



OPEN

DATA DESCRIPTOR

# Multiparametric MRI dataset for susceptibility-based radiomic feature extraction and analysis

Cristiana Fiscono<sup>1</sup>, Giovanni Sighinolfi<sup>2</sup>, David Neil Manners<sup>2,3</sup>✉, Lorenzo Motta<sup>2</sup>, Greta Venturi<sup>2</sup>, Ivan Panzera<sup>4</sup>, Fulvio Zaccagna<sup>5,6,7</sup>, Leonardo Rundo<sup>8</sup>, Alessandra Lugaresi<sup>1,4</sup>, Raffaele Lodi<sup>1,2</sup>, Caterina Tonon<sup>1,2</sup> & Mauro Castelli<sup>9</sup>

Multiple sclerosis (MS) is a progressive demyelinating disease impacting the central nervous system. Conventional Magnetic Resonance Imaging (MRI) techniques (e.g., T<sub>2</sub>w images) help diagnose MS, although they sometimes reveal non-specific lesions. Quantitative MRI techniques are capable of quantifying imaging biomarkers *in vivo*, offering the potential to identify specific signs related to pre-clinical inflammation. Among those techniques, Quantitative Susceptibility Mapping (QSM) is particularly useful for studying processes that influence the magnetic properties of brain tissue, such as alterations in myelin concentration. Because of its intrinsic quantitative nature, it is particularly well-suited to be analyzed through radiomics, including techniques that extract a high number of complex and multi-dimensional features from radiological images. The dataset presented in this work provides information about normal-appearing white matter (NAWM) in a cohort of MS patients and healthy controls. It includes QSM-based radiomic features from NAWM and its tracts, and MR sequences necessary to implement the pipeline: T<sub>1</sub>w, T<sub>2</sub>w, QSM, DWI. The workflow is outlined in this article, along with an application showing feature reliability assessment.

## Background & Summary

Quantitative Susceptibility Mapping (QSM)<sup>1,2</sup> is an advanced Magnetic Resonance Imaging (MRI) technique used to quantify and map the magnetic susceptibility ( $\chi$ ) of various structures within the body, with a primary focus on the brain. This technique is frequently employed to investigate and visualize medical conditions that alter the magnetic properties of bodily tissues, such as the presence of iron deposits, hemorrhages, or de-myelinating processes. Since the early 2010s, demand for more precise and quantitative assessments of magnetic tissue properties has driven interest in the application of QSM in neuroimaging.

The development of QSM builds on previous research in MRI and  $\chi$ -based techniques. In fact, numerous neurological conditions in both adults and children have demonstrated abnormal accumulations of blood-related substances or mineral deposits. Thus, Susceptibility-Based Imaging (SBI)<sup>3</sup> has found application in clinical settings. While traditional SBI methods, like Susceptibility Weighted Imaging (SWI)<sup>4,5</sup>, offer valuable insights, they come with certain limitations compared to QSM<sup>2</sup>, which provides quantitative measurements, enables comparisons within and between different groups, and facilitates the discrimination between diamagnetic and paramagnetic substances.

One of the most investigated applications of QSM is the examination of neurodegenerative and neuroinflammatory disorders<sup>6</sup>, including Multiple Sclerosis (MS), which is an autoimmune demyelinating disease affecting

<sup>1</sup>Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy. <sup>2</sup>Functional and Molecular Neuroimaging Unit, IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy. <sup>3</sup>Department for Life Quality Sciences, University of Bologna, Bologna, Italy. <sup>4</sup>UOSI Riabilitazione Sclerosi Multipla, IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy. <sup>5</sup>Department of Imaging, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom. <sup>6</sup>Department of Radiology, University of Cambridge, Cambridge, United Kingdom. <sup>7</sup>Investigative Medicine Division, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom. <sup>8</sup>Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano, Italy. <sup>9</sup>NOVA Information Management School (NOVA IMS), Universidade NOVA de Lisboa, Campus de Campolide, 1070-312, Lisbon, Portugal. ✉e-mail: [davidneil.manners@unibo.it](mailto:davidneil.manners@unibo.it)

	HC	MS
<b>N</b>	50	100
<b>F:M</b>	31:19	58:42
<b>Age (y)</b>	57.1 ± 16.7 (24–86)	47.6 ± 11.4 (23–73)

**Table 1.** Demographic details of the sample. Sex and age (mean ± standard deviation), with range in parenthesis, are given for healthy controls and patients with MS.

the Central Nervous System (CNS) and presenting a wide array of symptoms such as fatigue, motor or sensory loss in limbs, cognitive decline, and visual disturbances<sup>7</sup>. The current diagnosis of MS relies on the McDonald criteria, which combine clinical observations, laboratory biomarkers, and imaging data<sup>8</sup>. Traditional MRI has played a crucial role in these criteria and is widely used to monitor disease progression. Nevertheless, conventional MRI techniques reveal established lesions without providing detailed insights into the underlying mechanisms responsible for demyelination<sup>9</sup>: focal or confluent white matter (WM) alterations, which are found in over 95% of MS patients, can be detected, but the presence of these lesions is not an absolute requirement for diagnosing the disease, as non-specific WM lesions can also occur in healthy individuals over 50 years old or those affected by other neurological diseases.

The development of innovative quantitative MRI (qMRI) techniques capable of quantifying imaging biomarkers *in vivo* holds promise for exploring the microstructure of the brain<sup>10</sup> and its metabolic processes<sup>11</sup> to unravel the pathophysiology of MS and potentially reveal signs of pre-clinical inflammatory demyelination. QSM, as a qMRI technique, may aid in the investigation of both damaged and undamaged brain tissue in patients with MS<sup>12–15</sup>.

In the last twenty years, radiomics has emerged as a quantitative analytical tool for personalized medicine using medical imaging<sup>16,17</sup>. It includes a set of techniques that extract complex and multi-dimensional features from radiological images, including characteristics such as intensity histograms and textural patterns within the Region or Volume of Interest (ROI/VOI) analyzed<sup>18</sup>. Since its introduction, radiomics has found applications across several medical image modalities. In the field of MR neuroimaging, a significant focus has been directed toward the study of brain tumors, making use of anatomical T<sub>1</sub>- and T<sub>2</sub>-weighted (T<sub>1</sub>w and T<sub>2</sub>w) images for purposes such as tumor characterization and grading<sup>19,20</sup> or the assessment of treatment response and clinical outcomes<sup>21</sup>.

The dataset<sup>22</sup> we provide in this work includes QSM-based radiomic features from Normal Appearing White Matter (NAWM) and its tracts, extracted in a mixed group of patients with MS and healthy controls, and all the different MR sequences necessary to implement the pipeline: for each subject, morphological T<sub>1</sub>w and T<sub>2</sub>w, QSM and Diffusion Weighted Imaging (DWI) for Diffusion Tractography Imaging (DTI). An example of an application based on those data<sup>23</sup> is shown in the last section.

## Methods

**Participants.** For this work, data from 100 MS patients were obtained from the repository of the Neuroimaging Laboratory (IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy), along with data from 50 healthy control subjects. Inclusion criteria are as follows: patients had to be older than 18 years, with relapsing-remitting, primary progressive or secondary progressive MS course according to the 2017 revision of the McDonald diagnostic criteria, and to receive intravenous therapy with anti-CD20 monoclonal antibodies. Healthy controls had to be older than 18 years old; furthermore, the MRI exam had to be negative for abnormal cerebral atrophy, cortical and subcortical iron accumulation and CSF circulation disturbances. The demographic characteristics of the sample are summarized in Table 1.

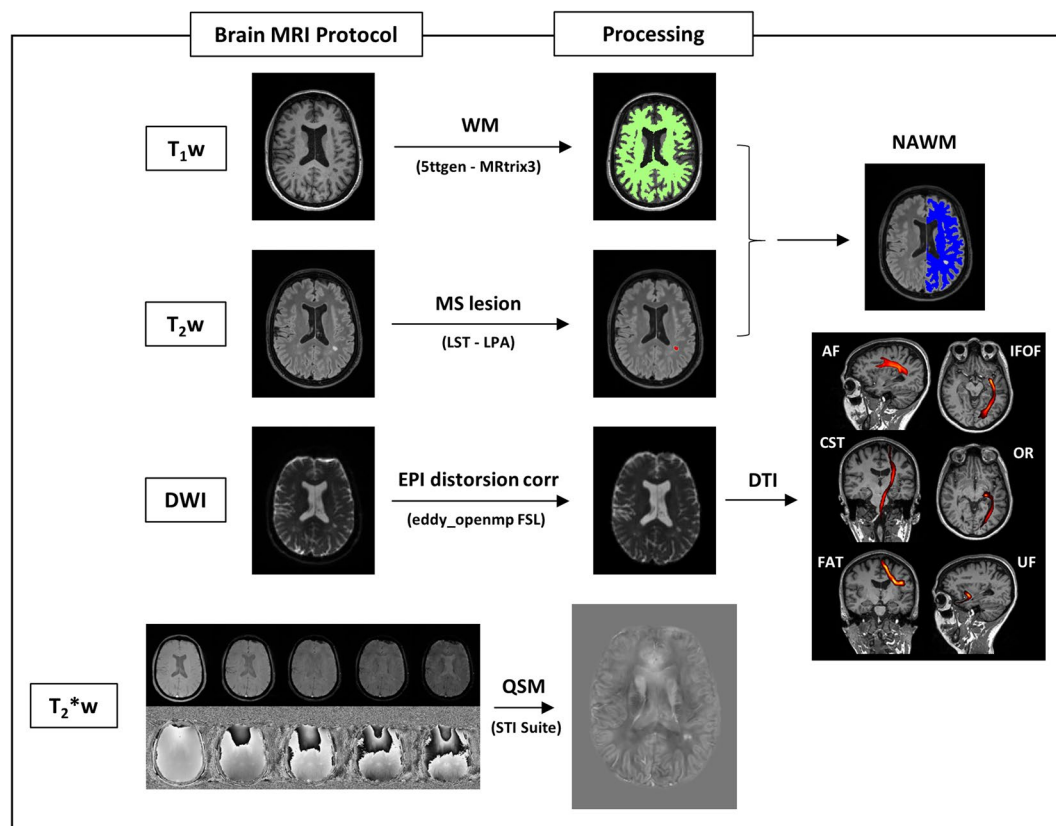
The data provided are related to scans performed between November 2020 and August 2023, except for one MR sequence for one healthy subject acquired in 2018. Acquisition and processing software and pipelines have basically remained the same throughout the relevant period, guaranteeing the homogeneity of the data used for analysis; small changes that were implemented are illustrated in the following sections.

**Ethical approval and informed consent.** Ethical approval for the study was obtained by the Institutional Review Board “Area Vasta Emilia Centro” (AVEC) (approval number AUSLBO 2023/CE 23043); written informed consent was obtained from all participants. Data has been anonymized and de-identified to ensure participating individuals cannot be identified from the information or images provided.

**MRI data acquisition.** Figure 1 shows the workflow for the acquisition and processing pipeline.

Scans were performed on a 3-T clinical scanner (Magnetom Skyra; Siemens Healthineers, Erlangen, Germany) equipped with a whole-body transmit and a 64-channel Head/Neck receiver coil. Sequences included in the MRI protocol and their details are shown in Table 2.

We modified the DWI sequence during the study, going from single-shell (5 unweighted volumes and 64 volumes with  $b = 2000 \text{ s/mm}^2$ ) to multi-shell (8 unweighted volumes, 12 volumes with  $b = 300 \text{ s/mm}^2$ , 30 volumes with  $b = 1000 \text{ s/mm}^2$  and 64 volumes with  $b = 2000 \text{ s/mm}^2$ ) acquisition. This change should have a negligible effect on tissue contrast for equivalently weighted volumes, given the processing method employed. In the dataset, the type of sequence is indicated for each subject. Sequences were acquired with Anterior-Posterior (AP) phase encoding. An additional sequence with inverted phase encoding (PA) (~ 4' scan time) was acquired to



**Fig. 1** Scheme of the acquisition and processing pipeline. The brain MRI protocol provided:  $T_1w$ ,  $T_2w$ , DWI and  $T_2^*w$  images (in this figure: patient with MS, F/38 years old): 1)  $T_1w$  images were used to obtain white matter segmentation using 5ttgen from MRtrix3 2) LPA algorithm from FSL was applied to  $T_2w$  images to obtain MS lesion mask. Merging WM segmentation with the MS lesion mask, the normal-appearing-white-matter segmentation mask was obtained. We run LST only on patients with MS; for controls, NAWM corresponds to WM from MRtrix3. 3) DWI images were pre-processed as explained in the specific section and diffusion tractography imaging automatic pipeline was applied, obtaining six white matter tracts (arcuate fasciculus, cortico-spinal tract, frontal aslant tract, inferior frontal-occipital fasciculus, optic radiation, uncinate fasciculus); VOIs from tractography reconstruction were merged with MS lesion mask to exclude damaged tissue. 4)  $T_2^*w$  images were processed to obtain QSM reconstructions, as explained in the specific section. All the images/masks were registered in the  $T_1w$  space. Radiomic features were extracted from QSM images in 14 volumes (six white matter tracts and total normal appearing white matter, left and right hemisphere).

	MPRAGE	FLAIR	HARDI DWI	QSM
	3D $T_1w$	3D SPACE $T_2w$	2D EPI single-shot	3D GRE $T_2^*w$
Plane	sagittal	sagittal	axial	axial
TR (ms)	2300	5000	4300	53
TE1 (ms)	2.98	428	98	9.42
$\Delta TE$ (ms)/n° TEs	—	—	—	9.42 / 5
TI (ms)	900	1800	—	—
SR (mm <sup>3</sup> )	1 × 1 × 1	1 × 1 × 1	2 × 2 × 2	0.5 × 0.5 × 1.5
FA (°)	9	120	90	15
Scan time	5'21"	5'55"	~9'	8'45"

**Table 2.** Details of the sequences included in the MRI protocol. (TE = echo time, TR = time of repetition, TI = Time of Inversion, FA = Flip Angle, MPRAGE = Magnetization Prepared RAPid Gradient Echo, SPACE = Sampling Perfection with Application optimized Contrast using different flip angle Evolution, FLAIR = FLuid Attenuated Inversion Recovery, GRE = GRAdient Echo, HARDI = High Angular Resolution Diffusion Imaging, EPI = Echo Planar Imaging).

correct EPI distortion artifacts in the EPI volumes: for single-shell measurements, three unweighted volumes were acquired, while for multi-shell, the sequence was re-acquired by repeating each volume acquisition with

weighting 0–1000 s/mm<sup>2</sup>. Concerning tractography reconstructions, we used volumes with a b-value of 2000 s/mm<sup>2</sup> for single- and multi-shell measurements without changing the pipeline. The extraction of radiomics features uses only the volumes of the reconstructed tracts, consistent between the two sequences. Raw data of unweighted and b = 2000 s/mm<sup>2</sup> volumes are available in the dataset<sup>22</sup>.

To qualify for inclusion in the dataset, complete and artifact-free MRI acquisition was essential for both patient and healthy control groups. Each examination comprises all the sequences (MPRAGE, FLAIR, DWI with both AP and PA phase encoding, QSM) necessary for the processing pipeline. The images were assessed to ensure they were of sufficient quality, devoid of significant motion artifacts or other distortions, to make them suitable for analysis using the standard pipeline.

MR sequences were derived from the same exam for all but three healthy subjects (subj-006 [F/57yo], subj-007 [F/46yo], and subj-040 [M/39yo]). In each of these instances, the FLAIR image has been acquired during a prior examination, approximately 30, 5, and 42 months before, respectively. For consistency, we include the FLAIR sequence for all the subjects, both patients and controls, even though it was used only to automatically segment patients' MS lesions; in this context, it was not used to analyze the healthy controls. Considering this and the fact that subj-006, subj-007 and subj-040 were healthy and not elderly, the mismatch between the days of the MRI exams does not affect either the pipeline or the extracted features.

**MRI data processing.** Raw DICOM data were converted to NIfTI format using `dcm2niix` (<https://github.com/rordenlab/dcm2niix>). Over the years, the software version changed from v1.0.20171215 to v1.0.20210317. This change does not affect recorded pixel intensities.

*Morphological T<sub>1w</sub> and T<sub>2w</sub> processing.* Original morphological images T<sub>1w</sub> and T<sub>2w</sub> were used without further processing; T<sub>2w</sub> images were linearly registered onto the corresponding T<sub>1w</sub> using Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL)<sup>24</sup> (v. 6.0.4) (FSL's Linear Image Registration Tool [FLIRT]<sup>25,26</sup>). All inter-modality registrations were performed using 12 degrees of freedom (affine transformations).

*QSM processing.* To obtain  $\chi$  maps, phase maps from the five echo times were processed individually by Laplacian unwrapping<sup>27</sup> and Variable kernel Sophisticated Harmonic Artifact Reduction for Phase data (V-SHARP) as background field removal<sup>28</sup>. The processing requires a brain binary mask obtained by skull-stripping the magnitude image of the first echo time using the FSL<sup>24</sup> Brain Extraction Tool (BET)<sup>29</sup>.

Multi-echo phase data were combined into a single-phase image through a weighted sum over echoes of processed phase maps, using as weights<sup>30</sup>:

$$w(TE_i) = \frac{TE_i e^{-TE_i/T_2^*}}{\sum_{j=1}^5 TE_j e^{-TE_j/T_2^*}}$$

The iterative least square (iLSQR) technique was used as a dipole inversion method<sup>31</sup>. STI Suite<sup>32</sup> was used for the processing. The resulting QSM image was linearly registered using FLIRT<sup>25,26</sup> to the corresponding morphological T<sub>1w</sub> image.

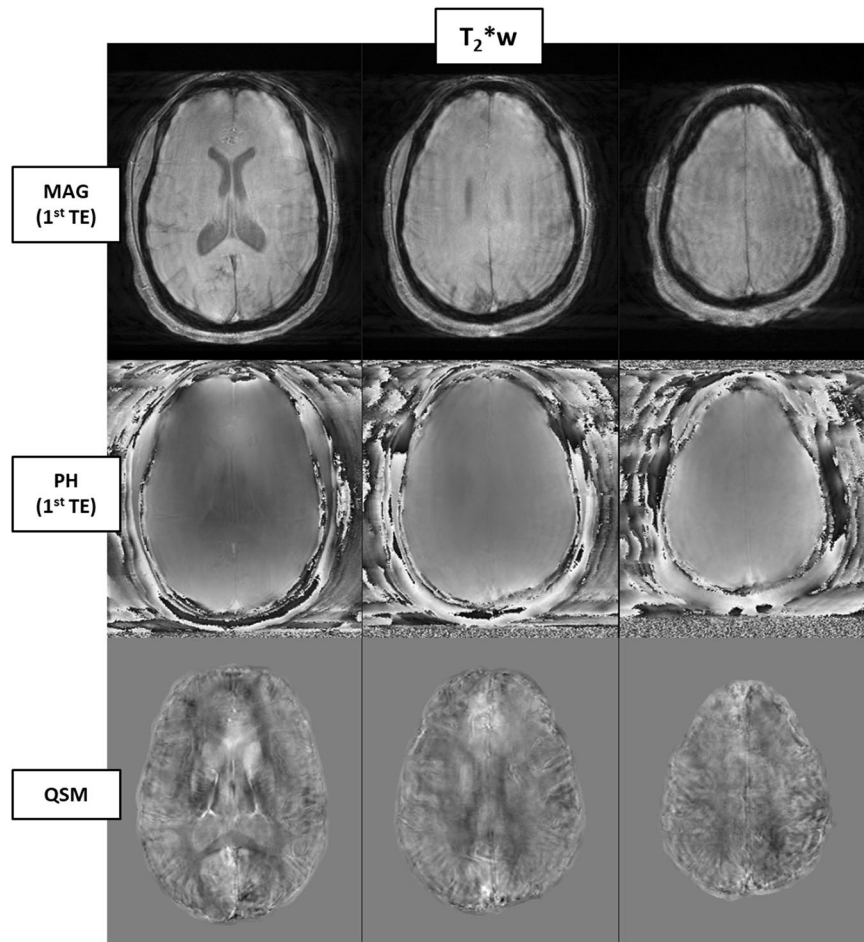
Cerebrospinal Fluid (CSF) served as the reference tissue, employing the *zero-referencing* method<sup>33</sup>. This choice is widely adopted because CSF susceptibility values do not show changes related to the subject's age or pathological condition, and orientation dependence can be excluded. However, since the sizes of the ventricles vary and the CSF signal often appears non-uniform in QSM reconstructions, we did not consider the entire ventricles but three small volumes instead (atrium, anterior horns, and central part), as suggested by Straub and colleagues<sup>34</sup>. An original atlas-based method was implemented to automatically identify the three volumes on each exam. Isotropic 1-mm MPRAGE maps of 60 subjects, both healthy controls and patients, were non-linearly registered (elastic registration) using FSL<sup>24</sup> FNIRT (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FNIRT>), after brain extraction using BET<sup>29</sup>, onto the space of a selected reference subject. The deformation matrix returned was combined with the linear transformation mapping QSM to T<sub>1w</sub> in the same exam so that all 60 QSM-magnitude maps were averaged in the same space.

Three small bilateral CSF volumes were manually defined on three adjacent slices of the atlas. The volumes were back-registered onto the original QSM map for each subject and the mean susceptibility value was calculated to establish the reference.

Motion artifacts pose a significant challenge for QSM due to its reliance on phase information extracted from MRI data. Even minor movements during image acquisition can introduce errors in phase consistency across slices or time frames, undermining the accuracy of susceptibility calculations. Notably, the QSM sequence is typically lengthy - lasting approximately 9 minutes in the current study -, allowing a lot of time for artifacts to occur. These artifacts manifest as geometric distortion, ghosting effects, and signal loss, ultimately compromising the quality and reliability of QSM images. To address this issue, rigorous quality control is essential. It was implemented by evaluating each QSM reconstruction individually.

Figure 2 shows an example of a QSM reconstruction excluded from the study due to motion artifacts. In this instance, the presence of motion was apparent even in the raw magnitude maps, facilitating the decision to exclude directly after the acquisition. However, QSM images are derived from phase data, which may not exhibit such evident artifacts in their raw form because of the discontinuities present before the unwrapping stage. Such artifacts can remain undetected until post-processing is performed.





**Fig. 2** Example of one exam (patient with MS, F/29 years old) excluded because of the quality of QSM image, showing movement artifacts and not considered suitable for the analysis: in the first and second row, magnitude and phase raw data from the first echo time; in the third row, the QSM reconstruction.

**DWI processing.** An automated diffusion and tractography processing pipeline was implemented as previously described<sup>35,36</sup>. DWI images underwent a process of skull-stripping using BET<sup>29</sup>. They were denoised using the **dwidenoise** function from MRtrix3<sup>37</sup> (v. 3.0.2) with a principal component analysis approach. To address susceptibility-related distortion, we employed the FSL<sup>24</sup> function **topup**. Subsequently, we addressed susceptibility effects, eddy currents, and signal dropout using the FSL<sup>24</sup> **eddy\_openmp** function<sup>38</sup>. Linear registration between diffusion measurements and  $T_1$ -weighted images was conducted using the **epi\_reg** function from FLIRT<sup>25,26</sup>. Diffusivity was modeled along the spatial eigenvector using the tensor model and a high-order fiber modeling technique. Moreover, we adopted a probabilistic streamline approach to evaluate crossing fibers.

Diffusion images were used to estimate fiber orientation distributions by the single tissue, single shell spherical deconvolution algorithm ‘**csd**’ implemented by the **dwi2fod** function<sup>39</sup> from MRtrix3. The Diffusion Tractography pipeline is described in the section VOI segmentation. In total, six WM tracts were defined: arcuate fasciculus (AF), cortico-spinal tract (CST), frontal aslant tract (FAT), inferior frontal-occipital fasciculus (IFOF), optic radiation (OR), and uncinate fasciculus (UF).

**VOI segmentation.** VOIs selected for feature extraction were segmented as follows:

- **White matter (WM):** WM segmentation was conducted using the MRtrix3 tool 5ttgen, which relies on FreeSurfer<sup>40</sup> (v. 6) segmentation, applied to the  $T_1$ -weighted ( $T_1w$ ) image.
- **MS lesions:** lesions were automatically segmented using the Lesion Prediction Algorithm (LPA)<sup>41</sup> from the Lesion Segmentation Tool (LST) (v. 3.0.0) ([www.statistical-modelling.de/lst.html](http://www.statistical-modelling.de/lst.html)), an open-source toolbox for Statistical Parametric Mapping (SPM) (v. 12). The toolbox generates an estimation of the lesion probability map, which was utilized to create a binary map of lesions. For each examination, the NAWM was identified by multiplying the inverse of this map with the WM mask. LST was applied only for patients’ exams; for controls, the NAWM corresponds to WM identified by MRtrix3.
- **DTI regions:** For each tract, seed and inclusion regions of interest for streamline generation and selection were defined on the Montreal Neurological Institute 152 (MNI152) standard brain. Non-linear registration between DWI and MNI spaces was performed using FNIRT from FSL<sup>24</sup>. Streamlines were generated using the

Tract	AF	CST	FAT	IFOF	OR	UF
<b>Seed</b>	Parietal lobe/peri-ventricular WM	Posterior Limb of Internal Capsule	Harvard-Oxford atlas supplementary motor area	Coronal plane at the anterior limit of the occipital lobe	Lateral geniculate nucleus	Superior temporal gyrus
<b>Obligatory waypoint(s)</b>	Frontal cortex and temporal cortex	Pons and precentral BN	Juelich atlas area BA44	Coronal plane, between the anterior limit of the genu of the corpus callosum and the anterior limit of the cingulate cortex	Visual cortex	VOI between insula and putamen
<b>Exclusion mask(s)</b>	Mid-sagittal plane	Mid-sagittal plane (*)	Mid-sagittal plane (*)	Mid-sagittal plane (*)	Mid-sagittal plane (*)	Mid-sagittal plane (*)
<b>Threshold (%)</b>	10	2	2	2	2	2

**Table 3.** Details for DTI implementation. For each tract, seed, inclusion and exclusion regions for streamline generation and selection were defined on the MNI152 standard brain. In the table, details for each tract are summarized. The mid-sagittal plane was used as an exclusion mask for all the tracts, adding VOIs for some of them (\*), specifically: squared VOI between the pons and middle cerebellar peduncles (axial/sagittal view) for CST; vertical plane at the anterior limit of the splenium of the corpus callosum, vertical plane at the anterior limit of the cingulate gyrus cortex, and VOI between insula and putamen for FAT; vertical plane anterior to mammillary bodies for IFOF; middle part of the pons/middle cerebellar peduncles (touching the posterior surface of the pons), WM/GM temporal lobe poles, under the superior temporal gyrus and laterally to uncus, mid-sagittal plane, superior to the pituitary gland and anterior to mammillary bodies and superior to posterior commissure/anterior to posterior limit of the splenium of the corpus callosum for OR; mid-coronal planes and lateral inferior limits of frontal lobes for UF. After the reconstruction, tracts were converted to a NIFTI image, and a minimum threshold was applied referring to the maximum value within the VOI. (AF = Arcuate Fasciculus, CST = Cortico-Spinal Tract, FAT = Frontal Aslant Tract, IFOF = Inferior Frontal-Occipital Fasciculus, OR = Optic Radiation, UF = Uncinate fasciculus, IS = inferior-superior, VOI = Volume of Interest, MNI152 = Montreal Neurological Institute's 152).

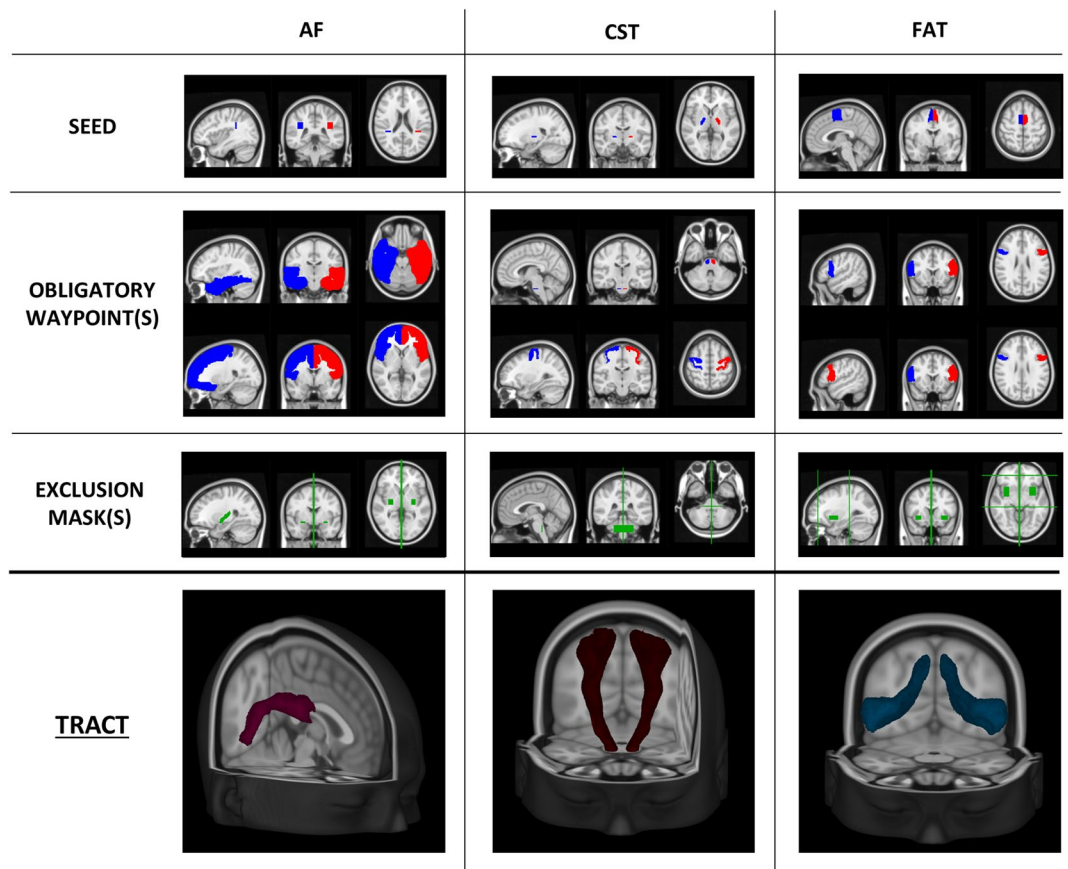
iFOD1 method tckgen from MRtrix3<sup>37</sup>, inclusion regions, FOD amplitude, and deviation angle. The criteria used to define each tract are presented in Table 3 and Figs. 3 and 4. Tracts were individually reconstructed for the two hemispheres and then combined into a single region, converted into a NIFTI image with voxel intensity representing the streamline count. To obtain a relative fiber count estimate image, a threshold was applied (referring to the maximum value). Similar to NAWM, tract VOIs were analyzed excluding the lesions.

**Feature extraction.** PyRadiomics<sup>42</sup> 3.0.1 (Python 3.7.6) was used to extract features from the VOIs overlain on the QSM images registered to the corresponding  $T_1w$  space. VOIs were likewise registered to the  $T_1w$  space, exploiting the previously obtained linear transformation matrices of FLAIR and DWI, using nearest neighbor interpolation to maintain binary masks. Since QSM and VOIs are in the same space, no additional interpolation operation was required for feature extraction, which was performed in 3D. MR exams were performed using the same clinical scanner and following the same acquisition protocol and processing pipeline, at the same center; this ensured sufficient homogeneity within the sample, obviating the need for histogram normalization steps.

Segment-based feature extraction was performed, meaning that each feature was extracted for each region of interest. Specifically, we considered 107 features, categorized as follows:

1. **First-order features (FO, # 18):** commonly used metrics to describe histogram intensity, including mean, median, 10<sup>th</sup> and 90<sup>th</sup> percentile, skewness, and kurtosis; FO measurements are independent of the number of Gray Levels (GLs);
2. **Shape 3D features (S3D, # 14):** descriptors of the 3D size and shape of the ROI (e.g., volume, surface, minimum and maximum axes); S3D measurements are independent of the number of GLs and their intensity distributions;
3. **Gray Level Co-occurrence Matrix features (GLCM, # 24):** describe the second-order joint probability function of an image region constrained by the mask;
4. **Gray Level Run Length Matrix features (GLRLM, # 16):** quantify GL runs (number of consecutive pixels that have the same grey level value);
5. **Gray Level Size Zone Matrix features (GLZM, # 16):** quantify GL zones in an image (number of the connected voxels that share the same grey level intensity);
6. **Neighboring Gray Tone Difference Matrix features (NGTDM, # 5):** quantify the difference between a GL value and the average grey value of its neighbors;
7. **Gray Level Dependence Matrix features (GLDM, # 14):** quantify GL dependencies in an image (number of connected voxels within a distance that are dependent on the central voxel).

Feature extraction was conducted with 64 as the number of gray levels (GLs), a parameter determined from the outcomes of a prior optimization study<sup>23</sup>. Categories 3 to 7 are referred to as texture and offer insights into the spatial distribution of intensity levels in the image. A complete list of the features can be found in the PyRadiomics documentation (<https://pyradiomics.readthedocs.io/en/latest/features.html>).



**Fig. 3** Seed, inclusion and exclusion regions for AF, CST and FAT. Regions were defined on the MNI152 standard brain and used for streamlining generation and selection to reconstruct WM tracts. Region descriptions are in Table 2. In the last row, the appearance of each tract reconstruction is shown in the MNI152 space, obtained by averaging tracts of 30 healthy controls (AF = Arcuate Fasciculus, CST = Cortico-Spinal Tract, FAT = Frontal Aslant Tract, OB WAY = Obligatory Waypoint(s), Excl = Exclusion mask(s), MNI152 = Montreal Neurological Institute's 152, WM = White Matter).

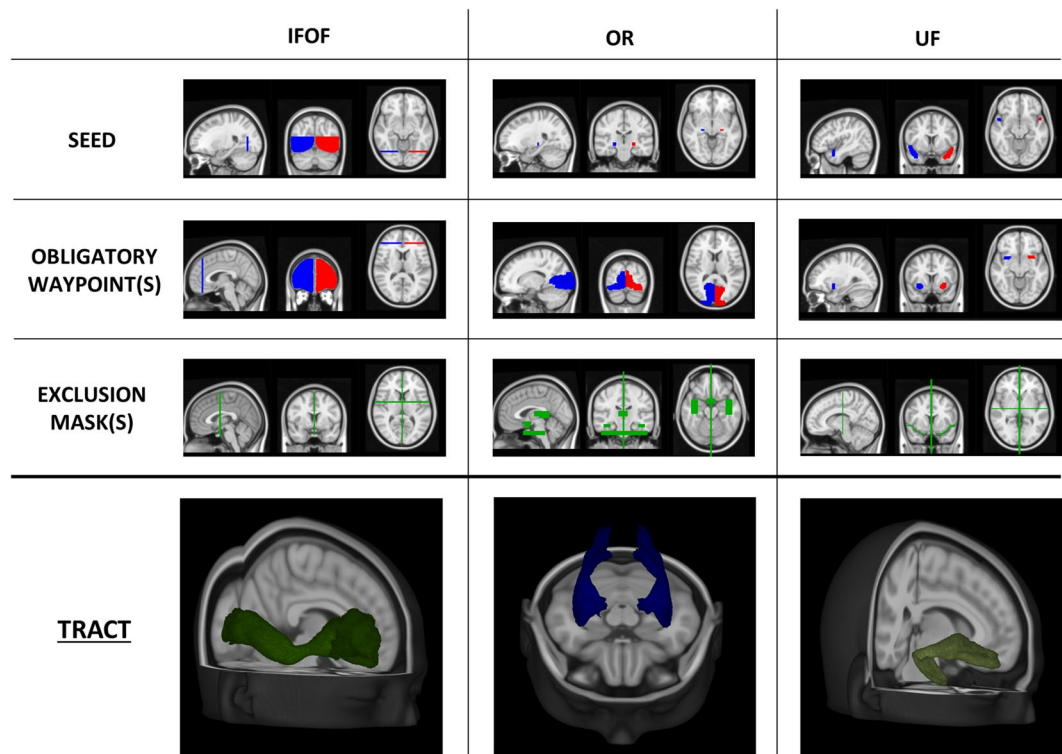
**Anonymization and de-identification.** All images were anonymized and de-identified. The anonymization mask was created by combining: (1) the mask obtained processing  $T_1w$  images with the automated defacing tools `mri_deface`<sup>43</sup> from FreeSurfer<sup>40</sup>; (2) the mask obtained processing  $T_1w$  images with SIENAX<sup>44</sup> from FSL<sup>24</sup>; (3) the mask obtained with BET<sup>29</sup> skull-stripping of the magnitude of the first echo time of the QSM sequence. Image files contain no metadata that can be used to identify study participants.

### Data Records

Files are organized according to the Brain Imaging Directory Structure (BIDS)<sup>45</sup> (Fig. 5).

In the dataset, files are grouped in 150 folders, each corresponding to an individual subject, and organized as follows:

- In the 'anat' folder there are: a)  $T_1$ -weighted ( $T_1w$ ),  $T_2$ -FLAIR, and multi-echo GRE  $T_2^*w$  images (MEGRE) for QSM reconstruction, serving as non-structural MR images for each subject. For MEGRE, both phase and magnitude contributions are specified for each echo time; b) The QSM reconstruction, denoted with the suffix 'chimap', as a parametric non-structural image, along with the affine registration matrix to  $T_1w$ . Specifically, the magnitude of the first echo time was used to register QSM to  $T_1w$ . Given that orientation with respect to  $B_0$  is essential information for QSM processing, raw images are stored without registration, with the registration matrix to  $T_1w$  provided separately, while the QSM reconstruction is already registered in  $T_1w$  space.
- In the 'dwi' folder, there are: a) the diffusion weighted image after the correction for EPI distortions, susceptibility effects, eddy currents, and signal dropout, b) the required bvals and bvecs files, providing gradient orientation information corresponding to DWI volumes available ( $b = 0$  and  $b = 2000$  s/mm<sup>2</sup>), and c) the affine registration matrix to  $T_1w$ .
- In the 'derivatives' folder there are the outputs from the processing pipeline. Two-subfolder were created: a) the segmentation folder, containing tracts and normal appearing white matter (NA-WM and tracts [AF, CST, FAT, IFOF, OR, UF] VOIs, divided for the two hemispheres) and b) the 'radiomic\_features' folder, containing the.csv with all the features for each volume of interest (seven features for each hemisphere, resulting in a total of 14 features).



**Fig. 4** Seed, inclusion and exclusion regions for IFOF, OR and UF. Regions were defined on the MNI152 standard brain and used for streamlining generation and selection to reconstruct WM tracts. Region descriptions are in Table 2. In the last row, the appearance of each tract reconstruction is shown in the MNI152 space, obtained by averaging tracts of 30 healthy controls (IFOF = Inferior Frontal-Occipital Fasciculus, OR = Optic Radiation, UF = Uncinate fasciculus, OB WAY = Obligatory Waypoint(s), Excl = Exclusion mask(s), MNI152 = Montreal Neurological Institute's 152, WM = White Matter).

- Outside the folder of individual subjects, there is the 'code' folder containing the scripts used for processing tasks such as image registration, radiomic feature extraction, and robustness evaluation.
- Additional files are stored at the top level: a) dataset\_description.json and README.txt files, describing the dataset; b) participants file (participants.xlsx and participants.json), containing the ID number, age, sex, scan date, clinical condition (denoted as 'HC' for healthy controls and 'MS' for multiple sclerosis patients), and the type of DWI sequence used (single-shell or multi-shell).

The size of the dataset<sup>22</sup> is ~30GB and it is available the Zenodo repository.

### Usage Notes

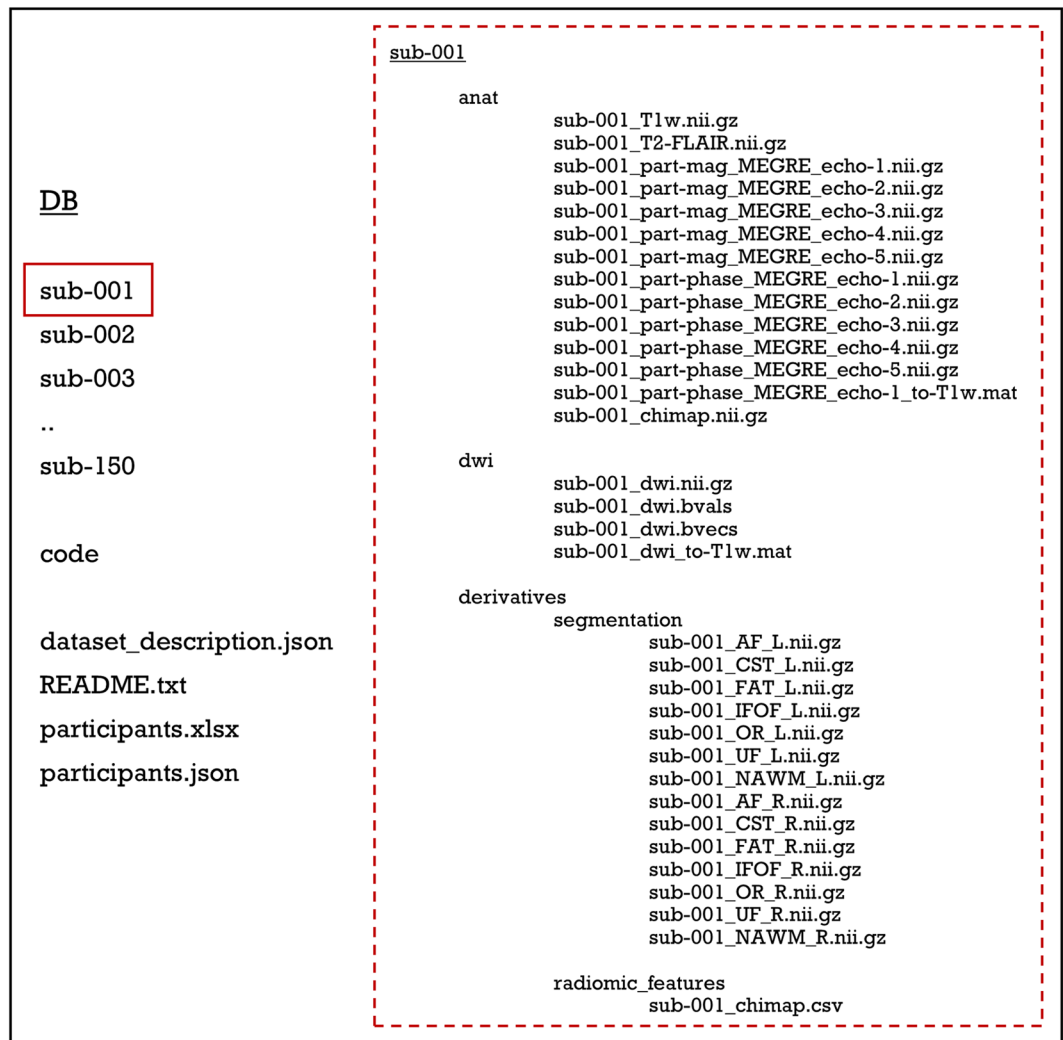
To obtain access to the dataset<sup>22</sup>, the following steps must be followed:

- 1) An account must be created on the Zenodo repository, ensuring that either the username or the email address (or both) are public.
- 2) The Data Use Agreement, provided in the description of the dataset, must be completed, signed and send to radiomicsdataset@live.unibo.it.
- 3) Access will be granted within two working days, providing 'Reader' status to a community in which the dataset is included; the status can be monitored directly on Zenodo or via email.
- 4) Upon acceptance, access to the dataset will be granted, allowing for data download.

### Technical Validation

**Assessing robustness of susceptibility-based radiomic features.** Hundreds of radiomic features can be extracted from MR images, enabling automated, high-throughput quantification of image characteristics<sup>45</sup>. This process provides a potentially extensive source of pathology-related biomarkers. It is essential to differentiate between image measurements and biomarkers<sup>46</sup>. Biomarkers are objective and quantifiable descriptors of biological processes capable of consistently predicting clinical outcomes and endpoints. Interpretability and reproducibility are two critical aspects in this context. In medical applications, it is crucial to maintain a clinical perspective as the driving force behind research<sup>47</sup>. The validity and relevance of a biomarker, along with its utility in clinical practice and its ability to offer valuable information, must be confirmed. Furthermore, when an informative and relevant feature is identified, it is vital to ensure the generalizability and replicability of outcomes.





**Fig. 5** Schematic diagram of the dataset organization. A list of the participants to the dataset is provided and for each subject: anatomical  $T_1w$  and  $T_2w$ ; DWI after the correction for EPI distortions and susceptibility effects, eddy currents and signal dropout, with b-values, b-vectors and the registration matrix to  $T_1w$ ; for QSM, original magnitude and phase maps for the five echo times, final QSM reconstruction registered in  $T_1w$  space, with registration matrix to  $T_1w$ ; VOIs (AF, CST, FAT, IFOF, OR, UF) for left and right hemisphere; csv files containing the 107 radiomic features for each region. Images were anonymized and de-identified ( $T_1w = T_1$ -weighted,  $T_2w = T_2$ -weighted, DWI = Diffusion-Weighted Imaging, QSM = Quantitative Susceptibility Mapping, VOI = Volume of Interest).

Since these features represent the outcome of physiological processes, they should remain consistent regardless of the imaging system or processing workflow used. The stages involved, from scanning to feature extraction, are numerous and complex. Thus, many parameters can impact the reliability of the results. In radiomic applications, robustness analysis is therefore essential to maintain the consistency of outcomes across different systems and multiple centers<sup>45,48</sup>.

In<sup>23</sup>, we describe a novel investigation into the robustness of susceptibility-based features derived from QSM. The study involved a cohort of 121 patients with MS and 30 healthy controls. We implemented an original and robust pipeline; we analyzed NAWM both as a whole and within the six clinically relevant tracts included in our dataset. To explore feature reliability, we varied the number of gray levels and echo times used for QSM reconstructions. After optimizing the number of GLs, set at  $n = 64$ , we found at least 65% of the features demonstrated robustness for each volume of interest. Notably, WM tracts exhibited higher levels of reliability, with over 75% of robust features in all of them. Differences among these tracts were explicable due to the volume of the structure and the susceptibility variance. No significant differences were observed between the left and right hemispheres.

The research confirmed the robustness of the data processing pipeline and established the reliability of QSM-based radiomic features against gray levels and echo times. This work paves the way for future investigations, where the identified set of reliable features may be used to characterize patients with MS, distinguish clinical phenotypes, and identify different responses to treatment.

## Code availability

We hosted the code used for the study in the Zenodo repository<sup>22</sup>; the scripts provided can be used for image registration, radiomic feature extraction, and robustness analysis.

Received: 12 December 2023; Accepted: 24 May 2024;

Published online: 04 June 2024

## References

- Deistung, A., Schweser, F. & Reichenbach, J. R. Overview of quantitative susceptibility mapping. *NMR Biomed.* **30**(4) (2017).
- Harada, T. *et al.* Quantitative susceptibility mapping: basic methods and clinical applications. *Radiographics.* **42**(4), 1161–1176 (2022).
- Soman, S. *et al.* Susceptibility-based neuroimaging: standard methods, clinical applications, and future directions. *Curr Radiol Rep.* **5**(3), 11 (2017).
- Haacke, E. M., Mittal, S., Wu, Z., Neelavalli, J. & Cheng, Y. C. Susceptibility-weighted imaging: technical aspects and clinical applications, part 1. *AJNR Am J Neuroradiol.* **30**(1), 19–30 (2009).
- Mittal, S., Wu, Z., Neelavalli, J. & Haacke, E. M. Susceptibility-weighted imaging: technical aspects and clinical applications, part 2. *AJNR Am J Neuroradiol.* **30**(2), 232–52 (2009).
- Ravanfar, P. *et al.* Systematic review: quantitative susceptibility mapping (QSM) of brain iron profile in neurodegenerative diseases. *Front Neurosci.* **15**, 618435 (2021).
- Reich, D. S., Lucchinetti, C. F. & Calabresi, P. A. Multiple sclerosis. *N Engl J Med.* **378**(2), 169–180 (2018).
- Thompson, A. J. *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* **17**(2), 162–173 (2018).
- Filippi, M. *et al.* MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. *Lancet Neurol.* **15**(3), 292–303 (2016).
- Sowa, P. *et al.* Restriction spectrum imaging of white matter and its relation to neurological disability in multiple sclerosis. *Mult Scler.* **25**(5), 687–698 (2019).
- Grist, J. T. *et al.* Imaging intraslesional heterogeneity of sodium concentration in multiple sclerosis: Initial evidence from <sup>23</sup>Na-MRI. *J Neurol Sci.* **387**, 111–114 (2018).
- Marcille, M. *et al.* Disease correlates of rim lesions on quantitative susceptibility mapping in multiple sclerosis. *Sci Rep.* **12**(1), 4411 (2022).
- Zivadinov, R. *et al.* Brain iron at quantitative MRI is associated with disability in multiple sclerosis. *Radiology.* **289**(2), 487–496 (2018).
- Wiggermann, V. *et al.* Susceptibility-sensitive MRI of multiple sclerosis lesions and the impact of normal-appearing white matter changes. *NMR Biomed.* **30**(8) (2017).
- Sibgatulin, R. *et al.* Magnetic susceptibility anisotropy in normal appearing white matter in multiple sclerosis from single-orientation acquisition. *Neuroimage Clin.* **35**, 103059 (2022).
- Abbasian Ardakani, A., Bureau, N. J., Ciaccio, E. J. & Acharya, U. R. Interpretation of radiomics features - a pictorial review. *Comput Methods Programs Biomed.* **215**, 106609 (2022).
- Guiot, J. *et al.* A review in radiomics: making personalized medicine a reality via routine imaging. *Med Res Rev.* **42**, 426–440 (2022).
- Gillies, R. J., Kinahan, P. & Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology.* **278**(2), 563–577 (2016).
- Kim, J. Y. *et al.* Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. *Neuro Oncol.* **21**(3), 404–414 (2019).
- Li, G. *et al.* An MRI radiomics approach to predict survival and tumor-infiltrating macrophages in gliomas. *Brain.* **143**(3), 1151–1161 (2022).
- Kickingereder, P. & Andronesi, O. C. Radiomics, metabolic, and molecular MRI for brain tumors. *Semin Neurol.* **38**(1), 32–40 (2018).
- Fiscone, C. *et al.* MRI dataset for susceptibility-based radiomic feature extraction in healthy controls and patients with multiple sclerosis [Dataset] Zenodo <https://doi.org/10.5281/zenodo.10931120> (2024).
- Fiscone, C. *et al.* Assessing robustness of quantitative susceptibility-based MRI radiomic features in patients with multiple sclerosis. *Sci Rep.* **13**(1), 16239 (2023).
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage.* **62**(2), 782–790 (2012).
- Jenkinson, M. & Smith, S. A global optimization method for robust affine registration of brain images. *Medical Image Analysis.* **5**(2), 143–156 (2001).
- Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *Neuroimage.* **17**(2), 825–841 (2002).
- Schweser, F., Deistung, A., Lehr, B. W. & Reichenbach, J. R. Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: An approach to *in vivo* brain iron metabolism? *Neuroimage.* **54**(4), 2789–2807, <https://doi.org/10.1016/j.neuroimage.2010.10.070> (2011).
- Li, W., Wu, B. & Liu, C. Quantitative susceptibility mapping of human brain reflects spatial variation in tissue composition. *Neuroimage.* **55**(4), 1645–1656 (2011).
- Smith, S. M. Fast robust automated brain extraction. *Human Brain Mapping.* **17**(3), 143–155 (2002).
- Biondetti, E. *et al.* Multi-echo quantitative susceptibility mapping: how to combine echoes for accuracy and precision at 3 Tesla. *MR Med.* **88**(5), 2101–2116 (2022).
- Li, W. *et al.* A method for estimating and removing streaming artifacts in quantitative susceptibility mapping. *Neuroimage.* **108**, 111–222 (2015).
- Li, W., Wu, B. & Liu, C. STI Suite. A software package for quantitative susceptibility imaging. *ISMRM-ESMRMB* <https://archive.ismrm.org/2014/3265.html> (2014).
- Wang, Y. & Liu, T. Quantitative susceptibility mapping (QSM): decoding MRI data for a tissue magnetic biomarker. *Magn Reson Med.* **73**(1), 82–101 (2015).
- Straub, S. *et al.* Suitable reference tissues for quantitative susceptibility mapping of the brain. *Magn Reson Med.* **78**(1), 204–214 (2017).
- Talozzi, L. *et al.* Along-tract analysis of the arcuate fasciculus using the Laplacian operator to evaluate different tractography methods. *Magn Reson Imaging.* **54**, 183–193 (2018).
- Zoli, M. *et al.* From neurosurgical planning to histopathological brain tumor characterization: potentialities of arcuate fasciculus along-tract diffusion tensor imaging tractography measures. *Front Neurol.* **12**, 633209 (2021).
- Tournier, J. D. *et al.* MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage.* **202**, 116–137 (2019).
- Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage.* **125**, 1063–1078 (2016).

39. Tournier, J. D., Calamante, F. & Connelly, A. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage*. **35**(4), 1459–72 (2007).
40. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. **33**(3), 341–355 (2002).
41. Schmidt, P. Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging. *PhD thesis, Ludwig-Maximilians-Universität München* <https://edoc.ub.uni-muenchen.de/20373/> (2017).
42. van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**(21), e104 (2017).
43. Bischoff-Grethe, A. *et al.* A technique for the deidentification of structural brain MR images. *Hum Brain Mapp.* **28**(9), 892–903 (2007).
44. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. **23**(Suppl1), S208–19 (2004).
45. Zwanenbrug, A. *et al.* The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. **295**(2), 328–338 (2020).
46. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* **69**(3), 89–95 (2001).
47. Smith, E. T. S. Clinical applications of imaging biomarkers. Part 1. The neuroradiologist's perspective. *Br J Radiol.* **84**(Spec Iss 2), S196–204 (2011).
48. Zwanenbrug, A. *et al.* Assessing robustness of radiomic features by image perturbation. *Sci Rep.* **9**(1), 614 (2019).

## Acknowledgements

This work was supported by national funds through the FCT (Fundação para a Ciência e a Tecnologia) by the project UIDB/04152/2020 (DOI: <https://doi.org/10.54499/UIDB/04152/2020>) - Centro de Investigação em Gestão de Informação – MagIC/NOVA IMS. This work was supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006)—A multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022). The publication of this article was supported by the “Ricerca Corrente” funding from the Italian Ministry of Health.

## Author contributions

All authors contributed to the work. C.F. and M.C. played a guiding role in designing, drafting and finalizing the manuscript, which was revised by D.N.M., C.T., R.L., A.L. and L.R.. Images processing pipeline was designed by C.F., G.S., G.V. and D.N.M., radiomic features extraction pipeline was designed by C.F., L.R. and M.C. Data were collected and selected from A.L. and I.P. L.M. evaluated image quality. C.F., L.R., F.Z. and M.C. designed and conducted the study illustrated to validate the dataset. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to D.N.M.

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

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



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

© The Author(s) 2024