Accelerated white matter lesion analysis based on simultaneous T_1 and T2* quantification using magnetic resonance fingerprinting and deep learning

Hermann, Ingo; Martínez-Heras, Eloy; Rieger, Benedikt; Schmidt, Ralf; Golla, Alena Kathrin; Hong, Jia Sheng; Lee, Wei Kai; Nagtegaal, Martijn; Weingärtner, Sebastian; More Authors

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Accelerated white matter lesion analysis based on simultaneous T_1 and T_2^* quantification using magnetic resonance fingerprinting and deep learning

Ingo Hermann[1,2] | Eloy Martínez-Heras[3] | Benedikt Rieger[1] | Ralf Schmidt[1] | Alena-Kathrin Golla[1,4] | Jia-Sheng Hong[5] | Wei-Kai Lee[5] | Wu Yu-Te[5,6] | Martijn Nagtegaal[2] | Elisabeth Solana[3] | Sara Llufriu[3] | Achim Gass[7] | Lothar R. Schad[1] | Sebastian Weingärtner[2] | Frank G. Zöllner[1,4]

[1]Computer Assisted Clinical Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2]Department of Imaging Physics, Delft University of Technology, Delft, the Netherlands

[3]Center of Neuroimmunology, Laboratory of Advanced Imaging in Neuroimmunological Diseases, Hospital Clinic Barcelona, Institut d'Investigacions Biomédiques August Pi i Sunyer (IDIBAPS) and Universitat de Barcelona, Barcelona, Spain

[4]Mannheim Institute for Intelligent Systems in Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[5]Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei, Taiwan

[6]Institute of Biophotonics and Brain Research Center, National Yang-Ming University, Taipei, Taiwan

[7]Department of Neurology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

Correspondence
Ingo Hermann, Computer Assisted Clinical Medicine, Medical Faculty Mannheim, University Heidelberg, Theodor Kutzer-Ufer 1-3, 68167, Mannheim, Germany.
Email: ingo.hermann@medma.uni-heidelberg.de

Purpose: To develop an accelerated postprocessing pipeline for reproducible and efficient assessment of white matter lesions using quantitative magnetic resonance fingerprinting (MRF) and deep learning.

Methods: MRF using echo-planar imaging (EPI) scans with varying repetition and echo times were acquired for whole brain quantification of T_1 and T_2^* in 50 subjects with multiple sclerosis (MS) and 10 healthy volunteers along 2 centers. MRF T_1 and T_2^* parametric maps were distortion corrected and denoised. A CNN was trained to reconstruct the T_1 and T_2^* parametric maps, and the WM and GM probability maps.

Results: Deep learning-based postprocessing reduced reconstruction and image processing times from hours to a few seconds while maintaining high accuracy, reliability, and precision. Mean absolute error performed the best for T_1 (deviations 5.6%) and the logarithmic hyperbolic cosinus loss the best for T_2^* (deviations 6.0%).

Conclusions: MRF is a fast and robust tool for quantitative T_1 and T_2^* mapping. Its long reconstruction and several postprocessing steps can be facilitated and accelerated using deep learning.

KEYWORDS

deep learning reconstruction, magnetic resonance fingerprinting, T_1 mapping, T_2^* mapping

1 | INTRODUCTION

White matter (WM) lesions are a common brain imaging finding in multiple sclerosis (MS) affecting the central nervous system. WM lesions are commonly characterized by increased T_1 and T_2^* relaxation times.[1] T_2-weighted imaging or fluid attenuated inversion recovery (FLAIR) is most commonly used in clinical MRI.[2] However, conventional magnetic resonance imaging (MRI) only provides limited insights into the pathological substrate of tissue changes (eg, axonal loss, inflammation, demyelination). Specifically, qualitative imaging inherently hampers standardization and reproducibility. Therefore, quantification of relaxation times, such as T_1, T_2, and T_2^*, is increasingly receiving interest for providing additional information beyond qualitative imaging.[3-5] However, most quantitative methods suffer from long acquisition times as the acquisition of multiple qualitative images is required. This renders quantitative MRI susceptibility to intra-scan motion. Furthermore, due to interscan motion and image distortion, multiple successive scans commonly need to be co-registered in order to allow for joint analysis.

Magnetic resonance fingerprinting (MRF) is a promising, time efficient approach for quantification of multiple tissue parameters in a single acquisition.[6] In MRF, characteristic magnetization evolutions are generated for tissues by varying sequence parameters including flip angle, echo time (TE), and repetition time (TR) throughout the acquisition. Thus, MRF has shown the potential to differentiate between healthy and pathological tissue and may, therefore, be useful for clinical MRI.[7] Rieger et al proposed an MRF sequence based on an echo-planar imaging (EPI) readout for simultaneous quantification of T_1 and T_2^* times covering the whole brain in less than 5 minutes.[8,9] Lower undersampling factors are applied compared with spiral MRF, reducing not only the noise per magnitude image but also the total number of magnitude images. This method was recently shown to provide clinically robust T_1 and T_2^* in neuro and renal applications.[8-10]

However, compared with other EPI scans high acceleration factors lead to a lower signal-to-noise ratio (SNR) than common for many clinical applications. Multiple denoising strategies have been proposed to improve the image quality and accuracy.[11-13] Recently, Marchenko–Pastur principal component analysis (MPPCA) was proposed to denoise EPI diffusion MRI images.[14] This is of particular interest, as a recent study demonstrated the value of denoising the acquired MRF magnitude images to improve the quality of the quantitative maps.[10] The large number of magnitude images in an MRF acquisition leads to long reconstruction times, which has been acknowledged as one of the drawbacks of the MRF methodology.[7,15] Additionally, several postprocessing steps hinder the practicability in clinical usage. In wake of recent developments, deep learning has superseded other approaches in many areas of data processing. Numerous publications have shown the benefits of using deep learning for medical imaging.[16-21] Specifically, deep learning accelerates processing steps and is capable of reconstructing MRI data.[22-24] Denoising plays an important role in MRI and several networks were evaluated to improve the visual image quality using generative adversarial networks and deep neuronal networks.[25-28] Furthermore, image synthesis has gained attention, which transforms a set of input images to a new set of image contrasts.[29,30] These image transformations can also contain deformable registration and artifact correction which showed good accuracy using CNN's.[31-33]

Especially for MRF, several models using fully connected neuronal network,[34,35] recurrent and convolutional neuronal network (CNN)[18,36-38] were analyzed showing promising results regarding the speed and accuracy of the reconstruction.[39] A deep learning reconstruction on MRF data using the spatiotemporal relationship between neighboring signal evolutions was proposed,[40,41] which showed an improvement in the reconstruction especially for undersampled complex MRF data. The U-Net has frequently been used to process medical data for segmentation and regression tasks.[42-45] Since most of the MRF acquisition techniques acquire a large number of highly undersampled images, the reconstruction problem is high dimensional. Therefore, a two-step deep learning approach was proposed in Ref. [19] to, first, reduce the dimensionality by using feature extraction with a fully connected network,[46] and second, a U-Net for spatially constrained quantification. The advantage of this learning-based model is that it contains tissue properties of the neighboring pixels which is more resilient to noise.[47]

In this study, we performed MRF-EPI for simultaneous quantification of T_1 and T_2^* in the whole brain on 50 patients with WM lesions and 10 healthy volunteers and analyzed the T_1 and T_2^* times in WM and gray matter (GM). Compared to conventional MRF methods, our MRF-EPI only slightly undersamples the k-space allowing for conventional parallel imaging reconstruction and yielding magnitude data that contains all relevant structural information. We developed a CNN for the MRF-EPI reconstruction of denoised and distortion corrected T_1 and T_2^* maps, and WM and GM probability maps. Furthermore, we compare different outputs, loss functions, and patches of the CNN for optimizing the entire reconstruction using deep learning.

2 | METHODS

This bicenter study was approved by the local institutional review board at both sites (2019-711N, BCB2012/7965), and written, informed consent was obtained prior to scanning. We performed MRF-EPI in 10 healthy volunteers (75% male, 22-30 (mean: 26) years, mean) and 18 patients (39% male, 23-73 (mean: 39) years) with MS on a 3T scanner (Magnetom

Skyra, Siemens Healthineers, Erlangen, Germany) at site 1 and in 32 patients (37% male, 1-63 (mean: 41) years) with MS at a 3T scanner (Magnetom Prisma, Siemens Healthineers, Erlangen, Germany) at site 2. Figure 1 depicts an overview of the MRF pipeline. The conventional steps (1-6) acquisition,

denoising, dictionary generation, reconstruction, distortion correction, and masking are depicted in the first part. The approach for standardization and acceleration using deep learning is shown in the second part, combining steps 2-6 to a single CNN.
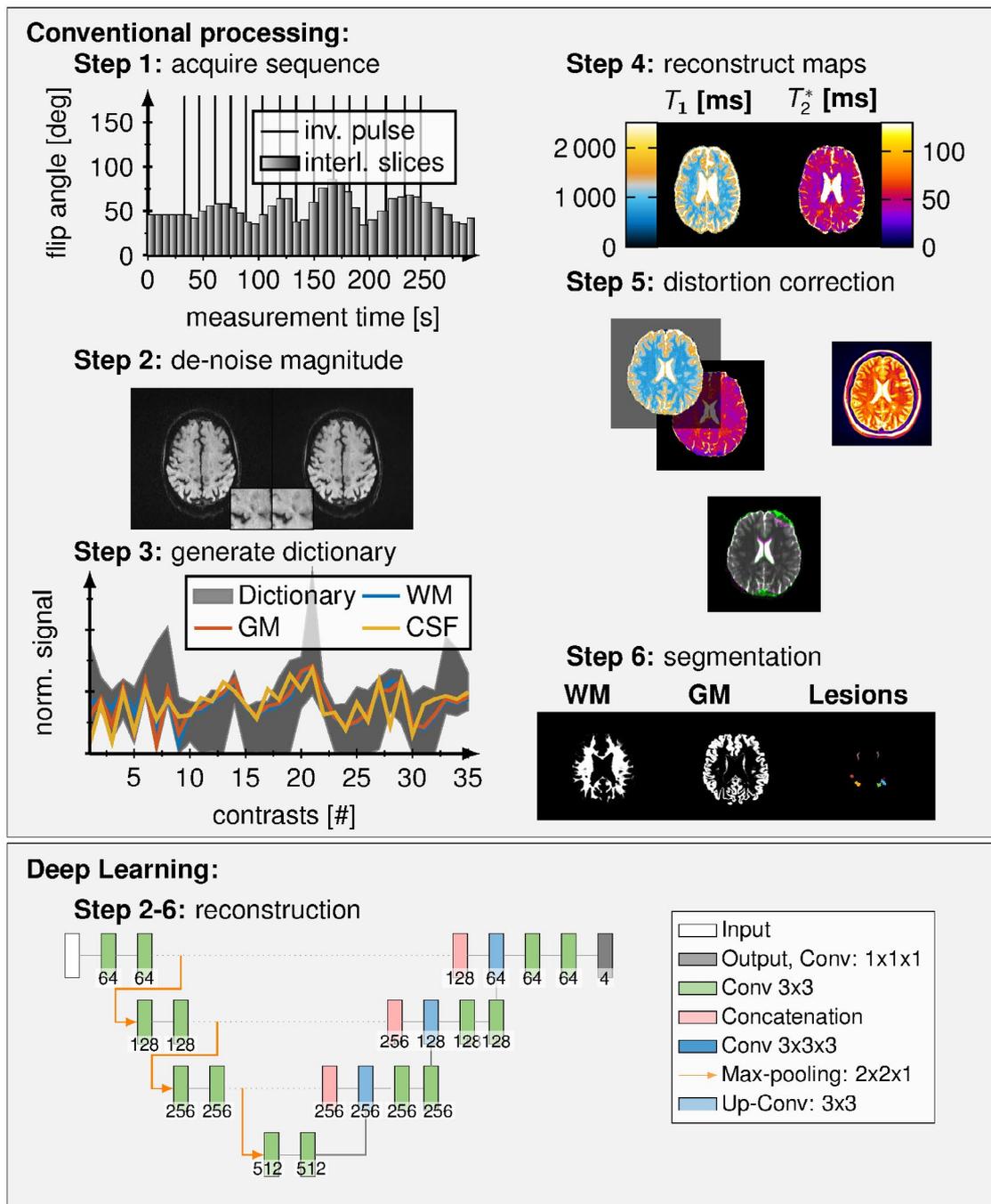
FIGURE 1 Schematic of the acquisition and postprocessing pipeline. **Step 1**: Varying flip angles, TE and TR with inversion pulses are played for the MRF-EPI sequence. **Step 2**: Denoising the magnitude data by MPPCA denoising. **Step 3**: Generation of the dictionaries for all T_1 and T_2^*. **Step 4**: Voxel-wise matching to generate the parametric maps via simple dictionary matching. **Step 5**: Distortion correction of the MRF maps using a restricted nonlinear registration onto T_2 weighted image due to susceptibility artifacts of the EPI readout. **Step 6**: WM and GM segmentation using SPM12 on the MRF T_1 maps. Manual lesion segmentation of the T_1-FLAIR data. **Steps 2-6**: Deep learning as a tool to integrate all postprocessing steps in a single operation

2.1 | Magnetic resonance fingerprinting

The acquisition was based on the previously proposed MRF-EPI technique for which accuracy and precision to gold standard methods were already evaluated.[8] Dictionaries were generated per slice using MATLAB (The MathWorks; Natick, MA, USA) consisting of 131,580 entries with T_1 (30-4000 ms) in 5% steps, T_2^* (5-3000 ms) in 5% steps, and flip angle efficiency B1+ (0.65-1.35) in steps of 0.05.

Sequence parameters for the MRF sequence were in-plane spatial resolution $= 1 \times 1$ mm^2, slice thickness $= 2$ mm, bandwidth $= 998$ Hz/px, GRAPPA factor $= 3$, partial fourier $= 5/8$, variable flip angle ($34° - 86°$), TE (21-81.5 ms), TR (3530-6570 ms), and fat suppression. At site 2, additionally, simultaneous multislice (SMS) imaging was used with an acceleration factor of 3. The acquisition time for site 1 was 4 minutes and 23 seconds and 1 minute and 52 seconds at site 2 covering all 60 slices. Additionally, T_1-FLAIR and T_2-weighted images were acquired for lesions segmentation and distortion correction, respectively, using the same spatial resolution.

2.2 | Principal component analysis denoising

We used MPPCA[14] to denoise the magnitude data of the MRF acquisition before reconstruction. Originally, the denoising strategy was proposed to estimate a non-Gaussian distribution on diffusion MRI data. The noise is estimated in a local neighborhood by the eigenvalues of principal component analysis using the Marchenko–Pastur distribution.[14]

Quantitative T_1 and T_2^* maps were compared with and without denoising. Denoising was performed on a per slice basis using a 2-dimensional (2D) kernel. As we are not interested in the actual image contrast but in the absolute T_1 and T_2^* times, we use standard deviation to describe the noise in these values.

2.3 | Distortion correction

Distortion correction was performed to correct for susceptibility artifacts, especially around the nasal cavities.[48] Rigid registration was computed from the T_2-weighted data to the MRF-magnitude data followed by a restricted nonlinear registration along phase-encode direction from the magnitude to the T_2-weighted data using ANTs.[49] Distorted maps were then visually compared to the FLAIR and T_2-weighted images to ensure that all modalities are properly registered.

2.4 | Data processing

White matter lesions were segmented manually by an expert radiologist on the FLAIR images. WM and GM were

TABLE 1 Different parameters for all the networks compared in this work are listed here

Network parameters			
Networks	Input	Outputs	Loss function
1	Patches	Single T_1	MAE
2	Patches	Single T_2^*	MAE
3	Patches	T_1, T_2^*	MAE
4	Patches	T_1, T_2^*, WM, GM	MAE
5	Patches	T_1, T_2^*, WM, GM	MSE
6	Patches	T_1, T_2^*, WM, GM	LCL
7	patches	T_1, T_2^*, WM, GM	HL
8	Full	Single T_1	MAE
9	Full	Single T_2^*	MAE
10	Full	T_1, T_2^*	MAE
11	Full	T_1, T_2^*, WM, GM	MAE
12	Full	T_1, T_2^*, WM, GM	MSE
13	Full	T_1, T_2^*, WM, GM	LCL
14	Full	T_1, T_2^*, WM, GM	HL

automatically segmented using SPM12 (Statistical Parametric Mapping version 12)[50] using the T_1 Maps acquired with MRF after denoising and distortion correction. The probability maps generated by SPM12 were transformed into binary masks by using a threshold (80%). Masks were visually analyzed and manually segmented WM lesions were extracted from the WM and GM mask to improve accuracy.

2.5 | Deep convolutional neural network for MRF reconstruction

Our network was a modified U-Net[42] implemented in Matlab 2020a (The MathWorks; Natick, MA) using the Deep Learning Toolbox. The network architecture is displayed in Figure 1. The training was performed on a GPU (Tesla K40m, Nvidia, Santa Clara, CA) for approximately 1 day per network. As inputs, the 35 differently weighted MRF-EPI magnitude images were used. The generated output were the T_1 and T_2^* maps and WM and GM probability maps. A brain mask was applied to exclude background noise. Data of 5 patients were randomly selected for testing, while the remaining data was chosen for training (49 datasets) and validation (5 datasets). Data of 6 healthy volunteers were acquired without T_2-weighted images and therefore, they are excluded from training. Two patients from site 1 and 3 patients from site 2 were chosen for the testing set. The 2D network was trained on individual slices. We trained half the networks with the full input resolution (240×240 voxels) and the other half using 32 random

patches (64 × 64 voxels) per slice (Table 1). We evaluated the following 4 different loss functions (the reference value y_i, predicted value y_i^p and the number of values n): the mean square error (MSE)

$$MSE = \frac{\sum_{i=1}^{n} (y_i - y_i^p)^2}{n}, \qquad (1)$$

the mean absolute error (MAE)

$$MAE = \frac{\sum_{i=1}^{n} |y_i - y_i^p|}{n}, \qquad (2)$$

the Huber loss (HL), which is a combination of MAE and MSE

$$HL = \begin{cases} \frac{1}{2}(y_i - y_i^p)^2, & \text{for } |y_i - y_i^p| \leq \delta \\ \delta|y_i - y_i^p| - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases} \qquad (3)$$

and the logarithm of the hyperbolic cosine (LCL)

$$LCL = \frac{\sum_{i=1}^{n} \log(\cosh(y_i - y_i^p))}{n}, \qquad (4)$$

We used Adam for optimization with a learning rate of 0.0001, L2-Regularization of 0.0001, 50 training epochs, and batch size = 64 for all networks, which was empirically determined to be optimal. Additionally, we trained the networks using patches and the full input resolution and MAE with 3 different types of outputs: (1) the network was trained with a single output once for T_1 and another for T_2^* (single); (2) the network was trained with both T_1 and T_2^* in a single network (dual); and (3) the network was trained with 4 output maps T_1, T_2^*, WM, and GM probability maps (4 outputs). Relative differences between dictionary matched and predicted maps were calculated and correlation coefficient of mean T_1 and T_2^* times between prediction and reference in WM, GM, lesions, and the whole brain was calculated. Reconstructions were executed on the CPU (Intel(R) Core(TM) i5-6500 @ 3.20 GHz).

2.6 | Statistics

Mean T_1 and T_2^* times with standard deviations were calculated and pair-wise comparison was performed using Student's t-tests and correlation R-values. P-values less than 0.05 were considered significant. The mean Dice similarity coefficient was used as a statistical validation metric for the predicted WM and GM probability maps after binarizing them into logical masks.

Computational time was measured using a standard desktop PC.

3 | RESULTS

The first part of this section presents the results of the conventional methods for denoising, distortion correction, and dictionary-based reconstruction and analysis. The second part presents the comparison to the deep learning-based reconstruction and analysis.

3.1 | Conventional reconstruction

Image denoising was successfully performed using MPPCA and resulted in up to 50% decreased variability in the magnitude data and 15% reduced standard deviations of T_1 and T_2^* (Supporting Information Figure S1). Overall, denoising the MRF magnitude data took about 10 minutes per subject on a standard CPU.

After the denoising and reconstruction of the parametric T_1 and T_2^* maps, EPI distortion correction was performed as exemplarily shown in Supporting Information Figure S2. Deviations in the relaxation times of up to 10% were observed in caudal slices next to the nasal cavities after applying the distortion correction. Distortion corrected mean T_1 and T_2^* times show only minor variations (<2%) in WM, GM, and WM lesions compared with T_1 and T_2^* times without distortion correction. The distortion correction of the MRF data takes around 1 hour for one whole brain on a standard CPU.

Representative T_1 and T_2^* maps including annotations are shown in Figure 2 for both sites. Reconstruction of the parametric maps using a pattern-matching algorithm took around 20 minutes per subject. Mean T_1 and T_2^* relaxation times for WM, GM, and WM lesions are depicted in Figure 3 and provided in the Supporting Information Tables S1-S3. Differences between healthy and diseased subjects from both sites were less than 4% for T_1 and less than 2% for T_2^* in WM and less than 7% for T_1 and less than 3% for T_2^* in GM. MRF acquired in site 2 had 15% higher standard deviations in T_1 and T_2^* due to increased scan time acceleration. Mean T_1 relaxation times in WM lesions are widespread ranging from 800 ms, comparable to WM, up to 2500 ms. Mean T_2^* times in WM lesions were consistently higher (70%) than WM and GM with mean T_2^* times up to 200 ms.

Clear separation between WM and GM was found in T_1 (Figure 3). We found a slight trend of increasing T_2^* (up to 10%) in WM and GM for increasing slice position ($R = 0.974$, $P < .0001$; Figure 3C). T_2^* was shorter and had higher standard deviations in caudal slices in the vicinity to the nasal cavities. No significant increase in T_1 and T_2^* with either age
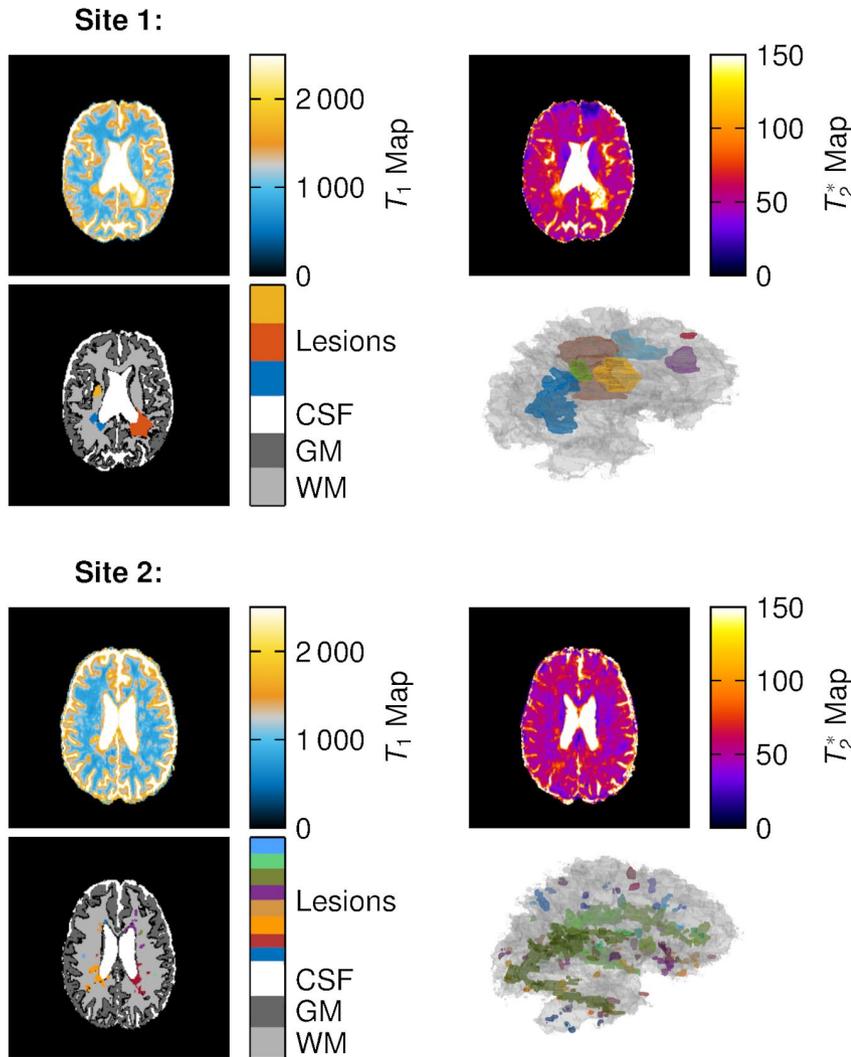
Site 1:

Site 2:

FIGURE 2 Representative T_1 and T_2^* map for 1 patient of site 1 (top) and 1 patient of site 2 (bottom). The segmented WM, GM, and CSF are shown combined with the manually segmented WM lesions. WM is shown semitransparent in 3D with the segmented lesions of representative different subjects from both sites. Different colors depict different lesions as they were automatically separated

or gender was observed (Figure 3D). T_1 and T_2^* times in WM lesions were highly heterogeneous and independent of their localization and size ($P > .2$).

3.2 | Deep convolutional network for MRF reconstruction

The computation time of the proposed CNN for 60 slices was about 5 seconds on a standard CPU workstation.

The performance of the reconstruction during the training process is depicted in Figure 4. Already after 5 epochs, the reconstructed maps have a visual good agreement with the dictionary-matched maps.

Figure 5 shows the 2D histogram of a representative slice in 1 subject for the CNN predicted T_1 and T_2^* times over the dictionary matching. The relative difference showed major noise with few anatomical structures and mean deviations of less than 6% for T_1 and T_2^*. Variations in the CSF are increased as seen around the ventricle and at the skull. T_1 and T_2^* times,

which exceed 3000 ms are cut, and therefore, the ventricle has variations of 0%.

The average correlation coefficient R and the relative difference for T_1 and T_2^* were calculated for different loss function and outputs (Figure 6). The P-value for all correlations was $P < .001$. For the 4 output models, the smallest relative difference for T_1 was observed when using the MAE with deviations of 5.8% in the whole brain and for T_2^* using the LCL with 6.0% deviations in the whole brain. Correlation coefficients in the whole brain were more than 0.99 except for the MSE (0.989) for T_1 and higher than 0.985 for T_2^* in the whole brain. The relative difference in T_1 and T_2^* was observed to be the highest in GM. All the relative differences and correlation coefficient are given in Supporting Information Table S4 for T_1 and Supporting Information Table S5 for T_2^*.

The difference in the different loss functions is visually depicted in Figure 7, where the MSE smooths the predicted maps the most as clearly seen in the WM and GM probability maps. The HL has increased T_2^* in WM and the MSE decreased T_1 in WM. In the WM probability maps, the LCL
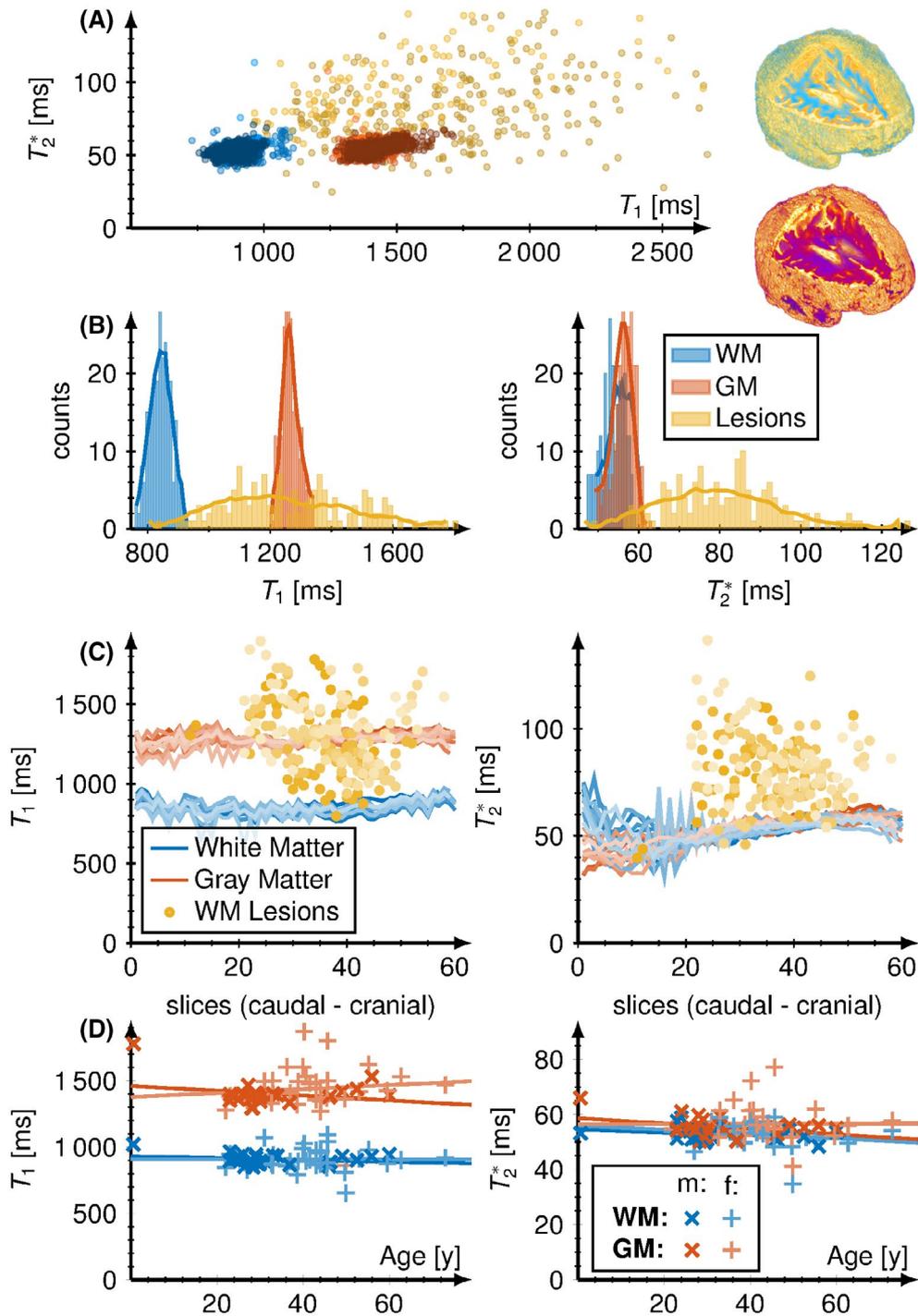
FIGURE 3 **A**, Mean T_2^* times over mean T_1 times for white matter (blue), gray matter (orange), and WM lesions (yellow) of all patients and subjects from both sites. Representative 3D T_1 and T_2^* maps were depicted on the right. **B**, Representative distributions of the T_1 and T_2^* times from A, which shows a much wider spread for the WM lesions considering T_1 and T_2^* times compared with WM and GM. In **C**, the mean T_1 (left) and T_2^* (right) over the slice position for white matter (blue) and gray matter (orange) are depicted. Color brightness encode different subjects. In **D**, the WM and GM T_1 and T_2^* times over the age and gender are shown

visually performed the best as seen in the prediction around the lesion.

The training with full image input showed significant increases in the relative error (25.8% for T_1 and 21.6% for T_2^*) and correlation coefficients of less than 0.90 for T_2^* in WM. Prediction in the WM performed better than in GM

with around 4% higher correlation coefficient and correlation coefficients in WM lesions were observed to be higher than 0.98. The mean Dice coefficient across the test data for WM was 0.9 and for GM 0.91 after conversion into logical masks with a threshold of 80% for both SPM and DL probability maps (Table 1, network 4). Dice coefficients
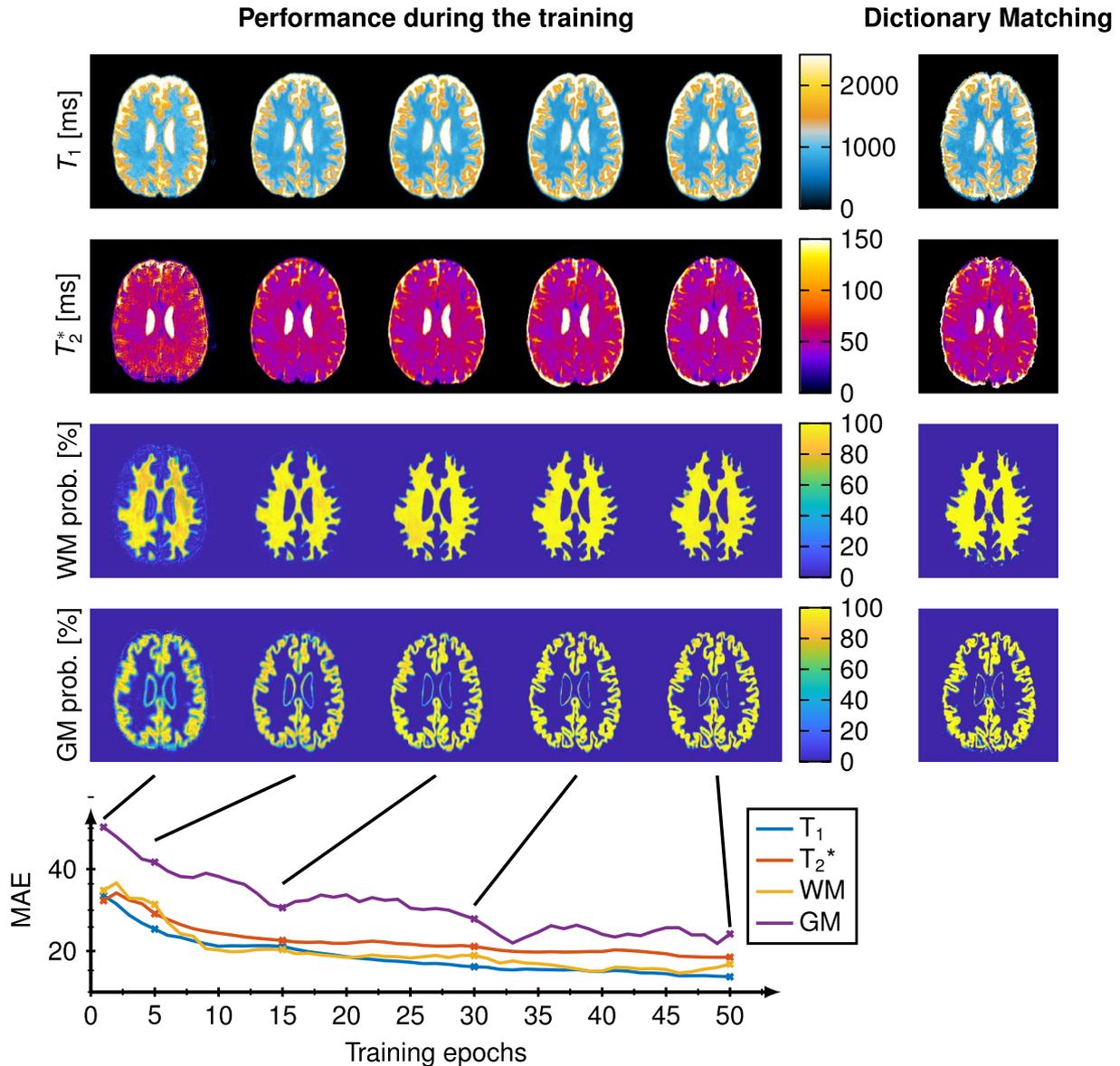
FIGURE 4 Visualization of the reconstruction during the training for 1 subject of the test data. The reconstructed T_1, T_2^*, WM, and GM probability maps are depicted for 1, 5, 15, 30, and 50 training epochs, and the dictionary matching reference maps are shown on the right side. On the bottom, the MAE is depicted over the 50 training epochs

decreased up to 15% when training was performed on the full input size without patches (mean WM: 0.81, mean GM: 0.79).

When training on T_1 and T_2^* as a dual output, prediction showed a slightly increased correlation coefficient (around 1%) and decreased relative difference compared with the 4 output models. Single T_1 and single T_2^* as outputs reached the highest correlation coefficients and smallest relative error among all other networks.

Figure 8 shows the mean T_1 and T_2^* times per subject between the DL and conventional reconstruction. A linear fit shows the correlation which was above 0.99 with $P < .0001$ for both T_1 and T_2^*. The bright colored markers depict the test data, which are aligned to the linear fit. We

observe a small offset in T_1 (55 ms) and in T_2^* (2.2 ms), which is within the standard deviations (100-200 ms $\sim [10]\%$, 3-5 ms $\sim 10\%$).

Figure 9 depicts the Dice coefficient between the WM and GM masks generated from the probability maps using SPM and our DL approach for different thresholds. The black line depicts the highest dice coefficients with close correlation to a straight line with a correlation coefficient of 0.9965 for WM and 0.9974 for GM with both $P < .0001$. For a commonly used threshold of 80% for SPM, the Dice coefficient is shown for different thresholds of the DL WM and GM maps. For a threshold of 80% of the DL reconstruction, the mean dice coefficient yields for both, WM and GM values of higher than 0.9.
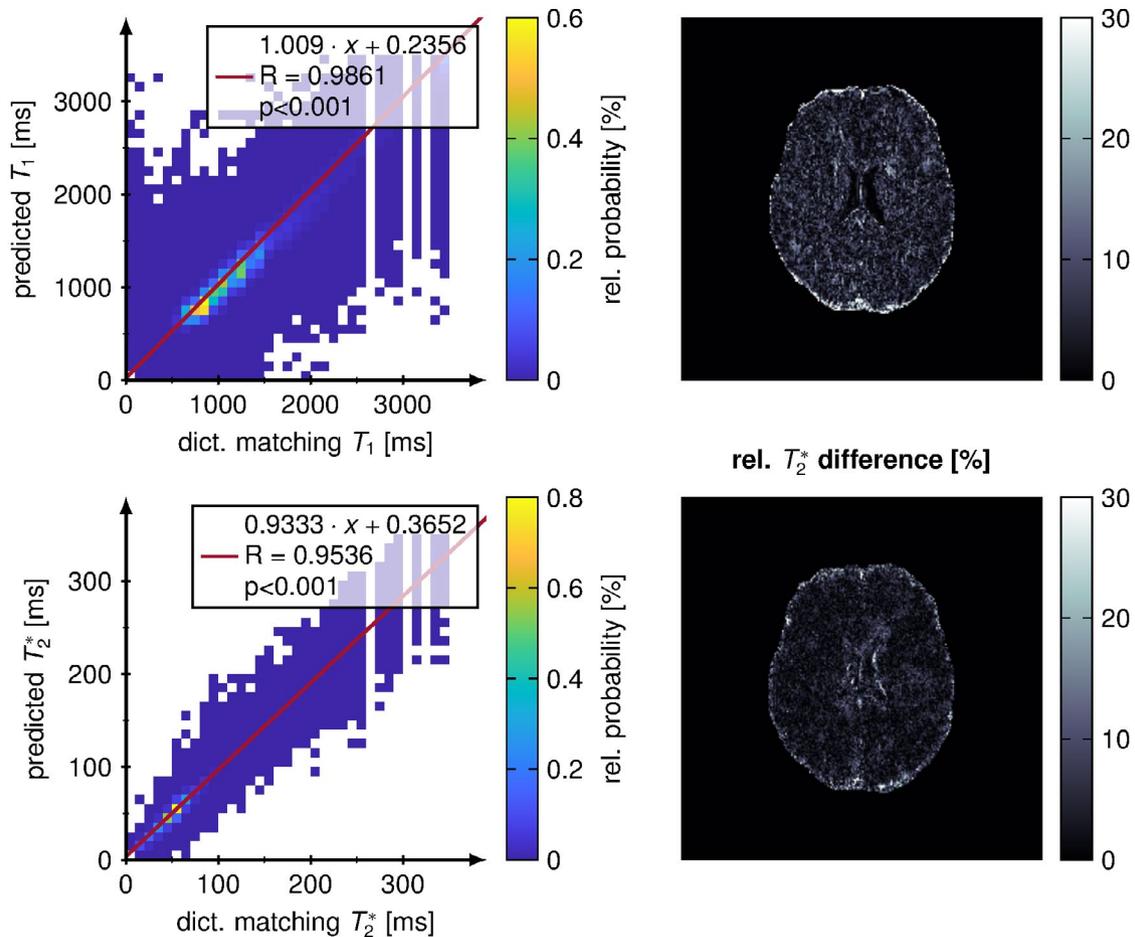
rel. T_1 difference [%]

rel. T_2^* difference [%]

FIGURE 5 Prediction of the CNN-network for 1 slice of a representative subject. The histograms (left panel) depict the predicted T_1/T_2^* (top/bottom) of 1 slice over the T_1/T_2^* generated by dictionary matching. The linear fit (red) with corresponding fit parameters and R and P-values is shown. On the right side, the relative difference of T_1 and T_2^* is shown between the predicted and dictionary matched parametric maps. Voxel-wise differences range up to 30% around the ventricles, because of the very high T_1 and T_2^* times for the CSF rendering the prediction difficult for the network

4 | DISCUSSION

We acquired MRF-EPI for simultaneous quantification of T_1 and T_2^* times in the whole brain. With a single convolutional neural network, we accelerated and combined several postprocessing steps as reconstruction, denoising, distortion correction, and masking.

MRF-EPI is a promising technique for quantification of T_1 and T_2^* of the whole brain in less than 5 minutes. T_1 and T_2^* times showed overall good agreement with literature.[1,3,5,51-55] However, as previously noted MR relaxation times for WM and GM show wide variability among studies due to different sequences, fitting procedures and natural variability among subjects.[51] Accuracy and precision measurements for the proposed MRF-EPI sequences were performed in previous work and therefore not analyzed in this study.[8,10] WM lesions exhibit a wide range of T_1 and T_2^* relaxation times. The relaxation times were independent of their localization

and size in the brain. WM lesions were successfully delineated from WM, GM, and CSF based only on quantitative MRF T_1 and T_2^* maps. Lesions which are difficult to separate from CSF on conventional images show a clear difference in the T_1 and T_2^* maps acquired with MRF due to long T_1 times in CSF of around 3000-4000 ms compared with T_1 times in lesions of around 1000-2000 ms. These high and widespread ranges of T_1 times in lesions might be due to altered interstitial fluid mobility and water content from edematous brain tissue.[56] Thus, the use of quantitative relaxometry obtained by MRF might potentially enhance the segmentation around the CSF. A fraction of WM lesions exhibit only a slight elevation of the T_1 times compared with WM and, therefore, yield similar or even smaller values compared with GM. This hampers the separation of WM lesions and GM. However, the additional assessment of T_2^* proved to be beneficial for the assessment of those lesions and showed improved separation against GM. The increased sensitivity
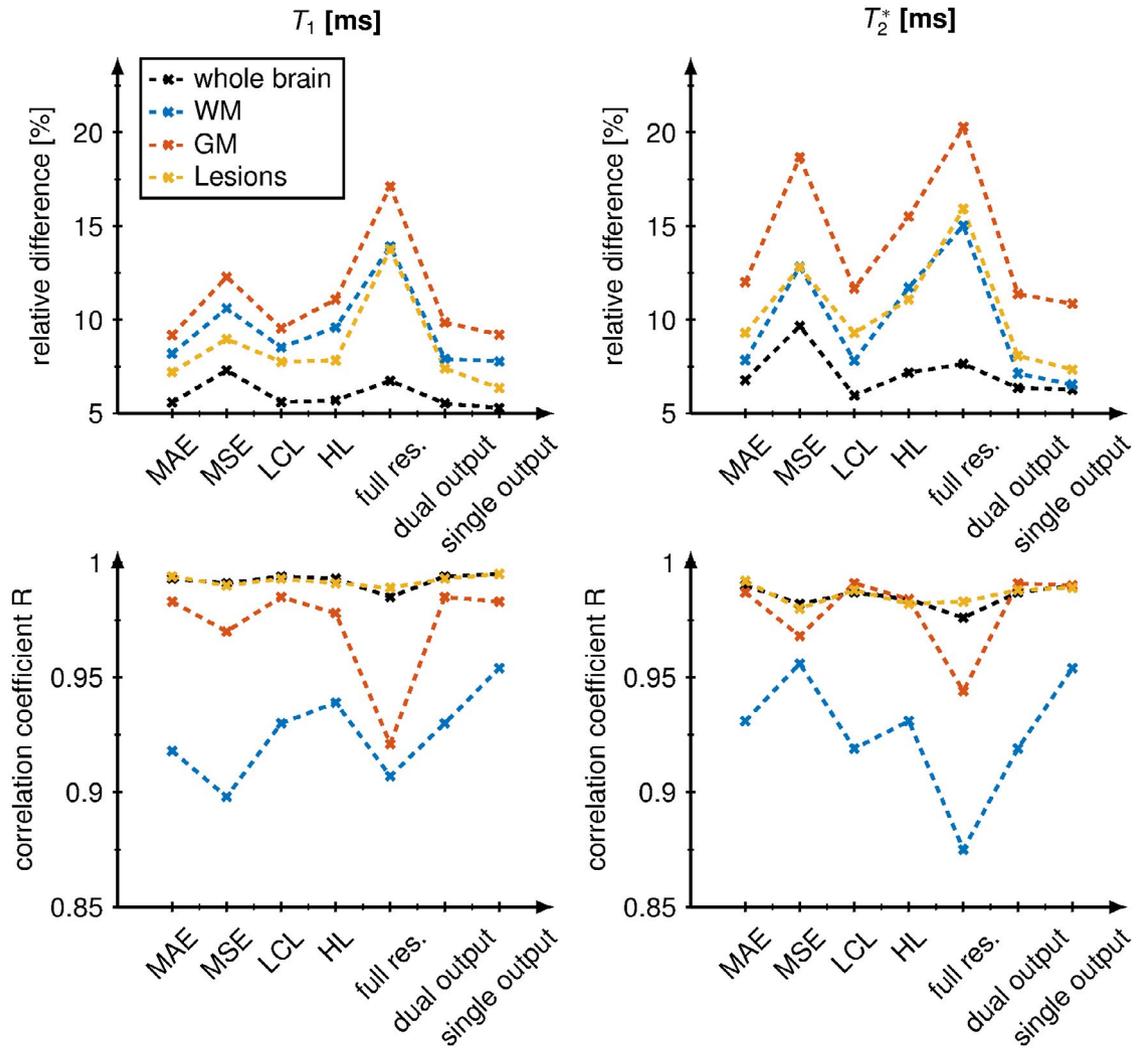
FIGURE 6 Relative difference between the predicted and dictionary matched T_1 (left) and T_2^* (right) for the whole brain (black), WM (blue), GM (orange), and lesions (yellow) compared to the different loss functions and network outputs. The first 4 data points (mean absolute error [MAE], mean squared error [MSE], logarithmic cosinus loss [LCL], Huber loss [HL]; Table 1, networks 4-7) are the networks trained with patches and 4 outputs. The fifth one (Table 1, networks 11) is trained with the full input resolution (full res.) and the MAE. The last 2 (dual output and single output; Table 1, networks 3 and 1+2) are trained using patches and the MAE loss function with 1 and 2 output maps, respectively. On the bottom correlation coefficients for the linear fit between predicted and dictionary matched T_1 (left) and T_2^* (right) is shown for the different network outputs using the MAE, MSE, LCL, and HL

in T_2^* might be explained by the fact that T_2^* times in WM and GM yield similar values, and hence, deviations in T_2^* in lesions benefit delineation against both WM and GM. This is a gain compared to conventional methods such as FLAIR or T_2-weighted images. We found no significant increase in T_1 and T_2^* with either age or gender, although a number of studies demonstrated that T_1 does change with age.[57,58] This might be due to the smaller number of subjects since we split between the healthy and diseased subjects and the narrow age range, especially for the healthy subjects. Further analysis of this might be performed when more subjects are measured. Only minor differences between data from site 1 and site 2 were observed (<7%), with no significant trends ($P > .2$). This demonstrates the potential of MRF as a quantitative method that is suitable for reproducible multicenter

studies and a pathway to standardization. A slight trend of increasing T_2^* was identified in the cranial direction. This is unlikely to be a result of the acquisition scheme, as due to the slice interleaving any inaccuracies would be expected to appear interleaved as well. Instead, this effect might be explained by increasing B0 inhomogeneities in the axial direction. In site 2, additional scan time acceleration was achieved with SMS factor 3, reducing the effective scan time by a factor of 2-3. However, the use of SMS acceleration inflicts an additional drop in SNR depending on the G-factor due to the coil geometry. Accordingly, the quantitative data was found to have increased standard deviations of up to 15% compared with data from site 1. This might be improved by extending the acquisition scheme when using SMS or by using regularized SMS reconstructions.[59]
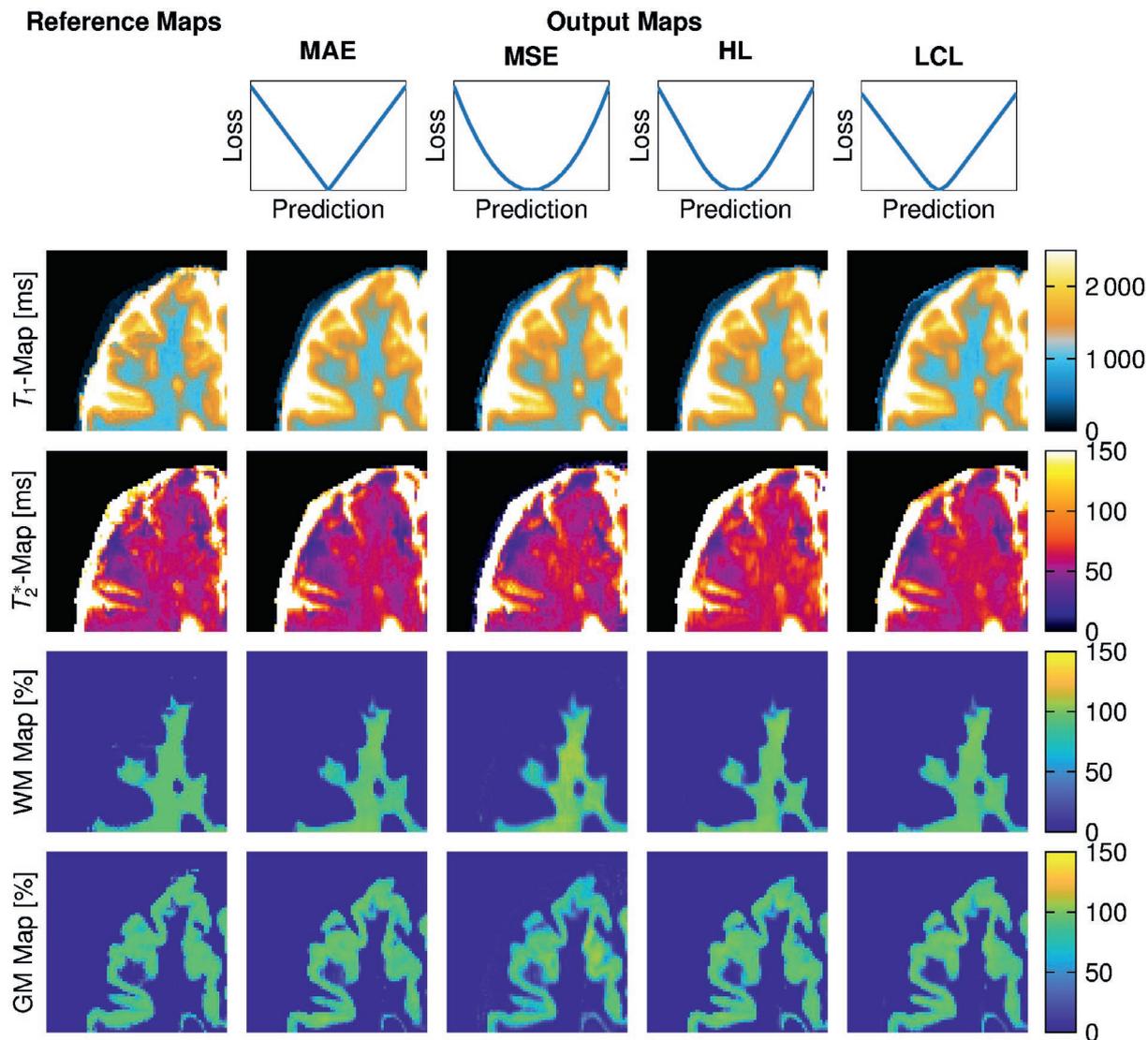
FIGURE 7 Comparison of the different loss functions to the dictionary matched input of 1 representative subject of the test data using the network 4 from Table 1. T_1, T_2^*, WM and GM probability maps are shown for the mean absolute error (MAE), mean squared error (MSE), logarithmic cosinus loss (LCL), and Huber loss (HL). A small patch (71×71) of 1 slice of a representative subject is shown. It is seen that the mean squared loss is smoothing the WM and GM probability maps the most

Our deep learning-based reconstruction yielded only minor differences between the T_1 and T_2^* times of WM, GM, and WM lesions compared with conventional dictionary matching. These mean deviations of 5.8% for T_1 and 6.0% for T_2^* are small and in the range of different approaches (2-8%).[19,36] Of note is that there is no ground truth data and, therefore, the dictionary matched data is the reference with a precision of 5%. Our deep learning approach is in the area of this precision and might be more precise since the output is continuous for all parameters. However, the deep learning reconstruction time was around 5 seconds for all slices as compared to 20 minutes dictionary matching, 10 minutes denoising, and 1-hour distortion correction (90 minutes in total).

We trained our networks with different loss functions and found that the MAE and LCL performed better regarding our regression task compared to the commonly used MSE function.[18,35] This might be due to the fact that in the MSE the CSF is weighted higher as it has longer T_1 and T_2^* times, and therefore, it is more difficult for the network to learn the relatively small differences in WM and GM. Since the T_1 and T_2^* times in the CSF are not of great clinical interest we accept the loss in accuracy for the CSF. The Dice coefficient for WM and GM was in the range of reported literature (0.82-0.93)[45,60,61] and in the range of SPM (0.76-0.83)[62,63] and above 0.87 for all loss functions if the training was performed with patches. This might be explained by the fact that data augmentation (random patch extraction) prevents overfitting and enriches the dataset. Overall improved performance was observed for the training using patches independent of the loss function and the output. We found an overall 25.8% decreased relative error for T_1 and 21.6% for T_2^* respectively.
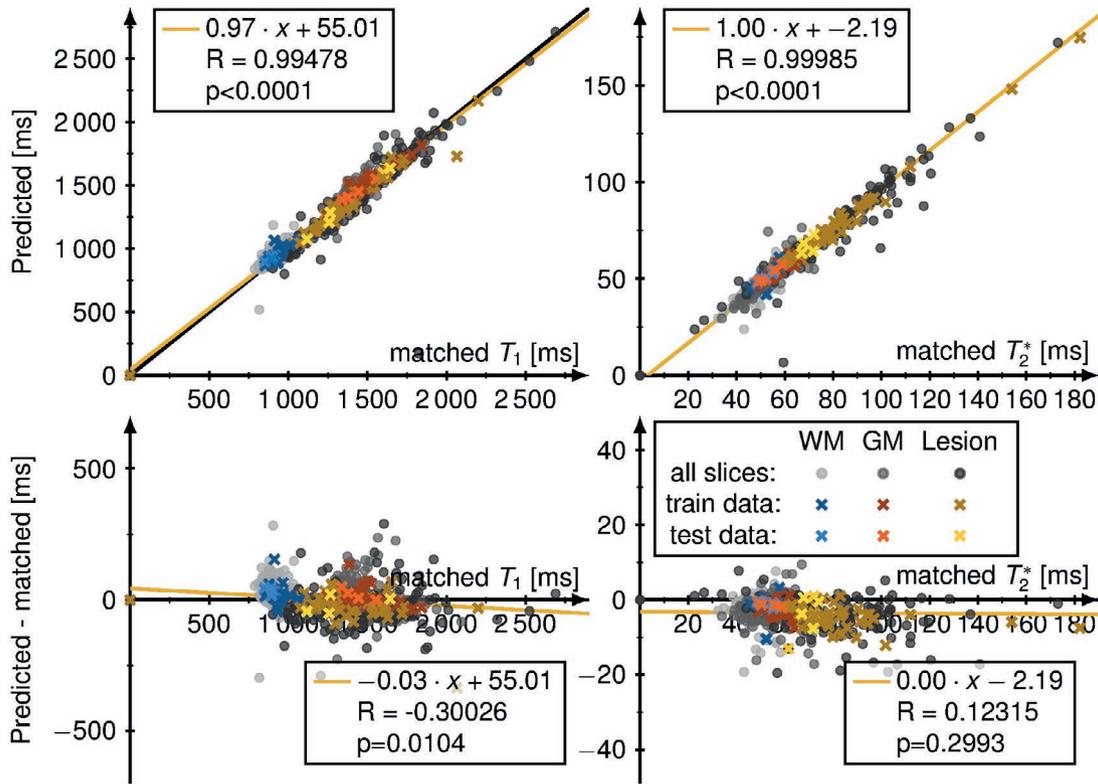
FIGURE 8 Predicted T_1 (left) and T_2^* (right) times over the dictionary matched T_1 and T_2^* times for the 4 output networks using patches for training and the mean absolute loss. Mean values per subject of WM are shown in blue, of GM are shown in orange and for the lesion are shown in yellow. The increased brightness of the representative colors depicts the test data and the reduced brightness depicts the training and validation data. In 3 different gray shades, the single T_1 and T_2^* times per slice are shown. A linear fit is used to correlate the predicted and the dictionary matched quantitative maps with corresponding R- and P-values

This might be due to the fact that training with the full input resolution takes longer to converge. Compared with conventional highly undersampled MRF acquisition in our MRF-EPI approach, we do not need to extract first the features and reduce the dimensionality of the network input as proposed in other MRF deep learning reconstruction approaches.[19,46,47] Since the anatomical structure is retained, the network has to solve an image to images regression task, which might have smaller computational requirements. Fang et al[19] used a U-Net after the dimensionality was reduced: for their dataset, 2304 time points were used compared to 35 time points for our MRF-EPI (66 times smaller). We also used the U-Net since it captures information of the input locally and globally. This is important since we also include denoising and distortion correction with the same and single network.

Our reconstruction task included denoising and distortion correction within the MRF reconstruction and therefore, training with patches (64 × 64 voxels) achieved better results since the observed distortion from the EPI readout is only local at the nasal cavities and the frontal lobe of the brain. We showed that it is possible to perform denoising, distortion correction, and MRF reconstruction with one network architecture with relative difference within the standard deviation of the quantitative parameters.

We were able to additionally generate the WM and GM probability maps as outputs with only slightly decreased accuracy of the test data considering T_1 and T_2^* in WM and GM. We have shown that the dice coefficient for the binarized WM and GM masks are in good correlation between our CNN and the reference SPM method. However, the network trained only on the T_1 and T_2^* maps (dual output) as an output performed better than the 4 output model. Using single T_1 and single T_2^* maps as an output performed the best with only minor improvement (<1%) compared to the dual output model. We compared the relative differences for different tissue types instead of using the RMSE as commonly used,[18,35] because outliers and variations of quantitative measures within single tissue types result in an overestimated error for a voxel-by-voxel comparisons, especially in the CSF.

We showed that the predicted values correlate very well with the reference dictionary matched values for T_1 and T_2^* ($R > .95$, $P < .0001$) with only a slight offset, which is within the standard deviation. The correlation coefficient was the lowest for only WM since the range of single WM T_1 and T_2^* times is denser compared with GM and especially compared with lesions as provided in Figure 3.

We achieved standardized results as we trained on data from both sites without significant differences between both
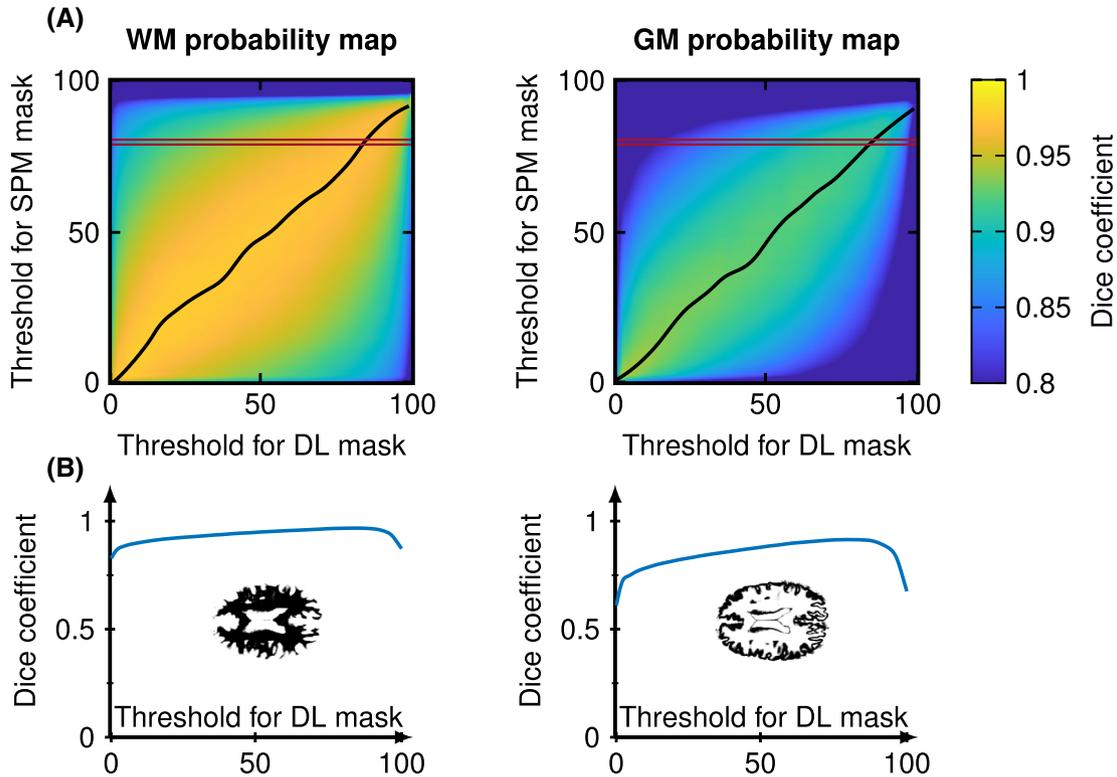
FIGURE 9 A, Dice coefficient for different thresholds of the SPM and DL (Deep Learning) WM (left) and GM (right) probability maps of a representative subject when using the MAE loss with 4 outputs and patches-wise training (Table 1, network 4). The black lines depict the maximum dice coefficient along with the different thresholds. The dice coefficient between both binary masks is shown in color encoding. **B**, The dice coefficient is shown for a fixed threshold of 80% of the SPM WM and GM masks dependent on the threshold of the DL mask, as marked in the red area A, which both show a maximum dice coefficient at around 80%. Exemplary WM and GM probability maps are depicted

$(P < .01)$, even though the magnitude data from both sites varies due to different accelerations. However, changing the sequence parameters changes the magnitude evolutions. Therefore, new dictionaries have to be calculated and different or retrained networks are required. Transfer learning may facilitate the possibility, to update the network when imaging parameters are changed.[24,64]

Our study has some limitations. As GM suffers from partial volume effects, calculating the mean T_1 and T_2^* times strongly depends on the segmentation and the used threshold on the probability maps. Lesion segmentation could be an extra output from a CNN similar to the one such as investigated here. However, to obtain reliable results from this, more WM lesions data would be required, due to the large variation in lesion tissue parameters and the small fraction of lesions compared to WM and GM. In our experiments, the training datasets did not provide enough lesion examples for the training to converge without significantly affecting other outputs. The strength of deep learning approaches commonly stems from the abundance of training data.[65,66] Therefore, the proposed reconstruction will likely benefit from larger datasets. Fractioning the full input into small patches is a first step to artificially generate more data, but data augmentation could be applied additionally. In this study, both sites operated on

the platform of a single MRI vendor. A multi-vendor study is required for more universal comparisons.

5 | CONCLUSIONS

MRF demonstrates to be an auspicious approach for quantifying T_1 and T_2^* in subjects with MS to obtain information in a standardized fashion along 2 clinical centers. This technique saves time by simultaneous acquisition of T_1 and T_2^* and might improve the segmentation pipeline of lesions as their quantitative measures are clearly separated from normal appearing brain tissue types. We showed that deep learning enables a drastic speed up in the postprocessing pipeline without a loss in accuracy and precision by combining denoising, distortion correction, reconstruction, and masking.

REFERENCES

1. Bonnier G, Roche A, Romascano D, et al. Advanced MRI unravels the nature of tissue alterations in early multiple sclerosis. *Ann Clin Trans Neurol*. 2014;1:423-432.
2. Trip SA, Miller DH. Imaging in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2015;76:iii11-iii18.

3. Blystad I, Håkansson I, Tisell A, et al. Quantitative MRI for analysis of active multiple sclerosis lesions without gadolinium-based contrast agent. *Am J Neuroradiol*. 2016;37:94-100.

4. Bauer S, Wagner M, Seiler A, et al. Quantitative T_2'-mapping in acute ischemic stroke. *Stroke*. 2014;45:3280-3286.

5. Hernández-Torres E, Wiggermann V, Machan L, et al. Increased mean R_2^* in the deep gray matter of multiple sclerosis patients: have we been measuring atrophy? *J Magn Reson Imaging*. 2019;50:201-208.

6. Ma D, Gulani V, Seiberlich N, et al. Magnetic resonance fingerprinting. *Nature*. 2013;495:187-192.

7. Panda A, Mehta BB, Coppo S, et al. Magnetic resonance fingerprinting—an overview. *Curr Opin Biomed Eng*. 2017;3:56-66.

8. Rieger B, Akçakaya M, Pariente JC, et al. Time efficient whole-brain coverage with MR fingerprinting using slice-interleaved echo-planar-imaging. *Sci Rep*. 2018;8:2045-2322.

9. Rieger B, Zimmer F, Zapp J, Weingartner S, Schad LR. Magnetic resonance fingerprinting using echo planar imaging joint quantification of T_1 and relaxation times. *Magn Reson Med*. 2017;78:1724-1733.

10. Hermann I, Chacon-Caldera J, Brumer I, et al. Magnetic resonance fingerprinting for simultaneous renal T_1 and mapping in a single breath-hold. *Magn Reson Med*. 2020;83:1940-1948.

11. Mohan J, Krishnaveni V, Guo Y. A survey on the magnetic resonance image denoising methods. *Biomed Signal Process Control*. 2014;9:56-69.

12. Nowak RN. Wavelet-based Rician noise removal for magnetic resonance imaging. *IEEEE Trans Image Process*. 1999;8:1408-1419.

13. Saladi S, Amutha PN. Analysis of denoising filters on MRI brain images. *Int J Imaging Syst Technol*. 2017;27:201-208.

14. Veraart J, Novikov DS, Christiaens D, Ades-aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. *Neuroimage*. 2016;15:394-406.

15. Bipin MB, Coppo S, Frances MD, et al. Magnetic resonance fingerprinting: a technical review. *Magn Reson Med*. 2019;81:25-46.

16. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2019;29:102-127. Special Issue: Deep Learning in Medical Physics.

17. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. *Z Med Phys*. 2019;29:86-101.Special Issue: Deep Learning in Medical Physics.

18. Hoppe E, Thamm F, Körzdörfer G, et al. Magnetic resonance fingerprinting reconstruction using recurrent neural networks. *Stud Health Technol Inform*. 2019;267:126-133.

19. Fang Z, Chen Y, Liu M, et al. Deep learning for fast and spatially constrained tissue quantification from highly accelerated data in magnetic resonance fingerprinting. *IEEE Trans Med Imaging*. 2019;38:2364-2374.

20. Akcakaya M, Moeller S, Weingärtner S, Ugurbil K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: database-free deep learning for fast imaging. *Magn Reson Med*. 2019;81:439-453.

21. Zhang C, Hosseini SAH, Weingärtner S, Ugurbil K, Moeller S, Akçakaya M. Optimized fast GPU implementation of robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction. *PLOS ONE*. 2019;14:1-14.

22. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055-3071.

23. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2018;37:491-503.

24. Knoll F, Hammernik K, Kobler E, Pock T, Recht MP, Sodickson DK. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med*. 2019;81:116-128.

25. Benou A, Veksler R, Friedman A, Riklin Raviv T. Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. *Med Image Anal*. 2017;42:145-159.

26. Bermudez C, Plassard AJ, Davis LT, Newton AT, Resnick SM, Landman BA. Learning implicit brain MRI manifolds with deep learning. *CoRR*. 2018.abs/1801.01847.

27. Manjon JV, Coupe P. MRI denoising using deep learning and non-local averaging. *ArXiv*. 2019. https://arxiv.org/abs/1911.04798. Accessed January 27, 2021.

28. You X, Cao N, Lu H, Mao M, Wang W. Denoising of MR images with Rician noise using a wider neural network and noise range division. *Magn Reson Imaging*. 2019;64:154-159.Artificial Intelligence in MRI.

29. Jog A, Roy S, Carass A, Prince JL. Magnetic resonance image synthesis through patch regression. In 2013 IEEE 10th International Symposium on Biomedical Imaging, San Francisco, CA. 2013:350-353.

30. Jog A, Carass A, Roy S, Pham DL, Prince JL. Random forest regression for magnetic resonance image synthesis. *Neuroimaging*. 2017;35:475-488.

31. Cao X, Yang J, Zhang J, Wang Q, Yap P, Shen D. Deformable image registration using a cue-aware deep regression network. *IEEE Trans Biomed Eng*. 2018;65:1900-1911.

32. Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Isgum I. *A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration*. 2018. https://arxiv.org/abs/1809.06130. Accessed January 27, 2021.

33. Ghosal S, Ray N. Deep deformable registration: enhancing accuracy by fully convolutional neural net. *Pattern Recogn Lett*. 2017;94:81-86.

34. Golbabaee M, Chen D, Gómez PA, Menzel MI, Davies ME. A deep learning approach for magnetic resonance fingerprinting. *ArXiv*. 2018.abs/1809.01749.

35. Cohen O, Zhu B, Rosen MS. MR fingerprinting deep reconstruction network (DRONE). *Magn Reson Med*. 2018;80:885-894.

36. Hoppe E, Körzdörfer G, Würfl T, et al. Deep learning for magnetic resonance fingerprinting: a new approach for predicting quantitative parameter values from time series. *Stud Health Technol Inform*. 2017;243:202-206.

37. Oksuz I, Cruz G, Clough J, et al. Magnetic resonance fingerprinting using recurrent neural networks. *IEEE Comput Soc*. 2019;1537-1540.

38. Chen Y, Fang Z, Hung S-C, Chang W-T, Shen D, Lin W. High-resolution 3D MR Fingerprinting using parallel imaging and deep learning. *NeuroImage*. 2020;206:116329.

39. Hamilton JI, Seiberlich N. Machine learning for rapid magnetic resonance fingerprinting tissue property quantification. *Proc IEEE*. 2020;108:69-85.

40. Balsiger F, Shridhar KA, Chikop S, et al. Magnetic resonance fingerprinting reconstruction via spatiotemporal convolutional neural networks. In F. Knoll, A. Maier, D. Rueckert (Eds.), *Machine Learning for Medical Image Reconstruction*; Cham: Springer International Publishing. 2018:39-46.

41. Balsiger F, Scheidegger O, Carlier PG, Marty B, Reyes M. On the spatial and temporal influence for the reconstruction of magnetic resonance fingerprinting, 102 of Proceedings of Machine Learning Research (London, United Kingdom). *PMLR*. 2019:27-38.

42. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9351 of LNCS*; Springer, 2015:234-241.(available on arXiv:1505.04597 [cs.CV]).

43. Ding PLK, Li Z, Zhou Y, Li B. Deep residual dense U-net for resolution enhancement in accelerated MRI acquisition. In E. D. Angelini, E. D. Angelini, E. D. Angelini, B. A. Landman (Eds.), *Medical Imaging 2019. Progress in Biomedical Optics and Imaging–Proceedings of SPIE, Hangzhou, China*. 2019.

44. Han C, Gesser F, Milacski Z. CNN-based MRI regression using u-net. 2018. https://www.semanticscholar.org/paper/CNN-based -MRI-Regression-Using-U-Net-Han-Gesser/c545b4c8f2cb5c5 dccf83bf3d5b3b49e17da8d29. Accessed January 27, 2021.

45. Moeskops P, de Bresser J, Kuijf HJ, et al. Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI. *NeuroImage*. 2018;17:251-262.

46. Golbabaee M, Chen D, Gómez PA, Menzel M, Davies M. Geometry of deep learning for magnetic resonance fingerprinting. ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK. 2019:7825-7829.

47. Fang Z, Chen Y, Hung S-C, Zhang X, Lin W, Shen D. Submillimeter MR fingerprinting using deep learning-based tissue quantification. *Magn Reson Med*. 2020;84:579-591.

48. Wang S, Peterson DJ, Gatenby JC, Li W, Grabowski TJ, Madhyastha TM. Evaluation of field map and nonlinear registration methods for correction of susceptibility artifacts in diffusion MRI. *Front Neuroinform*. 2017;11:17.

49. Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*. 2009;46:786-802.

50. Ashburner J, Balbastre Y, Barnes G, Brudfors M. SPM12. Wellcome Trust Centre for Neuroimaging. UCL. 2014. https://www.fil.ion.ucl. ac.uk/spm/software/spm12/. Accessed January 27, 2021.

51. Bojorquez JZ, Bricq S, Acquitter C, Brunotte F, Walker PM, Lalande A. What are normal relaxation times of tissues at 3 T? *Magn Reson Imaging*. 2017;35:69-80.

52. Feng X, Deistung A, Reichenbach JR. Quantitative susceptibility mapping (QSM) and R_2^* in the human brain at 3T: Evaluation of intra-scanner repeatability. *Z Med Phys*. 2018;28:36-48.

53. Gracien R-M, Maiworm M, Brüche N, et al. How stable is quantitative MRI? – assessment of intra- and inter-scanner-model reproducibility using identical acquisition sequences and data analysis programs. *NeuroImage*. 2020;207:116364.

54. Abbas Z, Gras V, Müllenhoff K, Keil F, Oros-Peusquens A-M, Shah NJ. Analysis of proton-density bias corrections based on T_1 measurement for robust quantification of water content in the brain at 3 Tesla. *Magn Reson Med*. 2014;72:1735-1745.

55. Baudrexel S, Volz S, Preibisch C, et al. Rapid single-scan T-mapping using exponential excitation pulses and image-based correction for linear background gradients. *Magn Reson Med*. 2009;62:263-268.

56. Wardlaw JM, Valdés Hernández MC, Munoz-Maniega S. What are white matter hyperintensities made of? *J Am Heart Assoc*. 2015;4:e001140.

57. Cho S, Jones D, Reddick WE, Ogg RJ, Steen RG. Establishing norms for age-related changes in proton T_1 of human brain tissue in vivo. *Magn Reson Imaging*. 1997;15:1133-1143.

58. Kupeli A, Kocak M, Goktepeli M, Karavas E, Danisan G. Role of T_1 mapping to evaluate brain aging in a healthy population. *Clin Imaging*. 2020;59:56-60.

59. Demirel ÖB, Weingärtner S, Moeller S, Akçakaya M. Improved Regularized reconstruction for simultaneous multi-slice cardiac MRI T_1 mapping. In 2019 27th European Signal Processing Conference (EUSIPCO), A Coruña, Spain. 2019:1-5.

60. Yogananda CGB, Wagner BC, Murugesan GK, Madhuranthakam A. Maldjian JA. A deep learning pipeline for automatic skull stripping and brain segmentation. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy. 2019:727-731.

61. Mahbod A, Chowdhury M, Smedby Ö, Wang C. Automatic brain segmentation using artificial neural networks with shape context. *Pattern Recognit Lett*. 2018;101:74-79.

62. Palumbo L, Bosco P, Fantacci ME, et al. Evaluation of the intra- and inter-method agreement of brain MRI segmentation software packages: a comparison between SPM12 and FreeSurfer v6.0. *Phys Med*. 2019;64:261-272.

63. Choi U-S, Kawaguchi H, Matsuoka Y, Kober T, Kida I. Brain tissue segmentation based on MP2RAGE multi-contrast images in 7 T MRI. *PLOS ONE*. 2019;14:1-13.

64. Dar SUH, Özbey M, Çatli AB, Çukur T. A transfer-learning approach for accelerated MRI using deep neural networks. *Magn Reson Med*. 2020;84:663-685.

65. Narayana PA, Coronado I, Sujit SJ, Wolinsky JS, Lublin FD, Gabr RE. Deep-learning-based neural tissue segmentation of MRI in multiple sclerosis: Effect of training set size. *J Magn Reson Imaging*. 2020;51:1487-1496.

66. Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *arXiv.org* [Internet]. 2015. https://scholar.harva rd.edu/synho/publications/how-much-data-needed-train-medical-image-deep-learning-system-achieve-necessary. Accessed January 27, 2021.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

FIGURE S1 Representation of magnitude images, T_1 and T_2^* maps before and after using Marchenko–Pastur principal component analysis (MPPCA) denoising. On the right side, the relative difference between acquired and denoised image and map is depicted. MPPCA (blue) compared with non-denoised (orange) 2D signals of 1 voxel is depicted over the different contrasts

FIGURE S2 Distorted (top) and corrected (middle) Maps (3 transversal slices and 1 sagittal slice) for T_1. Distortion correction was performed using ANTs (nonrigid deformation in phase-encoding direction only) onto the T_2-weighted images depicted as overlay, which shows in blue and red the deviations to the MRF-EPI before and after correction. The

difference in percentage before and after distortion correction is shown. Major improvements (areas marked by arrows and circles) are observed after distortion correction, especially around the nasal cavities and the frontal lobe

TABLE S1 Mean T_1 and T_2^* times for white matter (WM), gray matter (GM), and WM lesions are listed for all patients from site 1. The last column shows the respective mean values among the subjects and the intersubject variability. The amount of WM lesions per patient and patient's age and gender are provided

TABLE S2 Mean T_1 and T_2^* times for white matter (WM) and gray matter (GM) are listed for all healthy volunteers from site 1. The last column shows the respective mean values among the subjects and the intersubject variability. Subject's age and gender is provided

TABLE S3 Mean T_1 and T_2^* times for white matter (WM), gray matter (GM), and WM lesions are listed for all patients from site 2. The last column shows the respective mean values among the patients and the intersubject variability. The amount of WM lesions per patient and patient's age and gender are provided

TABLE S4 Mean absolute difference and correlation coefficients of T_1 for the different ne rks in the whole brain, WM, GM, and lesions. The loss functions mean absolute error

(MAE), mean squared error (MSE), logarithmic cosinus loss (LCL), Huber loss (HL) are used for the network with 4 outputs (T_1, T_2^*, WM, and GM). Additionally for MAE, the dual (T_1 and T_2^* as output) and single (combined single T_1 and single T_2^* as output) are listed. All networks are trained with patches and with the full input resolution (blue)

TABLE S5 Mean absolute difference and correlation coefficients of T_2* for the different networks in the whole brain, WM, GM, and lesions. The loss functions mean absolute error (MAE), mean squared error (MSE), logarithmic cosinus loss (LCL), and HL (Huber loss) are used for the network with 4 outputs (T_1, T_2^*, WM, and GM). Additionally, for MAE, the dual (T_1 and T_2^* as output) and single (combined single T_1 and single T_2^* as output) are listed. All networks are trained with patches and with the full input resolution (blue)