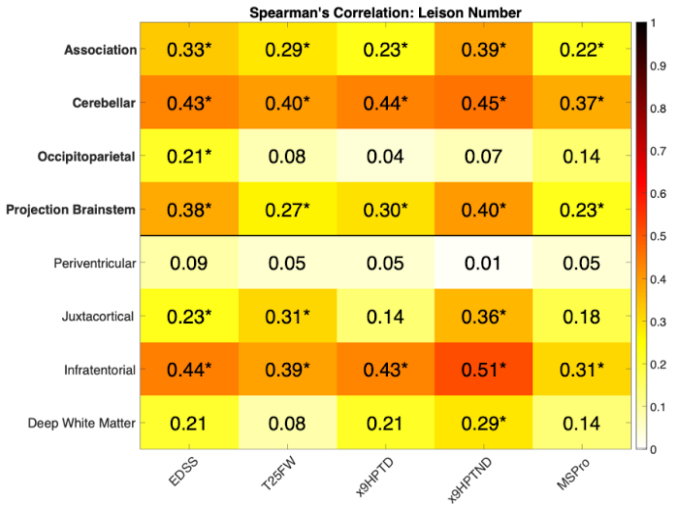

Figure S1 Spearman's correlation between lesion number in different locations (Tract-based vs Classical-based) and disability scores. In the heatmaps, darker shades indicate stronger absolute correlation values (range 0–1), with statistically significant associations annotated as (*).

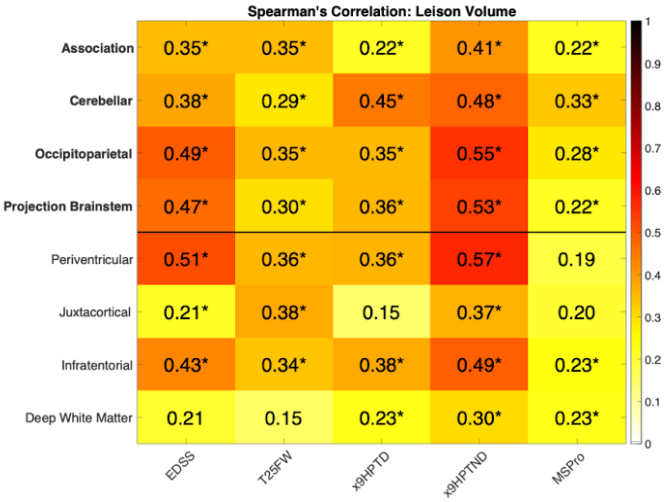

Figure S2 Spearman's correlation between lesion volume in different locations (Tract-based vs Classical-based) and disability scores. Statistically significant associations annotated as (*).

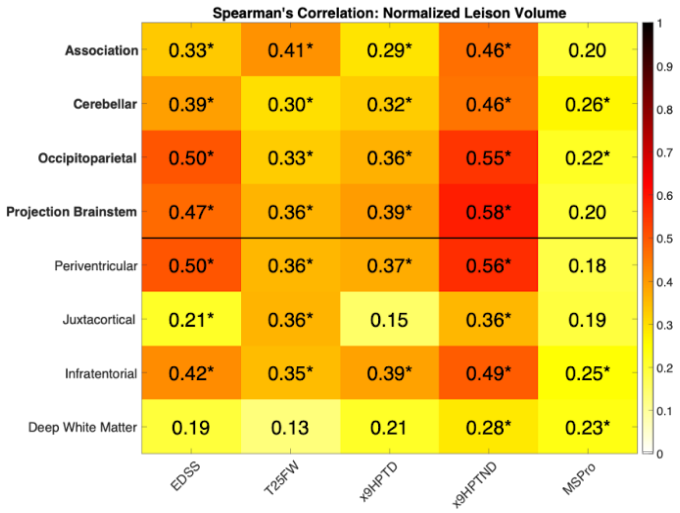

Figure S3 Spearman's correlation between normalized lesion volume in different locations (Tract-based vs Classical-based) and disability scores. Statistically significant associations annotated as (*).

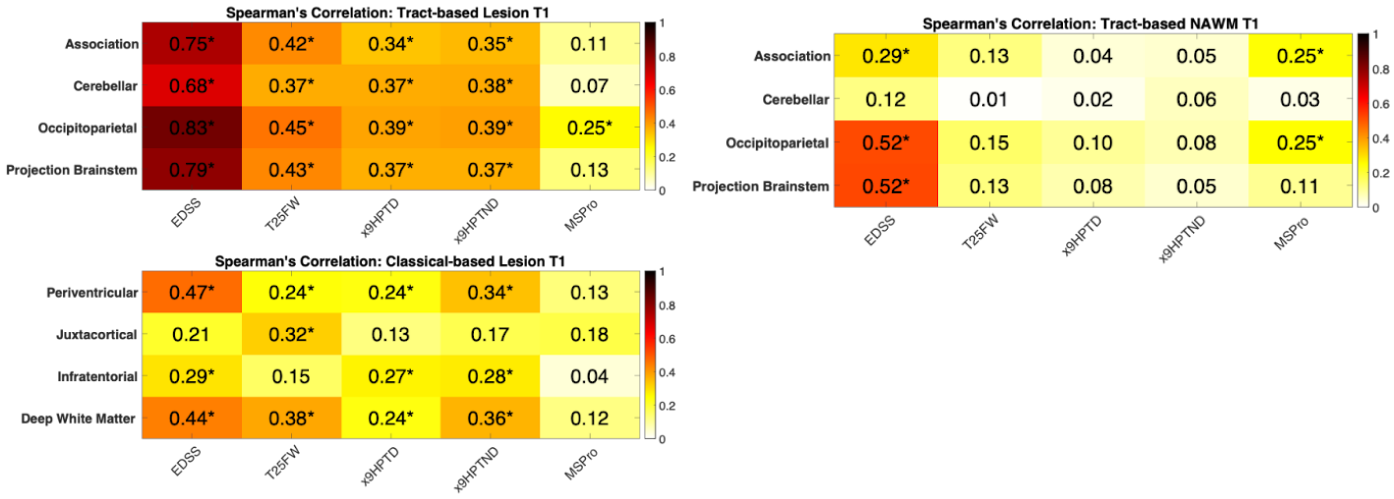

Figure S4 Spearman's correlation between T1 values of lesions (right) and NAWM (left), Tract-based vs Classical-based, and disability scores. Statistically significant associations annotated as (*).

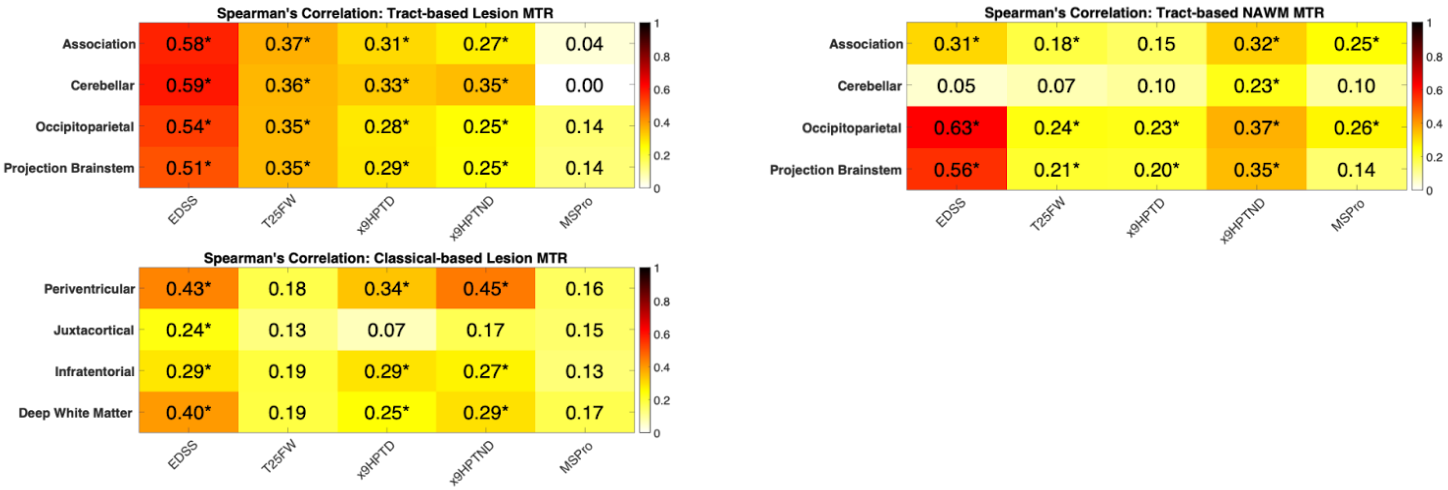

Figure S5 Spearman's correlation between MTR values of lesions (right) and NAWM (left), Tract-based vs Classical-based, and disability scores. Statistically significant associations annotated as (*).

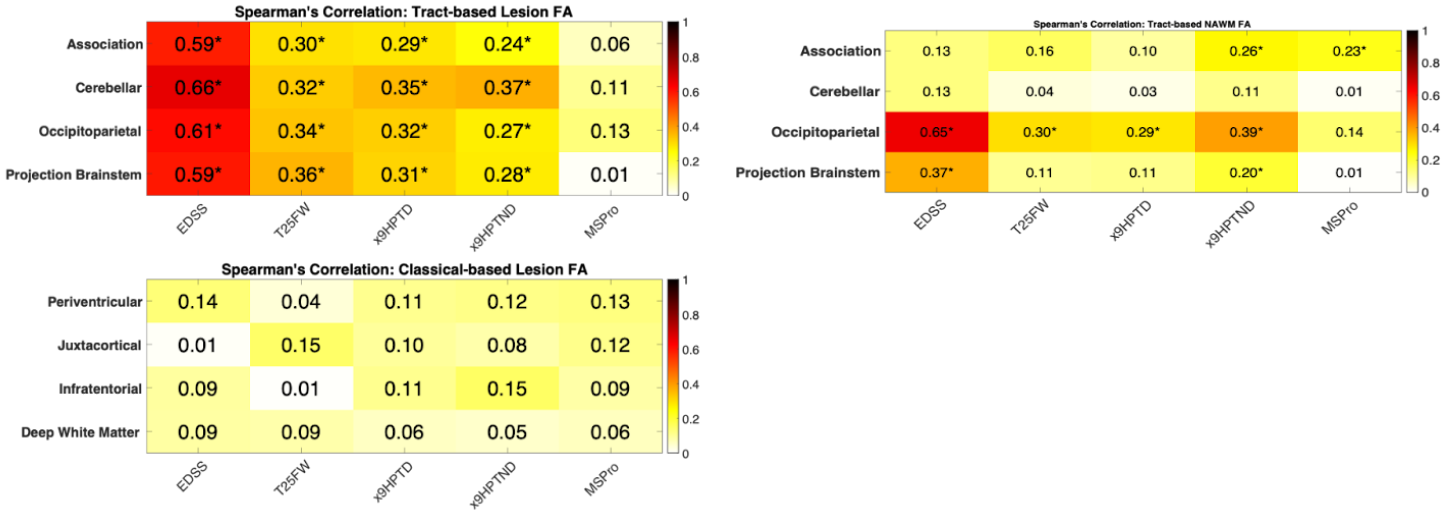

Figure S6 Spearman's correlation between FA values of lesions (right) and NAWM (left), Tract-based vs Classical-based, and disability scores. Statistically significant associations annotated as (*).

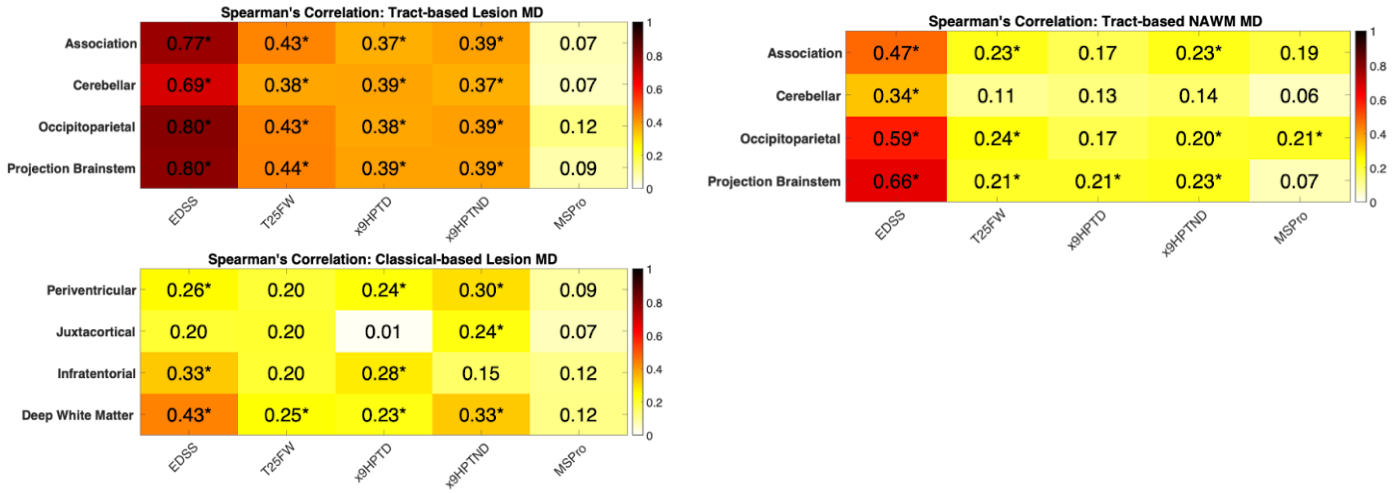

Figure S7 Spearman's correlation between MD values of lesions (right) and NAWM (left), Tract-based vs Classical-based, and disability scores. Statistically significant associations annotated as (*).

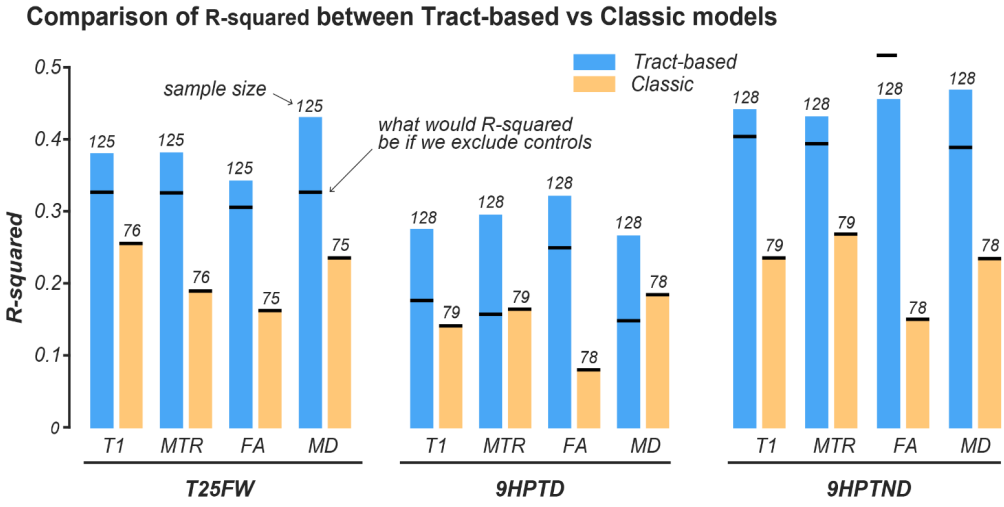

Figure S8 Comparison of explanatory power (R^2) between tract-based and classical models (with demographics) across continuous disability outcomes (T25FW, 9HPTD, and 9HPTND). *Each bar represents the proportion of variance in disability explained by the model for five imaging metrics (T1, MTR, FA, MD). Blue bars indicate tract-based models, and yellow bars represent classical models. Numbers above bars denote sample sizes used in each analysis, while black horizontal lines show R^2 values computed using matched samples (controls excluded, equal across models).*