



OPEN

A deep learning approach for fast muscle water T₂ mapping with subject specific fat T₂ calibration from multi-spin-echo acquisitions

Marco Barbieri¹✉, Melissa T. Hooijmans², Kevin Moulin³, Tyler E. Cork¹, Daniel B. Ennis¹, Garry E. Gold^{4,5}, Feliks Kogan¹ & Valentina Mazzoli^{4,5}

This work presents a deep learning approach for rapid and accurate muscle water T₂ with subject-specific fat T₂ calibration using multi-spin-echo acquisitions. This method addresses the computational limitations of conventional bi-component Extended Phase Graph fitting methods (nonlinear-least-squares and dictionary-based) by leveraging fully connected neural networks for fast processing with minimal computational resources. We validated the approach through *in vivo* experiments using two different MRI vendors. The results showed strong agreement of our deep learning approach with reference methods, summarized by Lin's concordance correlation coefficients ranging from 0.89 to 0.97. Further, the deep learning method achieved a significant computational time improvement, processing data 116 and 33 times faster than the nonlinear least squares and dictionary methods, respectively. In conclusion, the proposed approach demonstrated significant time and resource efficiency improvements over conventional methods while maintaining similar accuracy. This methodology makes the processing of water T₂ data faster and easier for the user and will facilitate the utilization of the use of a quantitative water T₂ map of muscle in clinical and research studies.

In understanding muscle health, the concept of "muscle quality" has become pivotal¹. "Muscle quality" is a term encompassing the compositional and functional alterations occurring in skeletal muscles due to factors like aging, diseases, injury, disuse, or training. These changes, which go beyond mere "muscle quantity," are critical in assessing overall muscle health. In this context, Quantitative Magnetic Resonance Imaging (qMRI) has emerged as a leading technique for *in vivo* study of muscle health. Its capability to assess various biomarkers relevant to both "muscle quality" and "muscle quantity" across different skeletal sites makes it exceptionally valuable²⁻⁴.

A notable compositional alteration is the replacement of muscle fibers by fat, which is observed in various physiological and pathological conditions, such as aging^{5,6}, muscle inactivity^{7,8}, and neuromuscular disorders³. Fat infiltration, however, marks an irreversible stage in the degeneration of muscle tissue, limiting its utility to only indicating the chronic or end-stage of muscle degeneration. Thus, once fat infiltration is observed therapeutic interventions are limited and aimed at decelerating or halting the further deterioration of muscle tissue. However, preceding these irreversible changes, muscle fibers typically experience damage characterized by cycles of degeneration and regeneration. This ongoing damage often leads to persistent inflammation, contributing to the progression of muscular degeneration^{9,10}. In this context, the transverse relaxation time (T₂) of the water present in the myocytic tissue has been shown to be a valuable biomarker of disease activity. T₂ has been shown to be sensitive to the inflammatory response indicative of early-stage muscular degenerative disorders, such as muscular dystrophies¹¹⁻¹³, as well as conditions related to aging¹⁴ and injury¹⁵. Unlike fat fraction measurements, which reflect irreversible changes, T₂ of the myocytic component provides information about processes that potentially can be reversed, offering valuable insights for timely and effective therapeutic interventions.

¹Department of Radiology, Stanford University, Stanford, CA, USA. ²Department of Radiology and Nuclear Medicine, Amsterdam University Medical Center, Amsterdam, The Netherlands. ³Department of Cardiology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. ⁴Department of Bioengineering, Stanford University, Stanford, CA, USA. ⁵Bernard and Irene Schwartz Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine, New York, NY, USA. ✉email: mb7@stanford.edu

Two-dimensional Multi-Echo Spin-Echo (MESE) MRI is the gold standard method to obtain a quantitative measure of the T_2 parameter. Typically, the experimental signal decay obtained from such sequences is analyzed voxel-wise using monoexponential modelling, which implicitly assumes that only a tissue component is present within an imaging voxel (“global T_2 ”). This way, when multiple tissues are present within the voxels, they all contribute to the signal and therefore affect the measured T_2 . Inflammatory processes and fat replacement often co-exist within the same muscle, early detection of disease progression or response to treatments requires to separate the contribution of fat and muscle to the measured T_2 . When fat replacement (myosteatosis) is present, it results in an overestimation of T_2 using a one component model for fitting. Elevations of T_2 , indicating disease activity⁴, are usually associated with myosteatosis, T_2 becomes then a biomarker of fat infiltration hindering its ability to inform on disease activity. For this reason, the correct estimation of the T_2 of lean muscle tissue (water T_2 or T_{2w}) requires multicomponent modeling. Furthermore, in 2D sequences the magnetic field used for excitation (B_1^+ field) generates an imperfect deposition of power across the slice thickness, the so-called slice profile effect. Thus, in MESE sequences this generates partial refocusing of the transverse magnetization across the slice. Contrary to perfect 180° refocusing pulses, pulses that are not exactly 180° not only partially refocus the magnetization, but also create longitudinal magnetization. This generates stimulated echoes¹⁶ that are added to the spin-echo pathway. This generally creates an oscillating behavior of the echo decay curve and overall lengthens the decay, biasing the true T_2 toward apparent longer T_2 times if this effect is not considered. The non-uniformity of the B_1^+ field across the field of view further exacerbates the aforementioned behavior, which implies that a pure exponential model is not adequate for signal processing.

A fitting procedure that addresses many of the mentioned issues in muscle exploits the extended phase graph (EPG) formalism to model the MESE signal decay with a two-component model: that is intramuscular fat and water. Such an approach is considered the gold-standard to estimate the T_2 of lean muscle tissue (T_{2w}), and it is robust to the presence of fat infiltration¹⁷. Two main classes of EPG methods exist: nonlinear least square fitting (NLSQ)-based procedures¹⁷ or dictionary-based procedures¹⁸. In the NLSQ approach, a non-linear iterative optimization procedure is used to minimize the sum of the squares of the difference between the experimental signal and the EPG-modeled theoretical signal, voxel-wise. In the dictionary-based approach, the experimental MESE signal is matched, voxel-wise, with a precomputed dictionary of simulated signals. Both techniques have been shown to accurately estimate T_2 of the muscle independently of fat infiltration¹⁷⁻¹⁹. It should be noted that to obtain accurate T_{2w} the slice profile used in the EPG model should match as close as possible the real slice profile in the MESE acquisition, which requires the knowledge of the pulse shapes used to acquire the data. This should be considered when different vendors apply different proprietary pulses.

Despite their robustness, both methods suffer from high computational burden and long computational time on standard computers, usually requiring several tens of minutes per subject, which limits the wide applicability of these methodologies for research and clinical applications. For the NLSQ procedure, for each pixel several iterations are necessary to find the optimal fit to the experimental data. For the dictionary-based approach, the computational burden arises from the creation of the dictionary and the subsequent exhaustive search across it. Reducing the size of the dictionary by decreasing its resolution is a straightforward strategy to limit the computational burden. However, this may introduce quantization artifacts, as the output MR parameters are limited to those present in the dictionary. The use of high-end computers with the assistance of graphical processing unit (GPU) can accelerate computing¹⁸, but at the cost of more high-end hardware. Ideally, one would want to maintain the accuracy of EPG fitting procedure while requiring limited computational resources and short processing times.

Long computational times and high computational resources are detrimental for several reasons. To increase fitting accuracy, the T_2 of the fat component (T_{2f}) should be calibrated for each subject applying an EPG fitting procedure in a region of subcutaneous fat. When a large population is studied, to decrease the computational burden associated with the need of computing multiple dictionaries in the dictionary-based approach, T_2 of the fat component is fixed to a value (estimated from literature or as average of the population under study) and used for all the subsequent muscle T_2 estimation. However, this procedure can lead to inaccuracy as it has been reported that even small variation in T_{2f} can cause significant differences in muscle T_2 estimation¹⁹. Another important motivation for shortening EPG fitting procedures with the use of minimal computational resources is that it could allow real-time processing of the data while a subject is undergoing the MRI exam. This could potentially suggest modifications to the MRI protocol informed by real-time examination of the quantitative T_2 maps.

A strategy to overcome the computational burden of current EPG fitting methods is to leverage deep learning (DL) algorithms such as neural networks (NN) to predict, voxel-wise, the T_2 of the lean muscle tissue given the experimental MESE as input after a supervised training procedure. NNs have been demonstrated to be universal function approximators given enough training data and model complexity²⁰. A NN can, in principle, learn an approximation of the inverse transfer function that maps the acquired MESE signal into the MR parameter space. Moreover, once the NN is trained, the prediction operation requires extremely limited computational resources. Further, under optimal training conditions, a NN can effectively process signals that were never seen during training. This limits quantization artifacts that can arise from a dictionary approach, in which template matching approximates MR parameters to those present in the dictionary.

The aim of this work was to develop a deep learning approach based on fully connected NN for fast muscle T_{2w} mapping with subject-specific T_{2f} calibration that (i) requires minimal computational resources, and (ii) takes only a few seconds to perform the computation. We validated our method in vivo using the MESE product sequence of two MRI vendors: Siemens and Philips.

The source code to set up and run the application is openly available at https://github.com/barma7/Deep_Learning_for_Muscle_T2_mapping.git.

Methods

EPG simulations

The pixel-wise 2D MESE signal was modeled using the EPG algorithm taking into account the effects of slice profiles as described by Lebel and Wilman²¹. A two-component model was used to represent fat and water pools within the muscle tissue according to Eq. 1

$$S(t) = FF \times EPG(T_{1f}, T_{2f}, B_1, TE, EXprofile, REFprofile, ETL) + WF \times EPG(T_{1w}, T_{2w}, B_1, TE, EXprofile, REFprofile, ETL) \quad (1)$$

where T_{1f} , T_{2f} , T_{1w} and T_{2w} are the fat and water longitudinal and transverse relaxation times, respectively; B_1 is the pulse efficiency ratio; TE is the echo time space; $EXprofile$ and $REFprofile$ are the excitation and refocusing slice profiles, respectively; ETL is the echo train length (number of refocusing pulses in the MESE sequence); FF and WF are the fat and water fractions, respectively.

T_{1f} and T_{1w} were fixed to 365 and 1400 ms, respectively¹⁷. Given the excitation and refocusing pulse shapes of Philips and Siemens MESE sequences used, the slice profiles were obtained using the Shinnar-Le Roux (SLR) algorithm²². All simulations were performed using an in-house MATLAB (MathWorks, release R2022b) library based on the StimFit toolbox²³.

Deep learning approach: NN model and training pipeline

The Neural Network application was developed in Python using Keras with TensorFlow²⁴ backend and is summarized in Fig. 1. Two NN models for pixel-wise processing were defined: Fat-Net (Fig. 1a) and Muscle-Net (Fig. 1b). Fat-Net estimates T_{2f} , T_{2w} and B_1^+ taking the MESE signals coming from regions of sub-cutaneous fat as input. Muscle-Net predicts the fat-fraction (FF), T_2 of the lean muscle component (T_{2w}) and B_1 given the MESE signal and an estimate of T_{2f} as inputs. Both models have 6 fully connected layers, the Rectified Linear Unit (ReLU) was used as the activation function for the neurons in the first 5 hidden layers, while a linear activation function was used for the output layer. A bottle-neck structure was chosen as it has been successfully applied in other quantitative MR applications^{25,26}.

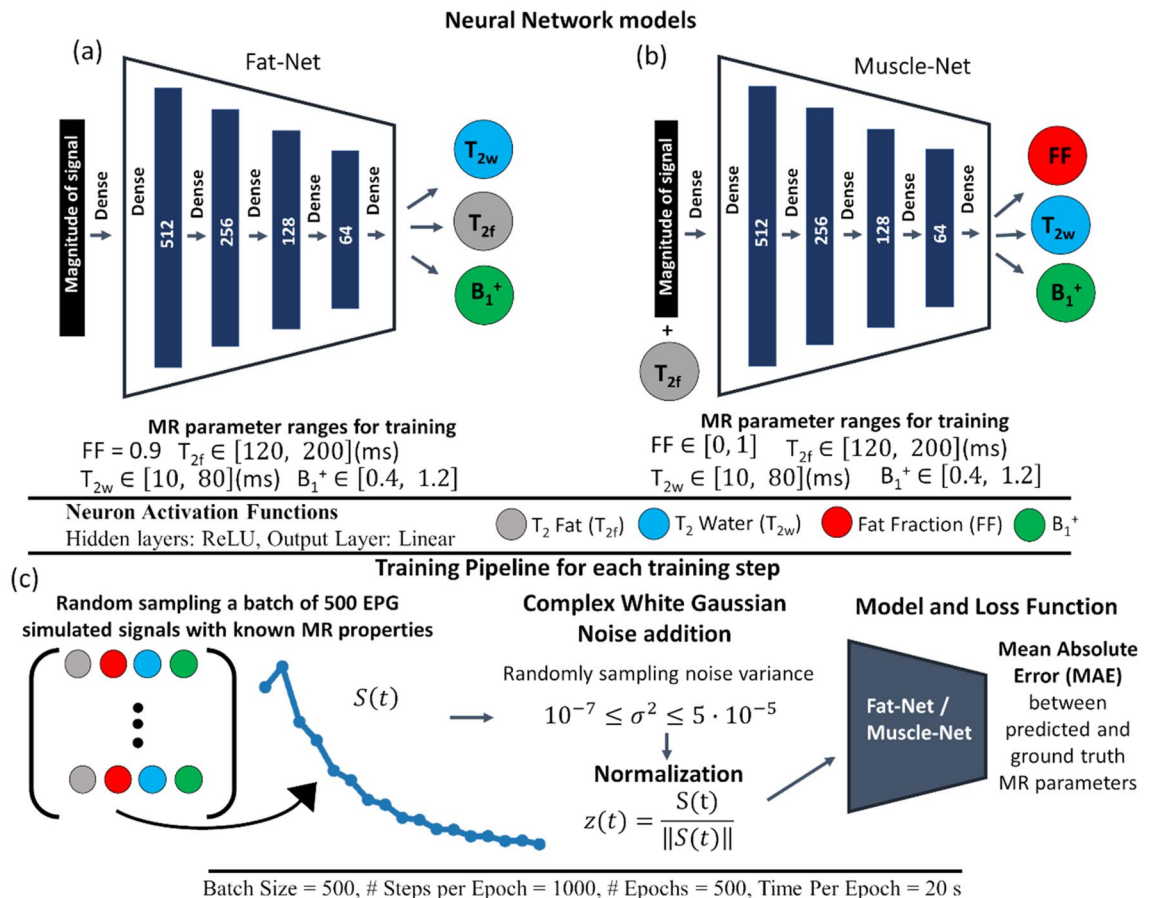


Figure 1. Schematization of the NN models and the training pipeline used for training. (A) Fat-Net and (B) Muscle-Net have the same architecture: 4 fully connected hidden layers with a progressively decreasing number of neurons. (C) During training, simulated data are augmented by injecting white gaussian noise with different noise variances. Signals are normalized before being fed to the NN.

Each model was trained using MESE data simulated using the EPG formalism with a supervised procedure, with the mean absolute error (MAE) between NN estimated parameters and ground-truth parameters as loss function. The Adam algorithm²⁷ was used as optimizer using 500 training epochs with 1000 gradient steps for each epoch, with a fixed batch size of 500. The training sets were composed of 100,000 MESE signals generated by randomly sampling the parameter space according to a uniform distribution. Specifically, for Fat-Net, subcutaneous fat signals were generated by randomly sampling T_{2f} and T_{2w} in the ranges [50, 250] ms and [10, 110] ms, respectively, using a fixed FF value of 0.9. For Muscle-Net, muscle tissue signals were generated by randomly sampling T_{2f} , T_{2w} and FF in ranges: [50, 250] ms, [10, 80] ms and [0, 1] respectively. For both training sets, B_1^+ was sampled within the range [0.4, 1.2].

With reference to Fig. 1c, to promote model generalization and robustness to noise, for each training step, simulated noiseless MESE signals were corrupted by white Gaussian noise with different variances, which are randomly sampled from a uniform distribution with $10^{-7} \leq \sigma^2 \leq 5 \cdot 10^{-5}$. Using a wide range of noise variances has been proved to be the best noise addition strategy when a similar NN architecture was used for MR pixel-wise parameter mapping²⁶. For training Fat-Net a narrower range of noise SNR ($10^3 \leq \text{SNR} \leq 10^4$) was used as pixels in the subcutaneous fat displays higher SNR than pixels in the muscles. Preliminary tests empirically showed that reducing the noise range to higher SNR values improved T_{2f} estimation. After the data augmentation step, the magnitude of each input was taken and normalized to the unit norm. When applied to real acquired data for pixel-wise processing, the experimental MESE data is first normalized to the unit norm, by dividing the signal with its norm, and then fed to the pre-trained network model.

Because the product MESE sequence of the two vendors utilized different pulse shapes, two deep learning applications were developed independently for Siemens and Philips by using training data simulated with their corresponding proprietary pulses.

Reference EPG fitting procedures

Two EPG fitting procedures conventionally adopted in the literature were used as reference against the proposed DL application. They were NLSQ¹⁷ and dictionary matching¹⁸ (DM) fitting. For the NLSQ, the MR parameter spaces for fat calibration and muscle fitting were constrained within the same ranges used for training Fat-Net and Muscle-Net, respectively. For the Dictionary application, two dictionaries were created for fat calibration and muscle processing, respectively. Specifically, for fat calibration FF was fixed to 0.9 and T_{2f} , T_{2w} and B_1^+ values spanned the ranges [50, 250] ms, [10, 110] ms and [0.4, 1.2], respectively, with steps sizes of 0.25 ms, 0.5 ms and 0.025 ms, respectively. For muscle processing, T_{2f} was fixed to the calibrated value while FF, T_{2w} and B_1^+ spanned the ranges [0, 1], [10, 80] ms and [0.4, 1.2], respectively, with steps sizes of 0.015, 0.2 and 0.025 ms, respectively. Both applications were developed in MATLAB R2022b and were optimized for computational efficiency through code vectorization and exploitation of MATLAB's automatic parallel support for multicore CPU. As for the deep learning approach, different dictionaries were created for Siemens and Philips by using their corresponding proprietary pulses used in the MESE sequence. Similarly, for the NLSQ fitting procedure the corresponding proprietary pulses were used for EPG modelling during least square minimization when analyzing Siemens and Philips data.

Computer configuration and computational time

All applications (deep learning, dictionary and NLSQ) were run on a standard laptop equipped with 16 Gb of Random Access Memory (RAM) using only Central Processing Units (CPUs): Intel Core i7-8565U. The Graphical Process Unit (GPU) integrated in the laptop, an Intel UHD Graphics 620, was not used at any point. With such configuration, each deep learning model required 20 s per epoch, resulting in a total of 2 h and 42 min to complete the training. For the dictionary approach, 250 s were needed to create the dictionary for fat calibration, whereas the dictionary for muscle T_{2w} estimation required 1.2 s once T_{2f} was calibrated.

MRI acquisitions

A multi-stack upper leg bilateral scan of ten healthy subjects was performed at 3 T. Five were scanned using a Philip scanner (Ingenia, Philips, Best, Netherlands) for an unrelated study²⁸, and five were scanned using a Siemens scanner (MAGNETOM Skyra, Siemens Healthineers, Erlangen, Germany). The MRI protocol included a Dixon water/fat-separated sequence for anatomical reference and for fat fraction estimation, and a MESE sequence for water T_2 mapping. Sequence parameters for each scanner are reported in Table 1. The variation in acquisition parameters between Siemens and Philips scanners reflects the default settings that each vendor set to optimize acquisition time and SNR given a specific product sequence. In this work, our goal was to use product sequences optimized by the vendor for the specific parameter of interest. Thus, the variation in acquisition parameters between Siemens and Philips scanners reflects the settings that each vendor set for optimizing acquisition time and SNR given a specific product sequence. For the MESE sequence, the focus of our analysis, the main difference resides in the TR (5633 vs 2870 ms for Siemens and Philips, respectively). TR needs to be long enough to allow full recovery of the longitudinal magnetization to its equilibrium value to minimize T_1 -weighting. To achieve such a goal, TR is normally set to a value between 2 and 4 times the longest T_1 of the component of interest. In our case T_1 of the water component is about 1400 ms while T_1 of the fat component is about 365 ms. TR for both vendors were long enough to minimize T_1 -weighting and it therefore unlikely that this difference in TR could significantly affect T_2 analysis.

Data processing and analysis

The data pre-processing pipeline is summarized in top panel of Fig. 2. For each subject, ten muscles (Vastus Medialis, Vastus Lateralis, Vastus Intermedius, Rectus Femoris, Biceps Femoris Long Head, Biceps Femoris

Scanner	PHILIPS		SIEMENS	
	T ₂ mapping	DIXON	T ₂ mapping	DIXON
Sequence name	MESE	MS-FFE	MESE	GRE-ME
Matrix size	160 × 160	320 × 320	132 × 192	172 × 288
Voxel size (mm ³)	3 × 3 × 6	1.5 × 1.5 × 6	2 × 2 × 6	1.4 × 1.4 × 5
Slice gap (mm)	6	–	6	–
Slices	17	31	15	52
TR (ms)	5633	210	2870	17.15
TE (ms)	7.6	2.6/3.36/4.12/4.88	7.5	1.3
ETL	17	4	17	12
Flip Angle	90°/180°	8°	90°/180°	9°
Acquisition Time (m:s)	4:32	1:25	5:55	0:21

Table 1. Sequence parameters used for the MRI protocols run using Philips and Siemens scanners. *FFE* Fast Field Echo, *GRE* Gradient Recalled Echo, *ME* Multi-Echo, *MESE* Multi-Echo Spin-Echo, *MS* Multi-Slice.

Short Head, Semimembranosus, Semitendinosus, Sartorius and Gracilis) in both legs were segmented in 10 slices of the out-of-phase Dixon images acquired at mid-thigh. For Siemens data, segmentation was carried out automatically using the Deep Anatomical Federated Network (DAFNE) collaborative platform²⁹ and subsequently manually verified and corrected to ensure accuracy of the segmentation. For Philips data, muscles were manually segmented. For both vendors and for each subject, the Dixon images were then rescaled to match the MESE resolution and registered to the first MESE echo using Elastix³⁰. Lastly, a region of subcutaneous fat was selected with an automatic K-means clustering segmentation based on the Lloyd's algorithm³¹ using the first echo of the MESE echo. Subcutaneous segmentations were also manually checked and corrected if necessary to avoid inclusion of background or muscle pixels within the subcutaneous mask.

Once the data underwent the pre-processing pipeline described above, quantitative analysis was performed as summarized in the bottom panel of Fig. 2. For each vendor and for each subject, to obtain T_{2w} of the muscle component, T_{2f} was first calibrated in the subcutaneous region using the pre-trained Fat-Net, and fed as input to the pre-trained Muscle-Net. The estimation was also repeated using both dictionaries matching and NLSQ fitting procedures for reference. The average T_{2w} value in each muscle was calculated. We used Bland-Altman analysis³² to compare the T_{2w} values obtained with our deep learning approach to the values obtained using the reference fitting methods. The limits of agreement (LOA) and the Lin's concordance correlation coefficient³³ (r_c) were used as global metrics of agreement between methods. A separate analysis was conducted for Siemens and Philips data as the MESE sequence parameters and the pulse shapes used differed between the two vendors. Furthermore, the reference average FF in each muscle was computed from the Dixon FF maps, which were obtained, pixel-wise, using Eq. 2, where I_w and I_f are the water-only and fat-only images output by the scanner. For each T_{2w} fitting method, a BA analysis comparing the reference Dixon-based FF values with the FF estimated by the bi-component fit was performed. Lastly, the correlation between average T_{2w} values and the Dixon-based FF values was also assessed.

$$FF = \frac{I_f}{I_w + I_f} \times 100\% \quad (2)$$

Results

Representative parameter maps reconstructed using the proposed DL approach and the reference methods for Siemens and Philips dataset are shown in Figs. 3 and 4, respectively. For both datasets the quantitative maps appear to be similar across fitting methods by visual inspection and no artefacts are noted. The average computational time required to perform fat calibration and process all the slices in a subject were 6 s, 200 s and 700 s for the DL, dictionary and NLSQ methods, respectively.

The BA plots for each scanner are presented in Fig. 5. The T_{2w} values estimated using the proposed DL approach were in very good agreement with those obtained with the dictionary approach based on dot-product metric and with those computed using the NLSQ fitting procedure. Specifically, in Siemens data when comparing the DL approach with the dictionary and NLSQ methods, respectively, r_c were equal to 0.89 and 0.92, respectively, and the widths of LOA were equal to ± 0.62 ms and ± 0.50 ms, respectively. Similarly, when looking at Philips data, r_c were equal to 0.96 and 0.97, respectively, and the widths of LOA were equal to ± 0.62 ms and ± 0.56 ms, respectively. In Siemens, small, but statistically significant, biases equal to − 0.4 ms and 0.3 ms were found between T_{2w} values evaluated using the DL approach and T_{2w} values obtained with the dictionary and the NLSQ reference methods, respectively. In Philips, only a small bias equal to − 0.1 ms between T_{2w} values evaluated using the DL approach and T_{2w} values obtained with the reference dictionary method was statistically significant. Similarly, very good agreement was found between B₁⁺ and FF values estimated using the proposed DL approach and the reference methods, as reported in the BA plots in Figures S1 and S2. Concerning FF accuracy, the BA plots comparing the FF values estimated using the EPG methods and the reference FF values from Dixon are reported in Figure S3. The EPG method overestimated the reference FF by about 3 percentage points (p.p.) regardless of the fitting procedure utilized. Further, a positive linear relationship between FF deviation and FF

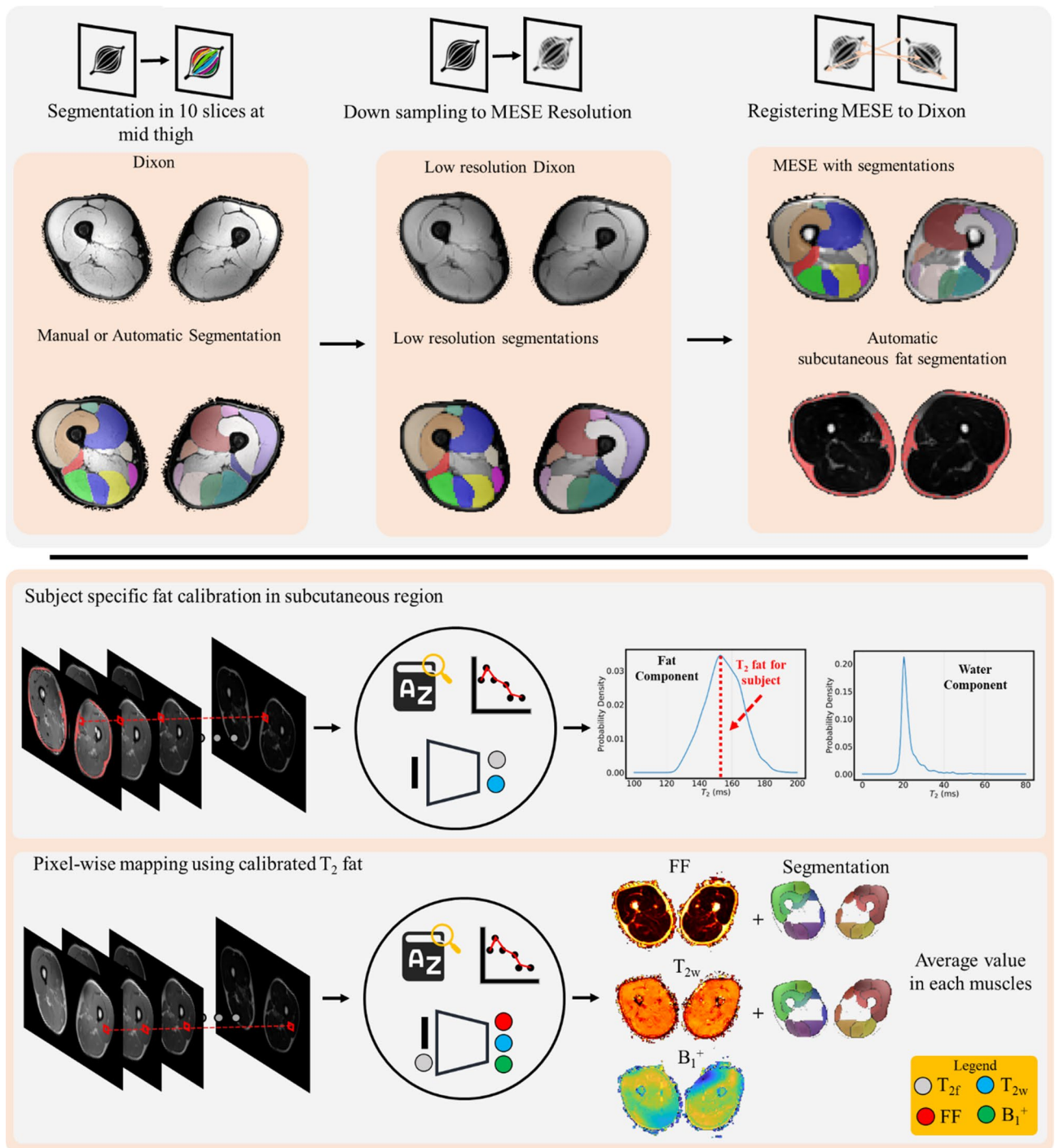


Figure 2. Processing pipeline used to analyze the data. Image pre-processing (top panel) and data analysis (bottom panel) steps used to produce the parametric maps for each subject: T_{2w} , FF and B_1^+ with subject specific T_{2f} calibration.

value was also observed. However, this was independent of the fitting method used (NLSQ, DM or NN). Finally, as highlighted in Fig. 6, no statistically significant correlation was found between the reference FF values obtained with Dixon and T_{2w} values obtained used the 3 EPG methods, indicating that they can successfully remove the contribution of fat from the MESE signal.

Discussion

In this work, we developed a neural network approach for fast muscle T_{2w} mapping with subject-specific T_{2f} calibration to overcome the computational burden and the lack of flexibility of current bi-component EPG fitting methods. This method was validated in vivo using the MESE product sequence of two MRI vendors, Siemens and Philips.

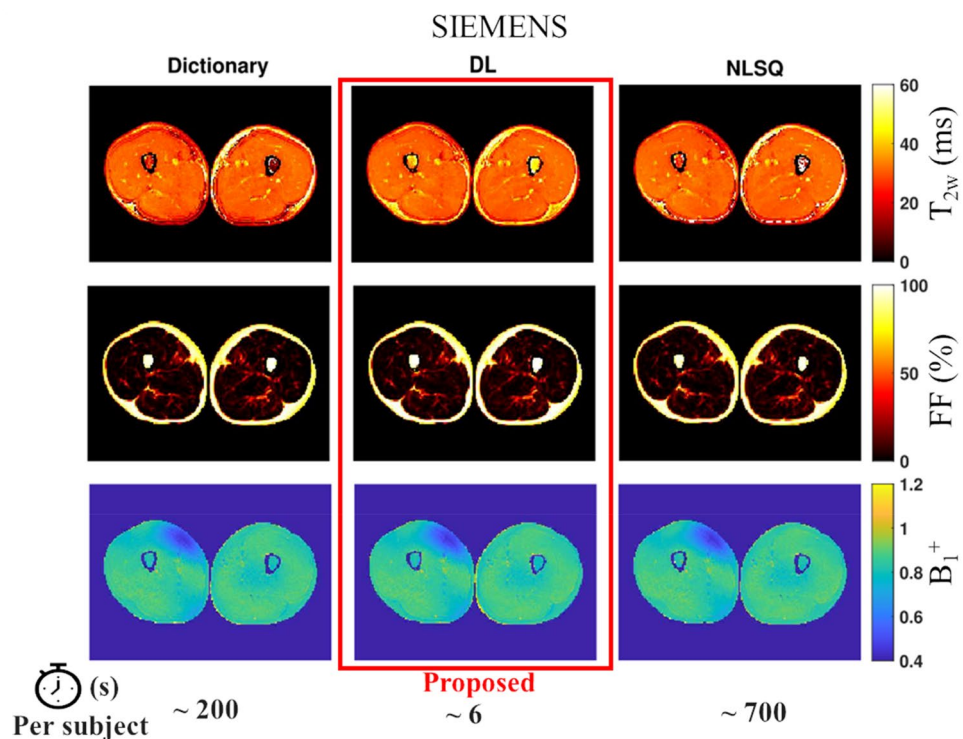


Figure 3. Example of T_{2w} , FF and B_1^+ quantitative maps obtained in the thigh from EPG fitting of MESE data acquired with Siemens scanner. The proposed deep learning approach (central column, red frame) is presented along with the maps obtained using the dictionary (left column) and NLSQ (right column) reference fitting methods.

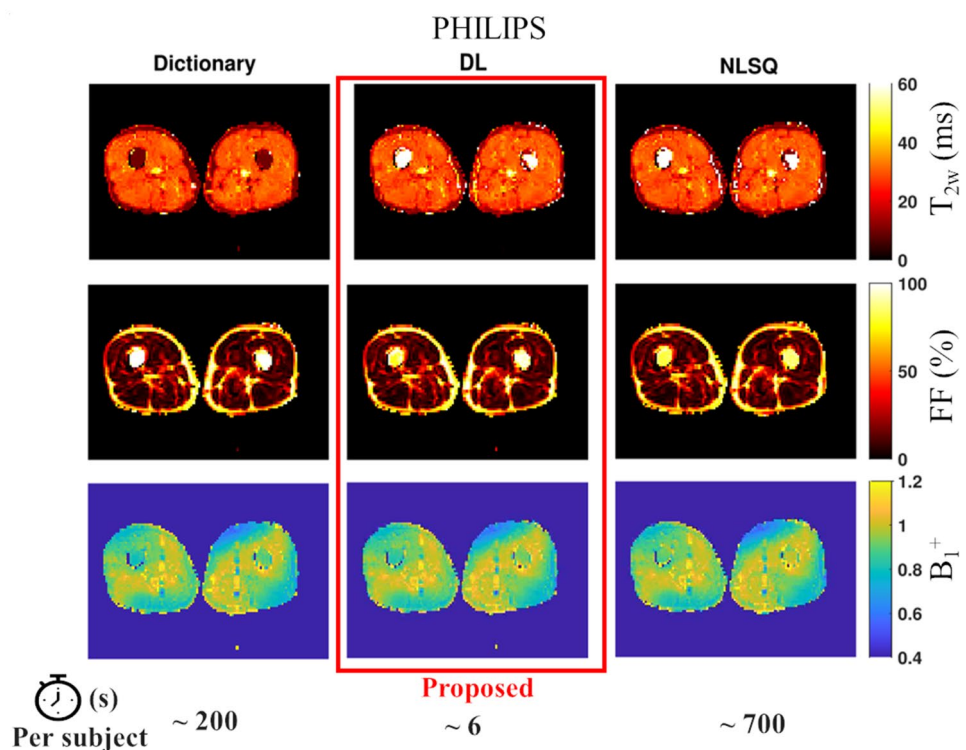


Figure 4. Example of T_{2w} , FF and B_1^+ quantitative maps obtained in the thigh from EPG fitting of MESE data acquired with Philips scanners. The proposed deep learning approach (central column, red frame) is presented along with the maps obtained using the dictionary (left column) and NLSQ (right column) reference fitting methods.

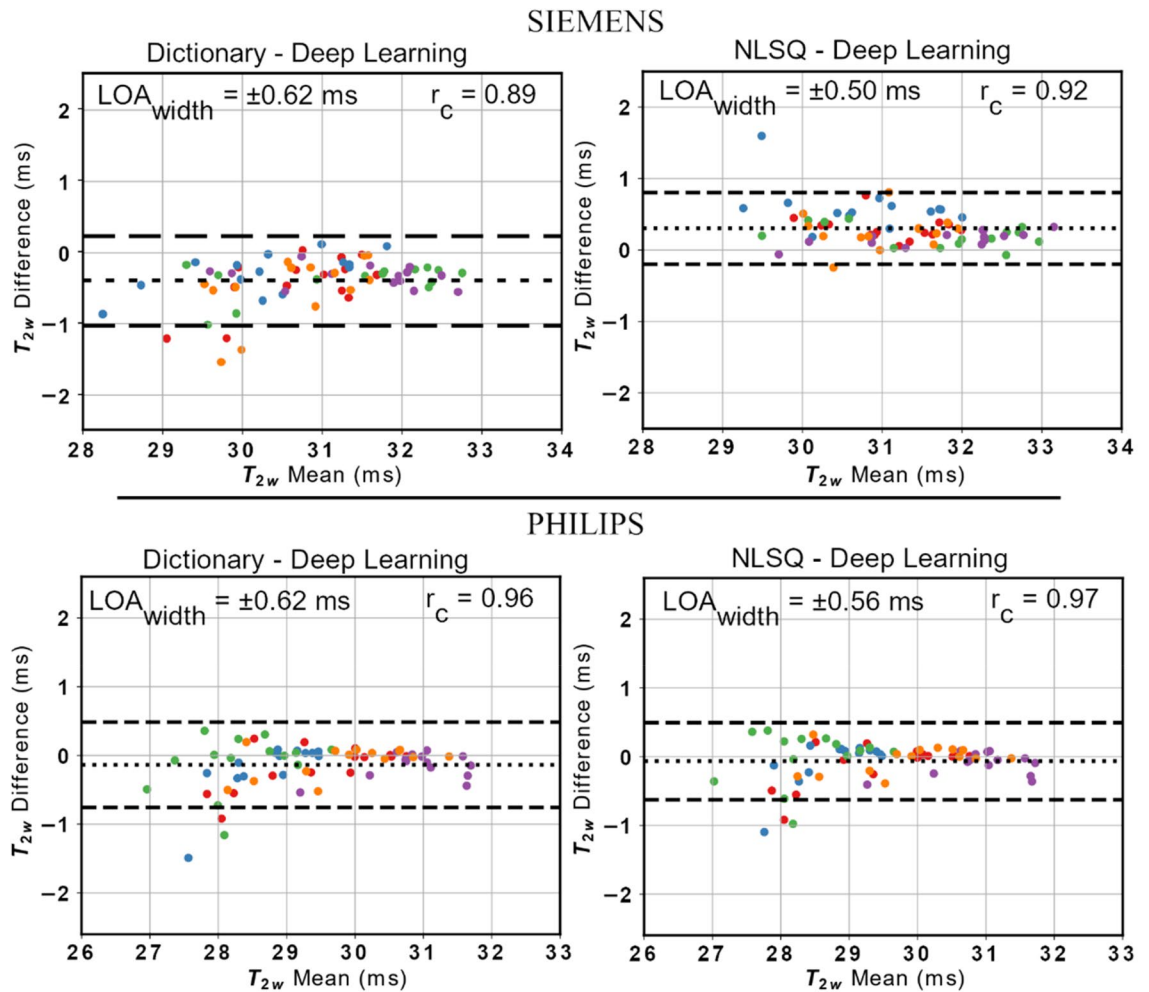


Figure 5. Bland–Altman plots reporting the comparison of T_{2w} values obtained using the proposed DL approach with the reference dictionary (left) and NLSQ (right) EPG fitting methods for data acquired using Siemens (top panel) and Philips (bottom panel) scanners. The width of the LOA is reported for each BA plot along with the Lin’s concordance coefficient. The different colors represent muscle T_{2w} values belonging to different subjects.

Our DL approach was able to predict T_{2w} values that were in strong agreement with those obtained using the reference NLSQ and dictionary approaches, while reaching a $116\times$ and $33\times$ computational time improvement, respectively. We observed a sub-millisecond bias in the T_{2w} negative obtained with our method compared to NLSQ and dictionary matching. However, these biases were smaller than the bias observed between these two reference methods. A bias of 0.9 ms between the dictionary matching method and the NLSQ methods has been previously reported in literature¹⁸ and it was consistent with the bias observed in this work. Furthermore, we observed that inter-method biases in the Siemens dataset were overall higher than the biases observed in the Philips dataset. Although we did not further explore the motivation of such difference, this could be related to the fact that, on average, B_1^+ efficiency observed in the Siemens acquisitions was about 15% less efficient than the one measured in Philips acquisitions. The EPG method accounts for B_1^+ inhomogeneity, but this could still have an impact as fitting metrics and optimization methods differ among the approaches. We observed that the EPG method overestimated the FF of about 3 p.p compared to the reference Dixon values and observed a positive relationship between FF bias and FF value. Although this is not of primary interest or concern, it is worth commenting on what could lead to such bias and behavior. Among several factors that might contribute to it, a major one could lie in the difference between the EPG-based and Dixon-based methods for water-fat separation. Dixon imaging is based on the chemical shift between water and fat. The EPG method does not consider the chemical shift between water and fat, as it models data acquired with MESE sequences. However, it has been recently shown that accounting for chemical shift artefacts in the slice direction due to misalignment of the slice profiles between water and fat improves accuracy of FF estimation in EPG fitting of MESE data¹⁹. Thus, such artefacts could be included in the modelling for improved FF accuracy. Nonetheless, the deviation between EPG-based and Dixon-based FF estimations did not adversely impact the obtained results in terms of muscle T_{2w} estimation, as no statistically significant correlation between T_{2w} values and reference Dixon-based FF was observed. This suggests that each EPG method successfully separated the T_{2w} from fat. For the NLSQ and dictionary method this behavior has been already observed¹⁸. The observation that also the proposed DL approach showed such behavior increases confidence on the method. However, it should be noted that only

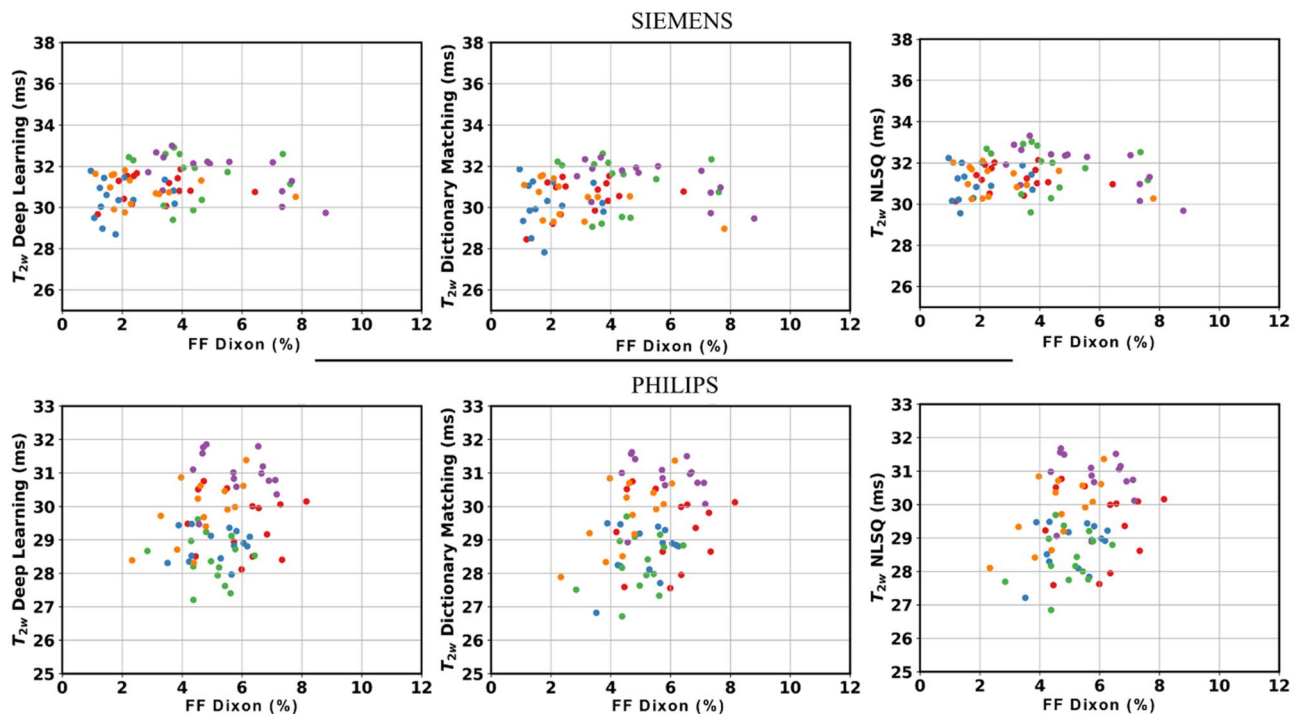


Figure 6. Scatter plots of estimated T_{2w} values against Dixon fat fraction values. The plots are reported separately for Siemens (top row) and Philips data (bottom row). A plot for each EPG fitting procedure investigated is reported: the proposed DL approach and the reference dictionary and NLSQ methods. The different colors represent values belonging to different subjects.

young healthy subjects were included in the current study. Thus, although we do not expect our DL approach to behave differently from the reference EPG methods, the absence of a positive correlation between T_{2w} and FF values should be verified when the method is applied to older subjects and patients with neuromuscular diseases, as a much higher FF is to be expected in this case.

It is worth noting that regardless of the method used, estimation of muscle T_{2w} in region where FF is close to 1 is not reliable. In such regions (for example bone marrow and subcutaneous fat), the amplitude of the water component is close to zero leading to very low SNR of the water signal compared to the SNR of the fat signal. In such cases, the fit tends to predict T_{2w} values equal to (or close to) either the lower or higher bound set for the fitting regardless of the method used. The metric and optimization algorithm used to perform the fitting influences whether the lower or higher bound is more likely to be selected. In our work all three methods are different in terms of metric and optimization algorithm used. Thus, it is expected to have different behaviors in regions where the water component has very low SNR compared to the fat component. As a result, the differences in T_{2w} estimation in the bone marrow region among methods observed in Figs. 3 and 4 are of no concern.

The most relevant benefit of our approach lies in the high computational efficiency and fast processing times (6 s per subject) to obtain muscle T_{2w} mapping with subject specific T_{2f} calibration. Even small variations in T_{2f} have been shown to cause differences up to 4 ms in T_{2w} ¹⁹, and we observed a relatively wide range of T_{2f} values in a cohort of healthy participants as reported in Figure S4, which is expected to be even wider in patients. However, in practice often a single fixed T_{2f} is assumed for dictionary matching approaches to keep storage and computational resources manageable, resulting in a bias in T_{2w} estimation. The approach can be speeded up with the use of parallel computing paired with Graphical Process Units (GPUs), which is the strategy adopted by the open-source toolbox MyoQMRI¹⁸. In the paper explaining the toolbox, Santini et al. reported an average processing time of 2 s/slice, which if applied to our case would have led to a total computational time of about 30 s per subject. This would be $6.6 \times$ faster than our CPU-based dictionary-matching approach, and about $5 \times$ slower than the proposed CPU-based deep learning approach. Furthermore, with increased dictionary size, more memory on the device is needed to exploit benefit from GPU parallel computation. Depending on the application, one may want to increase the resolution of the dictionary, and the higher required memory eventually becomes the limiting factor despite the use of GPUs in dictionary-based approaches. With the proposed DL approach this task is easily handled even with limited computational resources. For example, execution of our application was performed on a middle-range laptop equipped with 16 Gb of RAM and used only CPUs power (Intel Core i7-8565U). Another approach to increase computational efficiency of the dictionary matching approach is to use a low-rank approximation of the dictionary using singular value decomposition (SVD). In the case of MESE, three principal components allow to retain more than 99% of the total variance present within the dictionary as observed in Figure S5. We observed a 15% reduction in computational time (from about 200 s to about 175 s for a single subject) when using SVD compression compared to the “conventional” dictionary matching approach. It is worth pointing out that we did not systematically evaluate the effect of SVD compression on T_{2f} and T_{2w}

estimations when comparing it with the un-compressed dictionary method. Since more than 99% of the total variance is retained with 3 components, one should expect the deviation with the non-SVD compressed case to be negligible. However, such deviation should be quantitatively characterized before using SVD compression for muscle T_{2w} analysis. Notably, our NN method outperforms the dictionary-matching methods in terms of computation time, even when SVD compression is used.

One limitation of our approach is that a new training is required whenever MESE sequence parameters (TE, ETL, pulses) are changed. However, it is common to use only one MRI protocol in clinical and research studies. In such cases, fast processing with our DL approach outweighs the time required to train the NN models, which with our configuration required little less than 3 h. Such time can be significantly reduced with the use of GPUs. Thus, one could train the model on a high-end computer and exploit the benefits of DL processing with minimal required computational resources: potentially the processing could be directly performed at the scanner, facilitating implementation into clinical practice as no advanced post-processing would be needed. Another potential limitation is that estimation of T_{2w} values is sensitive to the selection of the slice-profile. Thus, it is important to obtain detailed information about the timing of the MESE sequences and the shapes of the excitation and refocusing pulses, as using incorrect pulse profiles may result in biases in T_{2w} estimation. However, it is worth mentioning that this limitation is inherent with any EPG fitting algorithm and not exclusive to our method.

In conclusion, we developed a DL approach for fast muscle T_{2w} mapping with subject specific T_{2f} calibration that outperformed cutting-edge literature methods in required computational time and resources while providing same quantitative accuracy. This methodology allows the processing of T_{2w} data in a few seconds, compared to tens of minutes required with current methods, and will facilitate the use of quantitative T_{2w} maps of muscle in clinical and research studies.

Data availability

The source code to develop and run the deep learning application, as long as the conventional EPG fitting methods is available at https://github.com/barma7/Deep_Learning_for_Muscle_T2_mapping.git. The MRI data and source code to run the data analysis to reproduce the findings of this study are openly available in *Zenodo* at <https://doi.org/https://doi.org/10.5281/zenodo.10520542>.

Received: 17 January 2024; Accepted: 3 April 2024

Published online: 08 April 2024

References

- McGregor, R. A., Cameron-Smith, D. & Poppitt, S. D. It is not just muscle mass: A review of muscle quality, composition and metabolism during ageing as determinants of muscle function and mobility in later life. *Longev. Heal.* **3**, 9 (2014).
- Smeulders, M. J. C. *et al.* Reliability of in vivo determination of forearm muscle volume using 3.0 T magnetic resonance imaging. *J. Magn. Reson. Imaging JMRI* **31**, 1252–1255 (2010).
- Strijkers, G. J. *et al.* Exploration of new contrasts, targets, and MR imaging and spectroscopy techniques for neuromuscular disease—A workshop report of working group 3 of the biomedicine and molecular biosciences COST action BM1304 MYO-MRI. *J. Neuromuscul. Dis.* **6**, 1–30 (2019).
- Hooijmans, M. T. *et al.* Compositional and functional MRI of skeletal muscle: A review. *J. Magn. Reson. Imaging* **n/a**.
- Farrow, M. *et al.* The effect of ageing on skeletal muscle as assessed by quantitative MR imaging: An association with frailty and muscle strength. *Ageing Clin. Exp. Res.* **33**, 291–301 (2021).
- Yoon, M. A. *et al.* Multiparametric MR imaging of age-related changes in healthy thigh muscles. *Radiology* **287**, 235–246 (2018).
- Bogdanov, J. *et al.* Fatty degeneration of the rotator cuff: Pathogenesis, clinical implications, and future treatment. *JSES Rev. Rep. Tech.* **1**, 301–308 (2021).
- Rostron, Z. P. J. *et al.* Effects of a targeted resistance intervention compared to a sham intervention on gluteal muscle hypertrophy, fatty infiltration and strength in people with hip osteoarthritis: Analysis of secondary outcomes from a randomised clinical trial. *BMC Musculoskelet. Disord.* **23**, 944 (2022).
- Mavrogeni, S. *et al.* Oedema-fibrosis in Duchenne Muscular dystrophy: Role of cardiovascular magnetic resonance imaging. *Eur. J. Clin. Investig.* **47**(12), e12843 (2017).
- Monforte, M. *et al.* Tracking muscle wasting and disease activity in facioscapulohumeral muscular dystrophy by qualitative longitudinal imaging. *J. Cachexia Sarcopenia Muscle* **10**, 1258–1265 (2019).
- Willcocks, R. J. *et al.* Longitudinal measurements of MRI-T2 in boys with Duchenne muscular dystrophy: Effects of age and disease progression. *Neuromuscul. Disord. NMD* **24**, 393–401 (2014).
- Wokke, B. H. *et al.* T2 relaxation times are increased in Skeletal muscle of DMD but not BMD patients. *Muscle Nerve* **53**, 38–43 (2016).
- Reyngoudt, H. *et al.* Three-year quantitative magnetic resonance imaging and phosphorus magnetic resonance spectroscopy study in lower limb muscle in dysferlinopathy. *J. Cachexia, Sarcopenia Muscle* **13**(3), 1850–1863 (2022).
- Azzabou, N., Loureiro Sousa, P., Caldas, E. & Carlier, P. G. Validation of a generic approach to muscle water T2 determination at 3T in fat-infiltrated skeletal muscle. *J. Mag. Resonance Imaging* **41**(3), 645–653 (2015).
- Biglands, J. D. *et al.* MRI in acute muscle tears in athletes: can quantitative T2 and DTI predict return to play better than visual assessment?. *Eur. Radiol.* **30**, 6603–6613 (2020).
- Majumdar, S., Orphanoudakis, S. C., Gmitro, A., O'Donnell, M. & Gore, J. C. Errors in the measurements of T2 using multiple-echo MRI techniques. I. Effects of radiofrequency pulse imperfections. *Magn. Reson. Med.* **3**, 397–417 (1986).
- Marty, B. *et al.* Simultaneous muscle water T2 and fat fraction mapping using transverse relaxation with stimulated echo compensation. *NMR Biomed.* **29**, 431–443 (2016).
- Santini, F. *et al.* Fast open-source toolkit for water T2 mapping in the presence of fat from multi-echo spin-echo acquisitions for muscle MRI. *Front. Neurol.* **26**(12), 630387 (2021).
- Keene, K. R. *et al.* T2 relaxation-time mapping in healthy and diseased skeletal muscle using extended phase graph algorithms. *Magn. Reson. Med.* **84**, 2656–2670 (2020).
- Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**, 251–257 (1991).
- Lebel, R. M. & Wilman, A. H. Transverse relaxation with stimulated echo compensation. *Magn. Reson. Med.* **64**, 1005–1014 (2010).
- Pauly, J., Le Roux, P., Nishimura, D. & Macovski, A. Parameter relations for the Shinnar-Le Roux selective excitation pulse design algorithm (NMR imaging). *IEEE Trans. Med. Imaging* **10**, 53–65 (1991).

23. R. Marc Lebel. StimFit: A toolbox for robust T2 mapping with stimulated echo compensation. in *Proc. Intl. Soc. Mag. Reson. Med.* 20 (2012). <https://archive.ismrm.org/2012/2558.html>
24. Martín Abadi *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems. (2015). Software available from <https://tensorflow.org>.
25. Barbieri, M. *et al.* A deep learning approach for magnetic resonance fingerprinting: Scaling capabilities and good training practices investigated by simulations. *Phys. Med.* **89**, 80–92 (2021).
26. Barbieri, M. *et al.* Circumventing the curse of dimensionality in magnetic resonance fingerprinting through a deep learning approach. *NMR Biomed* **35**(4), e4670 (2022).
27. Kingma, DP. & Ba, J. Adam: A method for stochastic optimization. *ArXiv E-Prints* (2014).
28. Hooijmans, M. T. *et al.* Quantitative MRI reveals microstructural changes in the upper leg muscles after running a marathon. *J. Magn. Reson. Imaging* **52**, 407–417 (2020).
29. Santini, F. *et al.* Deep anatomical federated network (Dafne): An open client/server framework for the continuous collaborative improvement of deep-learning-based medical image segmentation. Preprint at <https://doi.org/10.48550/arXiv.2302.06352> (2023).
30. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. W. Elastix: A toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **29**, 196–205 (2010).
31. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).
32. Bland, J. M. & Altman, D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **327**(8476), 307–310 (1986).
33. Lin, L.I.-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255–268 (1989).

Acknowledgements

This work was supported by GE Healthcare, NIH grants K99AG071735, R00AG071735, R01AR0796431 and R01AR074492.

Author contributions

MB—study conception and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, and critical revision. M.T - acquisition of data and critical revision. K.M — technical support in data acquisition. T.C - acquisition of data and critical revision. D.E. - critical revision. G.G - critical revision. FK - interpretation of data and critical revision. VM — study conception and design, analysis and interpretation of data and critical revision. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-58812-2>.

Correspondence and requests for materials should be addressed to M.B.

Reprints and permissions information is available at 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