Physics-informed Deep Learning Approach to Quantification of Human Brain Metabolites from Magnetic Resonance Spectroscopy Data

Amir M Shamaei[1,2], Jana Starcukova[1], Zenon Starcuk Jr[1]

1 Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

2 Department of Biomedical Engineering, Brno University of Technology, Czech Republic

Github: https://github.com/isi-nmr/Deep-MRS-Quantification

Word count: 5346

Corresponding Author:

Amir M Shamaei,

Magnetic Resonance Group,

Institute of Scientific Instruments of the Czech Academy of Sciences,

Kralovopolska 147, CZ 61264, Brno, Czech Republic

Email: amirshamaei@isibrno.cz

Highlights

- A novel self-supervised deep learning method for quantifying metabolite concentrations in MR spectroscopy signals.
- Providing a unique opportunity to quantify complex-valued MRS data in the time domain.
- Quantifying MR spectroscopy signals with comparable performance to traditional methods in a significantly shorter amount of time.
- Investigating the impacts of the dataset size and neural network design on our proposed model

Declaration of competing interest

The authors have no conflict of interest to declare.

Credit authorship contribution statement

Amirmohammad Shamaei: Conceptualization, Methodology, Software, Formal Analysis, Writing - Original Draft, Writing - Review & Editing, Visualization.

Jana Starcukova and Zenon Starcuk Jr: Conceptualization, Methodology, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

Abstract (~250/250)

Purpose:

While the recommended analysis method for magnetic resonance spectroscopy data is linear combination model (LCM) fitting, the supervised deep learning (DL) approach for quantification of MR spectroscopy (MRS) and MR spectroscopic imaging (MRSI) data recently showed encouraging results; however, supervised learning requires ground truth fitted spectra, which is not practical. Moreover, This work investigates the feasibility and efficiency of the LCM-based self-supervised DL method for the analysis of MRS data.

Method:

We present a novel DL-based method for the quantification of relative metabolite concentrations, using quantum-mechanics simulated metabolite responses and neural networks. We trained, validated, and evaluated the proposed networks with simulated and publicly accessible in-vivo human brain MRS data and compared the performance with traditional methods. A novel adaptive macromolecule fitting algorithm is included. We investigated the performance of the proposed methods in a Monte Carlo (MC) study.

Result:

The validation using low-SNR simulated data demonstrated that the proposed methods could perform quantification comparably to other methods. The applicability of the proposed method for the quantification of in-vivo MRS data was demonstrated. Our proposed networks have the potential to reduce computation time significantly.

Conclusion:

The proposed model-constrained deep neural networks trained in a self-supervised manner can offer fast and efficient quantification of MRS and MRSI data. The rapid processing of large datasets such as high-resolution MRSI datasets, which may have thousands of spectra, is a crucial aspect of clinical practice.

Keywords (6/6)

MR spectroscopy, inverse problem, deep learning, machine learning, convolutional neural network, metabolite quantification.

1 Introduction

Accurate spectral analysis of in vivo magnetic resonance spectroscopy (MRS) data is complicated by overlapping of resonance lines from different metabolites, which precludes or inhibits accurate identification of low-concentration metabolites, and also by the presence of a broad background of rapidly decaying macromolecule (MM) and lipid signals (Kreis et al., 2021; Robin A. de Graaf, 2019).

Linear combination model (LCM) fitting, peak fitting, and peak integration are the three most often used methods for spectral analysis of an MRS signal (Near et al., 2020a). In LCM fitting, each metabolite in the spectrum is represented by a "basis spectrum," which is a description of the spectral shape of an individual metabolite (Near et al., 2020a; Poullet et al., 2008; Robin A. de Graaf, 2019; Stagg and Rothman, 2013). Contrary to LCM fitting, peak fitting uses a simple lineshape model function to fit isolated peaks within a spectrum. Peak fitting is highly dependent on the prior knowledge of the parameters of peaks, and imposing a large amount of prior knowledge may be burdensome in crowded spectra like ^{1}H-MRS of the brain due to the excessive number of metabolites and peaks per metabolite (Near et al., 2020a; Poullet et al., 2008; Robin A. de Graaf, 2019; Stagg and Rothman, 2013).

LCM fitting is the expert-recommended approach (Near et al., 2020b) owing to its shown efficacy, adaptability, and relative simplicity of usage. The choice of basis spectra over individual peak components (peak fitting method) reduces the number of model functions necessary to adequately describe the spectrum, which leads to fewer fitting parameters. Basis spectra are realistic since they are derived directly from phantom experiments or numerical simulations (Landheer et al., 2019; Near et al., 2020a; Poullet et al., 2008; Robin A. de Graaf, 2019; Stagg and Rothman, 2013; Starčuk and Starčuková, 2017).

A long-standing interest in LCM fitting has resulted in the development of a variety of fitting techniques, including time-domain and frequency-domain algorithms (Daniel G.Q. Chong et al., 2011; Clarke et al., 2021; Oeltzschner et al., 2020; Poullet et al., 2008; Provencher, 2001; Ratiney et al., 2005; Soher et al., 1994; Wilson et al., 2011).

The recent success of deep learning (DL), one of the latest machine learning approaches, in a variety of tasks, including applications with a low signal-to-noise ratio (SNR) (Goodfellow et al., 2016; Lundervold and Lundervold, 2019), suggests that it might also handle the spectral analysis of an MRS signal. Hatami et al. (Hatami et al., 2018) and Lee et al. (Lee and Kim, 2019) applied supervised DL approaches to metabolite quantification and presented results comparable to conventional LCM approaches. Chandler et al. (Chandler et al., 2019) also applied a supervised DL approach to study metabolite quantification in edited human brain MRS spectra.

Moreover, supervised DL-based approaches have been used for ghosting artifacts detection and removal (Kreis and Kyathanahally, 2018), spectral reconstruction (Lee et al., 2020), automatic peak picking (Klukowski et al., 2018), magnetic resonance spectroscopic imaging (MRSI) spatial resolution enhancement (Iqbal et al., 2019), localized correlated spectroscopy acceleration (Iqbal et al., 2021), and poor-quality spectra identification (Gurbani et al., 2018).

These studies showed results competitive with traditional methods using supervised learning, in which the input and the output were simulated spectra and known values, respectively. The true output values are unknown in in-vivo MRS data. Moreover, a network trained in a supervised manner using simulated data might be prone to overfit training data (Higham and Higham, 2019); thus, any discrepancy between the in-vivo and the simulated training spectra, such as the presence of nuisance peaks, frequency, and phase shifts, and line-broadening, may result in errors in metabolite quantification.

Self-supervised learning may eliminate the drawbacks of supervised learning. Gurbani et al. (Gurbani et al., 2019) presented a self-supervised DL architecture that integrates a convolutional neural network (CNN) with peak fitting. In their approach, a deep autoencoder (DAE), introduced by Hinton et al. (Hinton and Salakhutdinov, 2006) as a specific type of neural network, is used as a framework for unsupervised learning.

A DAE is typically designed to encode the input into a low-dimensional embedding space and then reconstruct the input from the encoding (Goodfellow et al., 2016). A DAE has also been successfully used for metabolite and MM separation in MRS signals, as well as for quantification and noise removal of MRSI signals (Lam et al., 2020; Li et al., 2020).

This paper presents a novel DL-based LCM algorithm for MRS data quantification, which uses the advantages of LCM and the capabilities of DAE in self-supervised learning. We designed a DAE network that can learn in a self-supervised manner to fit a linear combination of the basis spectra to the acquired MR spectrum. The proposed model utilizes the LCM fitting and integrates it into a DAE to estimate relative concentrations of metabolites. The proposed network was trained, validated, and evaluated using simulated MRS data, and in-vivo MRS data obtained from the publicly accessible Big GABA repository (Mikkelsen et al., 2019, 2017). The performance of our proposed model was compared with other LCM packages, namely, QUEST (Ratiney et al., 2005) and FiTAID (Daniel G.Q. Chong et al., 2011), in terms of accuracy and time efficiency.

A common belief among traditional statisticians and DL practitioners is that "*more data is always better*" (Nakkiran et al., 2019). The effects of model complexity, dataset size, and data augmentation were empirically investigated in computer vision and natural language processing (Nakkiran et al., 2019)(Belkin et al., 2019). In this study, we designed an experiment to evaluate the effect of the dataset size and neural network architecture in our proposed model.

2 Material and methods

2.1 Data Sets

2.1.1 The simulated dataset

One significant challenge in metabolite quantification is that the ground truth values are unknown; thus, it is hard to compare the results of different quantification methods. This problem can be addressed by creating a dataset with known ground truth values using prior physical knowledge and basis spectra.

The MRS signal can be described as a linear combination of amplitude-scaled metabolite signals, the baseline, and noise. The simplified model describing a time-domain MRS signal $S(t)$ as a combination of several metabolite profiles is:

$$S(t) = \left(\sum_{m=1}^{\mathrm{M}} A_m X_m(t) + A_b B(t) \right) e^{(\Delta\alpha + 2i\pi\Delta f)t} e^{i\Delta\theta} + \xi(t), \qquad [1]$$

where $i = \sqrt{-1}$ and A_m and $X_m(t)$ are the scaling factor (amplitude) and the model (basis) function (signal) for the m-th metabolite, respectively. $\Delta\alpha$, Δf, and $\Delta\theta$ are the global damping factor, the global frequency shift, and the global phase shift, respectively. M is the number of metabolites. A_b and $B(t)$ is the scaling factor and the signal for MMs, and $\xi(t)$ is noise.

Our simulated dataset, containing 96000 signals, was generated using Eq. 1 and the publicly available metabolite basis set from the ISMRM MRS study group's fitting challenge (Marjańska et al., 2022) (19 metabolites signals (see Table 1) and one MM signal, 3T, PRESS, TE = 30 ms, spectral width = 4000 Hz, 2048 points). The range of parameters (A_m, $\Delta\alpha$, Δf, and $\Delta\theta$) were determined according to the literature (Govindaraju et al., 2000; Robin A. de Graaf, 2019) (listed in Table 1). The parameters were chosen randomly from the defined ranges with a uniform distribution. The mean SNR of the dataset (largest metabolite peak height in the magnitude spectrum to standard deviation of noise in the frequency domain) was set to 67 ± 15 by introducing random complex Gaussian white noise (ξ). Signals in the simulated dataset were

shuffled randomly, and 90% of each dataset was allocated to the training subset, 9% to the validation subset, and the rest 1% to the test subset.

Table 1 Parameters used in the simulated dataset. $A_{(.)}$, $\Delta\alpha$, Δf, and $\Delta\theta$ are the metabolite amplitudes, the global damping factor, the global frequency shift, and the global phase shift, respectively.

Parameter	Min	Max
$A_{(\mathrm{Ala})}$	0.1	1.8
$A_{(\mathrm{Asc})}$	0.2	1.8
$A_{(\mathrm{Asp})}$	1	2
$A_{(\mathrm{Cr})}$	4.5	10.5
$A_{(\mathrm{GABA})}$	1	2
$A_{(\mathrm{Glc})}$	1	2
$A_{(\mathrm{Gln})}$	3	6
$A_{(\mathrm{Glu})}$	6	12.5
$A_{(\mathrm{GPC})}$	0.5	1.8
$A_{(\mathrm{GSH})}$	1.5	3
$A_{(\mathrm{Gly})}$	0.2	1
$A_{(\mathrm{Ins})}$	4	9
$A_{(\mathrm{Lac})}$	0.2	1
$A_{(\mathrm{NAA})}$	7.5	17
$A_{(\mathrm{NAAG})}$	0.5	2.5
$A_{(\mathrm{PCho})}$	0.2	1
$A_{(\mathrm{PCr})}$	3	5.5
$A_{(\mathrm{PE})}$	1	2
$A_{(\mathrm{sIns})}$	0.2	0.5
$A_{(\mathrm{Tau})}$	3	6
A_b	14	17
Δα (s^{-1})	0	15
Δf (Hz)	-3	3
Δθ (˚)	-π/10	+π/10

2.1.2 The in vivo dataset

Data from the public repository Big GABA (Mikkelsen et al., 2019, 2017) were used to demonstrate the applicability of the proposed model for the quantification of in-vivo signals.

We selected 48 single-voxel short-TE PRESS brain datasets (subjects) from 48 healthy volunteers acquired on Siemens scanners from 4 different sites (S1, S5, S6, and S8, 3 Tesla field strength, spectral width = 4000 Hz, 4096 points, TE = 35 ms, TR = 2000 ms; 64 transients; 30 × 30 × 30 mm^3; medial parietal lobe voxel). Water reference scans without water suppression were obtained with similar parameters, except for 8–16 averages. It is well recognized that the sufficient size and diversity of data are crucial for the success of the majority of DL models (Goodfellow et al., 2016). However, rich and adequate datasets are uncommon (Lam et al., 2020) in the field of MRS and MRSI. Data augmentation is a possible technique for simulating realistic data by making slight modifications to a small existing dataset.

A two-step data augmentation was applied. In the first step, 16 transients of 64 transients of each subject dataset were randomly selected. Then, the transients were frequency- and phase-aligned using the spectral registration method (Near et al., 2015; Simpson et al., 2017) and were processed into an averaged signal. Eddy-current effects in the averaged signal were corrected using the corresponding water reference signal. Then, the residual water components were removed from the averaged signal with the Hankel Lanczos singular value decomposition (HLSVD) method (Cabanes et al., 2001). The first step was repeated 64 times, producing 64 signals for each subject and thus producing a new extended subject dataset. (Cabanes et al., 2001)

Totally, 3072 (48 × 64) signals, each representing a corrected 16-transient average, were generated and stacked together to create a new in vivo dataset from all subjects.

In the second step of augmentation, first, each complex signal $S(t)$ from the new in vivo dataset generated in step 1 was rescaled as

$$S(t)_{normalized} = \frac{S(t)}{\max_{\tau}(|S(\tau)|)}. \quad [2]$$

Then five extended subject datasets (320 (5 × 64) signals) were included in a test dataset and excluded from the in vivo dataset. Signals with substantial lipid contamination were removed from the in vivo dataset in the following way: Time-domain signals were Fourier-transformed, then the mean of the magnitude of each signal between 0 and 1.85 ppm was calculated. Finally, based on prior testing on the dataset, signals with a mean value above six were labeled as strongly contaminated signals and discarded from the dataset. Further, we denote this dataset as the filtered dataset (2624 signals).

In the following, ten signals were generated from each signal of the filtered dataset, such that signals were apodized by dampings corresponding to Lorentzian linewidths drawn from a uniform distribution over [0 Hz, 2 Hz]. Then artificial frequency and phase offsets were drawn from a uniform distribution in the ranges of −3 to 3 Hz and -9° to 9°, respectively, and added to apodized signals (ranges were extracted from in vivo brain spectra). The final training dataset consisted of 28864 (10 × 2624 + 2624) signals. The mean SNR of the training dataset was set to 256 ± 72 by introducing random complex Gaussian noise with a uniform spectral power. Finally, all signals were dedicated to the training set.

To decrease the complexity of the network (number of weights) and the contribution of noise, which typically predominates in the later part of time domain signals (FIDs), the FIDs were shortened to maintain the first 1024 points.

2.2 The deep model

2.2.1 The proposed DAE

DAE is a sort of deep artificial neural network designed to learn an efficient data coding in a self-supervised way (Goodfellow et al., 2016; Hinton and Salakhutdinov, 2006). The basic idea behind DAEs is to utilize the input data as the target that should be reconstructed in the output layer (Goodfellow et al., 2016). A DAE is typically composed of two parts: an encoder and a decoder. The encoder function $\boldsymbol{h} = f(x)$ maps the n-dimensional input vector ($\boldsymbol{x} \in R^n$) to the n'-dimensional latent vector ($\boldsymbol{h} \in R^{n'}$), while the decoder function $\hat{\boldsymbol{x}} = g(h)$ aims to reconstruct the n-dimensional output vector ($\hat{\boldsymbol{x}} \in R^n$) from the latent space representation. Typically, the latent space representation has significantly lower dimension than the input ($n' \ll n$).

Interestingly, the LCM approach toward the quantification of MRS signal can be seen as a DAE, where the few parameters of a model are the interpretable latent space parameters, and the metabolites basis set and prior knowledge are the decoder. For the DAE to be usable for quantifying MRS signals, its latent-space representation must be interpretable as metabolite concentrations. Therefore, it is natural to desire that it includes parameters such as all individual relative metabolite concentrations (amplitudes), damping factors, frequencies, and phases. However, a DAE with the typical architecture (Hinton and Salakhutdinov, 2006) is not useful for quantifying MRS signals.

To this end, a convolutional encoder–model decoder (Gurbani et al., 2019) was employed, incorporating a parametric analysis into a DL model. Our proposed DAE consisted of a conventional encoder and a model decoder. The encoder was composed of sequential layers of convolutional layers (Fukushima, 1980; Schmidhuber, 2015), Gaussian error linear units (GELU) (Hendrycks and Gimpel, 2020), and a fully connected (FC) layer (Goodfellow et al., 2016; Schmidhuber, 2015). Furthermore, a Softplus function (Goodfellow et al., 2016) was applied to estimated amplitudes to ensure that the relative metabolite concentrations (amplitudes) were non-negative. The model decoder was composed of a model function (Eq. 1) that

reconstructs the representation of the input signal using the assigned latent space parameters and basis spectra. The proposed DAE architecture is depicted in Figure 1a. The encoder computes four types of parameters: amplitude, frequency, phase, and damping. The input and output of the proposed DAE were complex signals in the time domain.

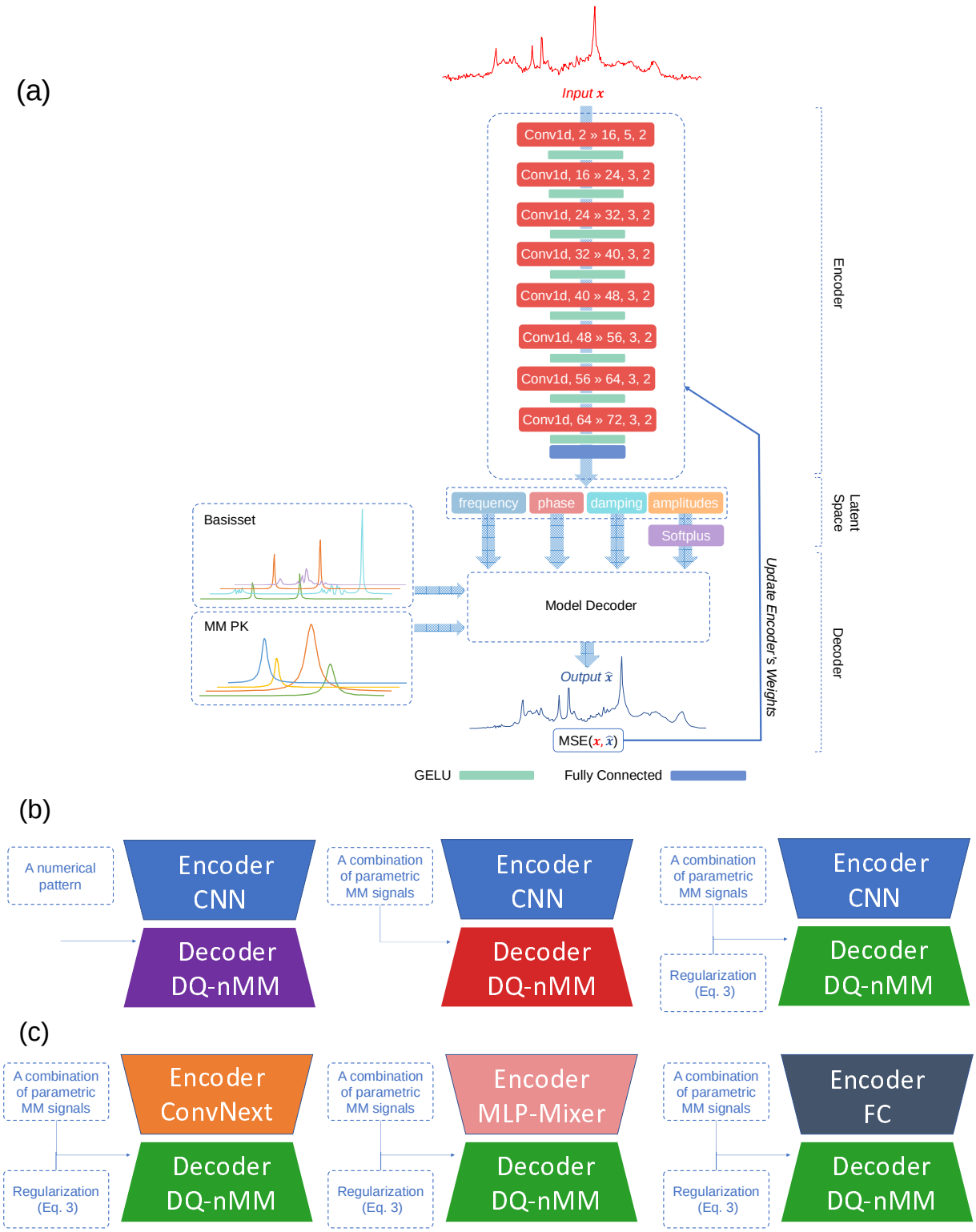

Figure 1. (a) Illustration of the proposed convolutional encoder–model decoder network. The input of the network is a complex signal (x) in the time domain, which

is fed to the encoder. The encoder consisted of eight convolutional blocks and an FC layer (see details in Supporting Information Table S2). A convolutional block is composed of a 1D convolution (Conv1d) layer followed by a GELU layer. The model decoder reconstructs the output signal ($\hat{x}$) using Eq. 1, basis set, and MM prior knowledge. The DAE was trained to encode the input vector x in the time domain into parameters (the amplitudes, damping factor, resonance frequency, and zero-order phase of the basis spectra) that can be used to reconstruct the output vector $\hat{x}$ in the time domain. The proposed network is trained by minimizing the mean square error (MSE) between x and $\hat{x}$.

(b) Illustration of three proposed models (DQ-nMM, DQ-pMM, and DQ-rpMM).

(c) Illustration of three variants of the DQ-rpMM model with different encoder architectures (ConvNext, MLP-Mixer, and our proposed network with FC layers).

CNN, Convolutional Neural Network; Conv1d, (in channels » out channels, kernel size, stride); FC, Fully Connected; GELU, Gaussian Error Linear Unit; MM PK, Macromolecule prior knowledge; DQ-nMM, deep learning-based quantification using a numeric MM; DQ-pMM, deep learning-based quantification using parametric MM; DQ-rpMM, deep quantification using parametric MM and a regularization term.

Training our proposed network is a self-supervised learning task that does not require ground truth values. It minimized the differences between the original input and the consequent reconstruction. In each iteration step of training, the parameters of the encoders were adjusted according to the gradient of the loss function with respect to the given parameters of the model ($A_m, A_b,\ \Delta\alpha, \Delta f$, and $\Delta\theta$). The mathematical representation of training can be written as follows:

$$\widehat{w} = \arg\min_{w} L(x, \hat{x}(w)), \qquad [3]$$

$$\frac{\partial L(x,\hat{x})}{\partial w_i} = \sum_{\rho \in P} \frac{\partial L(x,\hat{x})}{\partial \hat{x}} \cdot \frac{\partial \hat{x}}{\partial \rho} \cdot \frac{\partial \rho}{\partial w_i}, \quad [4]$$

where $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ are the n-dimensional input vector ($\boldsymbol{x} \in C^n$) and the n-dimensional output vector ($\hat{\boldsymbol{x}} \in C^n$), respectively. Eq. 3 is the objective function of training, L is the loss function, w_i is an element of the parameters set (weights and biases) of the encoder, and $P = \{A_1, A_2, \ldots, A_M, A_b, \Delta\alpha, \Delta f, \Delta\theta\}$. Gradients were computed with the help of Pytorch using its automatic gradient computation (Paszke et al., 2019), which is a reverse automatic differentiation system.

2.2.2 Modeling of macromolecules

Short-TE MRS signals contain a contribution of MMs that overlap with metabolites (Cudalbu et al., 2021). Thus, MM modeling or MM removal from the signal should be included in the spectra analysis.

The presence of MMs may be accounted for in a variety of ways (Near et al., 2020b). MMs can be included as a numerical pattern (found, for instance, from metabolite-nulling inversion recovery measurements) (D G Q Chong et al., 2011). Another approach is to include a set of parametric patterns in the spectral fitting algorithm (Near et al., 2020a; Robin A. de Graaf, 2019).

We implemented both approaches to MM handling in the proposed DAE, which resulted in two models. The first model (referred to as deep learning-based quantification using a numeric MM pattern [DQ-nMM]) includes a MM signal as a numerical pattern (D G Q Chong et al., 2011) in Eq. 1. The second model (referred to as deep learning-based quantification using parametric MM

components [DQ-pMM]) includes a MM signal as a combination of parametric MM components in Eq. 1 (the detailed model is provided in Supporting Information Text S1).

However, the increased number of fitted parameters without constraints increases the risk of over-parametrization [42] in the DQ-pMM model. Another approach to handling MM is to remove the initial part of the FID (Poullet et al., 2008), which eliminates MM signals thanks to their decay being faster than that of metabolites. Nevertheless, the disadvantages of this approach are neglecting important information in the early part of the time domain signal and difficulty in determining the number of removed initial points (Poullet et al., 2008).

To overcome these disadvantages and handle the over-parametrization problem in the DQ-pMM model, we proposed a novel approach for MM modeling, which resulted in the third model (referred to as deep learning-based quantification using parametric MM components and a regularization term [DQ-rpMM]).

DQ-rpMM includes a combination of parametric MM components in Eq. 1 and adds a regularization term (Eq. 5) to the loss function.

$$\left\| \left(\left(x_{input}(t) - \hat{x}_{metabolites}(t) \right) \left(\frac{1 + \mathrm{sgn}[t - t_r]}{2} \right) \right) \right\|_2^2 + \lambda \| t_r \|_1 \,, \qquad [5]$$

where sgn is a sign function, its shift t_r is a parameter determined by the network, x_{input} and $\hat{x}_{metabolites}$ are the input signal and the reconstructed signal from the metabolite components, respectively. Parameter λ tunes the significance of the regularization term. Brackets $\|.\|_2^2$ and $\|.\|_1$ represent squared L2-norm and L1-norm, respectively.

The first term promotes consistency between $\hat{x}_{metabolites}$ and x_{input}. Intuitively, the cutoff time t_r should increase to a point where the contribution of MM is low, and the contribution of metabolites is reasonably high. Excessive cutoff of useful signal (that would lead to fitting the non-informative end tail of $\hat{x}_{metabolites}$ to the noisy part of x_{input}) should be prevented by the contradictory force of the second term. The architectures of the proposed models are depicted in Figure 1b.

2.3 Accuracy analysis (for the simulated data set)

For the simulated data set, in which the true amplitudes were known, we compared the results, i.e., the estimated amplitudes, obtained with our approach with those obtained by traditional LCM methods using mean absolute percentage error (MAPE) and the coefficient of determination (R^2) metrics. It has been reported that R^2 is more informative than MAPE (Chicco et al., 2021). In fact, MAPE does not reveal much information about the effectiveness of regression with regard to the distribution of the ground truth values.

2.4 Monte Carlo analysis

Monte Carlo (MC) studies were carried out to investigate the bias and the standard deviation of the proposed DQ-nMM estimates. A signal without noise was generated using Eq. 1 with known global frequency (2 Hz), phase (45˚), and damping (5 Hz), and known amplitudes of metabolites and MMs. Second, 256 realizations of a normally distributed random complex noise were added to the signal such that SNR was in the range of ~50.

Finally, the amplitudes of metabolite components were estimated for the 256 signals using the network trained with the simulated dataset, and the estimation errors (MAPE) were measured.

2.5 Implementation Details and Training

All steps were run on a computer with a dual EPYC 7742 (2×64 cores) processor and one graphics processing unit (NVIDIA A100 40 GB). The DAE was implemented in Python with the help of the Pytorch lightning interface (Lightning and Lightning, 2020; Paszke et al., 2019).

The initial architecture of the network and the training parameters were optimized using the Bayesian Optimization HyperBand algorithm (Falkner et al., 2018) with the help of the Tune framework (Liaw et al., 2018). The optimization details are given in Supporting Information Text S2. All training was performed using the mean-squared error loss (MSE) and the Adam optimizer (Kingma and Ba, 2015) with a batch size of 128, a learning rate of 0.001, and 200 epochs. An early-stopping strategy (Paszke et al., 2019) was performed by monitoring the MSE of the validation subset at the end of every epoch and stopping the training when no improvement was observed in 50 epochs. As reported (Loshchilov and Hutter, 2016), reducing the learning rate is beneficial when learning becomes stagnant. The learning was reduced twice by a factor of 10 during training.

All proposed models (DQ-nMM, DQ-pMM, and DQ-rpMM) were tested on the simulated dataset. Moreover, The DQ-rpMM model was tested on the in vivo dataset. For the simulated dataset, the basis set and the numeric pattern of MMs from the ISMRM MRS study group's fitting challenge (Marjańska et al., 2022) were utilized in the model decoder. The provided numeric pattern of MMs was parametrized using AMARES (Vanhamme et al., 1997) algorithm and utilized in the models with parametrized MMs, i.e., DQ-pMM and DQ-rpMM (the details of the parametrization are provided in Supporting Information Table S1). For the in vivo dataset (Big GABA), the publicly available metabolite basis set (Zöllner et al., 2021) consisted of 19 metabolites was used in the model decoder: alanine (Ala), aspartate (Asp), creatine (Cr), negative creatine methylene (-CrCH2), γ-aminobutyric acid (GABA), glycerophosphocholine (GPC), glutathione (GSH), glutamine (Gln), glutamate (Glu), water (H2O), myo-inositol (mI), lactate(Lac), NAA, N-acetylaspartylglutamate (NAAG), phosphocholine (PCho), PCr, phosphoethanolamine, scyllo-inositol, and taurine. MM components (13 gaussian lineshapes, previously reported in (Birch et al., 2017)) were used to generate the lipid and MM basis signals.

2.6 Evaluation

All proposed models were validated using 128 signals selected randomly from the test signals of the simulated dataset, and the DQ-rpMM model was evaluated using the test subset of the in vivo datasets (320 signals).

Based on the results and interpretation of the fitting challenge conducted by the MRS study group of ISMRM (Marjańska et al., 2022), in this work, two LCM fitting methods were used, namely, QUEST (Ratiney et al., 2005; Stefan et al., 2009) and FiTAID (Daniel G.Q. Chong et al., 2011), in order to compare the results obtained for the simulated data and determine whether our proposed models are generally comparable to existing fitting procedures.

The purpose of utilizing QUEST was for its optimal performance in the challenge for the artifact-free healthy-brain-like dataset (Marjańska et al., 2022). FiTAID performed close to the top in the challenge (Marjańska et al., 2022) and also offers fitting in both time and frequency domains, allowing for a wider comparison.

The simulated test signals were processed using the QUEST method (the QUASAR plugin in the jMRUI software package (Stefan et al., 2009)) in two distinct ways. The first procedure (labeled as QUEST) included the numeric MM signal in the basis set. In the second procedure (labeled as QUEST-Subtract), the MM signal was not included in the basis set but estimated with the QUEST subtract technique (Ratiney et al., 2005) that was used to quantify the metabolite signals (the first 100 points were cut off).

The simulated test signals were processed using the FiTAID software in the time (labeled as FiTAID [time]) and frequency (labeled as FiTAID [freq.]) domains over specified ranges (TD = [1..1024]) points, FD = [0,5] ppm). Eq.1 were utilized as a model in the QUEST and the FiTAID software for quantification.

2.7 Data size effect

In this experiment, five variants of the simulated dataset with different sizes (1000, 12000, 24000, 96000, and 384000 signals) were generated using the same procedure mentioned for generating the simulated dataset. 90% of each dataset was allocated to the training subset and

10% to the validation subset. Then, the first proposed model (DQ-nMM) was trained using each training dataset. The performance of each dataset in terms of R^2 of each metabolite and the reconstruction loss (MSE) was monitored using its validation subset during training.

2.8 The regularization term effects

The significance of the regularization term (i.e., λ) in Eq. 5 determines from which time point, $\hat{x}_{metabolites}$ and x_{input} must be consistent. An experiment was designed to address the effect of λ on the performance of the proposed model. DQ-rpMM methods were trained seven times with varying λ ($\{0,1,2,3,4,6,8\}$). All runs used the simulated dataset with a size of 24000 (90% of each dataset was allocated to the training subset and 10% to the validation subset) and were trained for 65000 iterations. All other training parameters were identical to those indicated in the Implementation Details section.

2.9 The architecture effects

To test the applicability of successful machine-vision architectures, we used two cutting-edge architectures, ConvNext (Liu et al., 2022) and MLP-Mixer (Tolstikhin et al., 2021), as the encoder of the DQ-rpMM model, which resulted in two variants of the DQ-rpMM model.

To test the effect of convolutional layers on the performance of our proposed models, we substituted FC layers for the convolutional layers in the DQ-rpMM model, which resulted in the third variant of the DQ-rpMM model.

We modified ConvNext (Liu et al., 2022) and MLP-Mixer (Tolstikhin et al., 2021) architectures to process 1D signals. The architectures of the proposed variants are depicted in Figure 1c. The details of ConvNext, MLP-Mixer, and our proposed network with FC layers architectures are provided in Supporting Information Table S3, S4, and S5.

These three variants of the DQ-rpMM model were trained for 100 epochs using the simulated dataset and the parameters mentioned in section 2.4 for DQ-rpMM, except that the initial learning was decreased to 1×10^{-4}.

3 Results

Table 2 and Figure 2 illustrate a comparison of our DL-based proposed models and traditional methods using the test signals of the simulated dataset. On average, DQ-nMM produced better performance in terms of R^2 and MAPE than QUEST and FiTAID (time and freq.). QUEST-Subtract showed the worst performance in terms of R^2 and MAPE. DQ-rpMM presented a slightly better performance compared with DQ-pMM in all terms. The DQ-pMM method with parametrized MM showed poor performance.

DQ-nMM outperformed other methods in the quantification of most metabolite concentrations. DL-based methods showed an excellent performance in the estimation of Cr concentration. However, DQ-pMM and DQ-rpMM struggled to estimate NAA concentration. In general, all methods reported relatively high MAPEs in the estimation of GPC, PCho, Tau, and NAAG concentrations. All methods showed a poor performance in estimating low-concentration metabolites, for example, Ala, Asp, GABA, Gly, and PE. The effect of damping, frequency, phase and SNR of spectra on quantification accuracy is provided in Supporting Information Figure S1.

The training time for DQ-nMM, DQ-pMM, and DQ-rpMM were 40.7, 73.9, and 89.6 minutes, respectively. The processing time per signal for DQ-nMM, DQ-pMM, and DQ-rpMM were 0.031, 0.507, and 0.517 milliseconds, respectively. The processing time per signal for FiTAID (time), FiTAID (freq.), QUEST, and QUEST-Subtract were 2500, 2500, 560, and 1518 milliseconds, respectively.

Figure 2 illustrates scatterplots between estimated and ground truth values for eight metabolites (Cr, Gln, GPC, Glu, Ins, NAA, NAAG, PCho, PCr, and Tau). The values of R^2, MSE, and t_r during training are reported in Supporting Information Figure S3. DQ-nMM revealed a very good performance in estimation of Cr (R^2=0.88), Glu (R^2=0.68), GPC (R^2=0.67), Ins (R^2=0.71), NAA

Table 2. Testing results of the simulated dataset using DQ-nMM, DQ-pMM, DQ-rpMM, QUEST, and FiTAID. R^2 and MAPE are calculated between the true and estimated values. All metabolites and seven methods are reported in rows and in columns, respectively.

MAPE, mean absolute percentage error; R^2, coefficient of determination; MM, macromolecule; DQ-nMM, deep quantification using numeric MM; DQ-pMM, deep quantification using parametric MM; DQ-rpMM, deep quantification using parametric MM and a regularization term; QUSET (Sub.), Subtract-QUEST; FiTAID (Time), FiTAID with the time-domain fit; FiTAID (freq.), FiTAID with the frequency-domain fit.

(R^2=0.98), PCho (R^2=0.62), PCr (R^2=0.54) and Tau (R^2=0.66) concentration and a good performance in estimation of NAAG (R^2=0.46). Interestingly, DQ-nMM outperformed QUEST in estimation of GPC (R^2=0.10), Ins (R^2 =0.35), PCho (R^2 =0.10), PCr (R^2 =0.19), and Tau (R^2 =0.42).

	DQ-nMM		DQ-pMM		DQ-rpMM		FITAID (Freq.)		FITAID (Time)		QUEST		QUEST (Sub.)	
	R^2	MAPE (%)	R^2	MAPE (%)	R^2	MAPE (%)	R^2	MAPE (%)	R^2	MAPE (%)	R^2	MAPE (%)	R^2	MAPE (%)
Ala	**-0.020**	87.87	-2.850	109.19	-1.680	98.43	-1.690	90.19	-1.370	**84.46**	-1.540	91.07	-2.370	100.99
Asp	**-11.800**	**47.52**	-105.510	149.18	-125.810	166.13	-24.720	80.52	-24.400	80.83	-22.820	78.58	-106.810	149.09
Cr	**0.880**	**6.02**	0.730	9.07	0.760	8.55	0.800	7.85	0.690	8.43	0.830	7.12	0.350	15.45
GABA	**-3.890**	**33.04**	-71.350	101.61	-57.360	101.28	-26.130	99.65	-26.270	100	-10.070	52.73	-72.010	146.23
Glc	**-4.750**	**32.97**	-16.150	70	-14.550	65.5	-9.270	51.01	-8.720	50.07	-8.860	48.17	-22.980	78.23
Gln	**-0.530**	**19.11**	-2.820	31.1	-3.180	33.49	-2.020	25.95	-1.940	25.51	-1.840	24.54	-6.710	47.05
Glu	**0.680**	**8.88**	-2.010	29.88	-1.560	27.31	0.380	13.01	0.440	12.13	0.550	11.05	-1.270	24.54
Gly	**-2.420**	**41.69**	-32.730	131.96	-14.910	93.06	-19.870	94.65	-15.060	100	-18.970	92.7	-48.420	138.09
GPC	**0.670**	**16.27**	0.230	24.01	0.400	21.97	0.030	27.32	0.030	27.39	0.100	26.06	-0.220	29.8
GSH	**-0.290**	**17.47**	-3.490	49.55	-5.960	43.34	-1.080	22.68	-0.920	21.75	-0.880	21.48	-2.930	31.21
Ins	**0.710**	**8.3**	0.210	14.89	0.450	12.06	0.340	12.8	0.600	10.57	0.350	12.88	-0.260	17.84
Lac	**-0.470**	**41.37**	-4.940	70.78	-4.340	63.8	-5.530	75.75	-6.090	75.05	-5.970	78.09	-7.440	86.95
NAA	**0.980**	**2.96**	0.560	12.17	0.570	11.55	0.960	3.5	0.950	3.92	0.970	3.36	0.510	14.38
NAAG	**0.460**	**19.15**	-5.310	78.63	-4.000	68.75	0.240	25.84	0.110	27.52	0.300	24.84	-2.190	56.88
PCho	**0.620**	**18.22**	0.340	25.56	0.500	21.05	-0.060	27.02	-0.010	27.15	0.100	25.75	-0.300	31.95
PCr	**0.540**	**9.12**	-0.700	18.07	-0.390	16.63	0.130	12.29	-0.470	13.55	0.190	11.79	0.000	13.4
PE	**-9.260**	**51.21**	-40.080	124.28	-35.000	118.87	-16.850	78.81	-13.710	69.89	-16.180	76.5	-95.470	157.56
sIns	**-2.370**	**26.94**	-2.340	27.1	-2.760	27.54	-4.080	35.48	-4.420	36.96	-4.000	35.05	-8.210	51
Tau	**0.660**	**14.91**	-0.230	30.51	0.030	26.62	0.450	18.78	0.450	18.84	0.420	19.21	-2.340	47.7

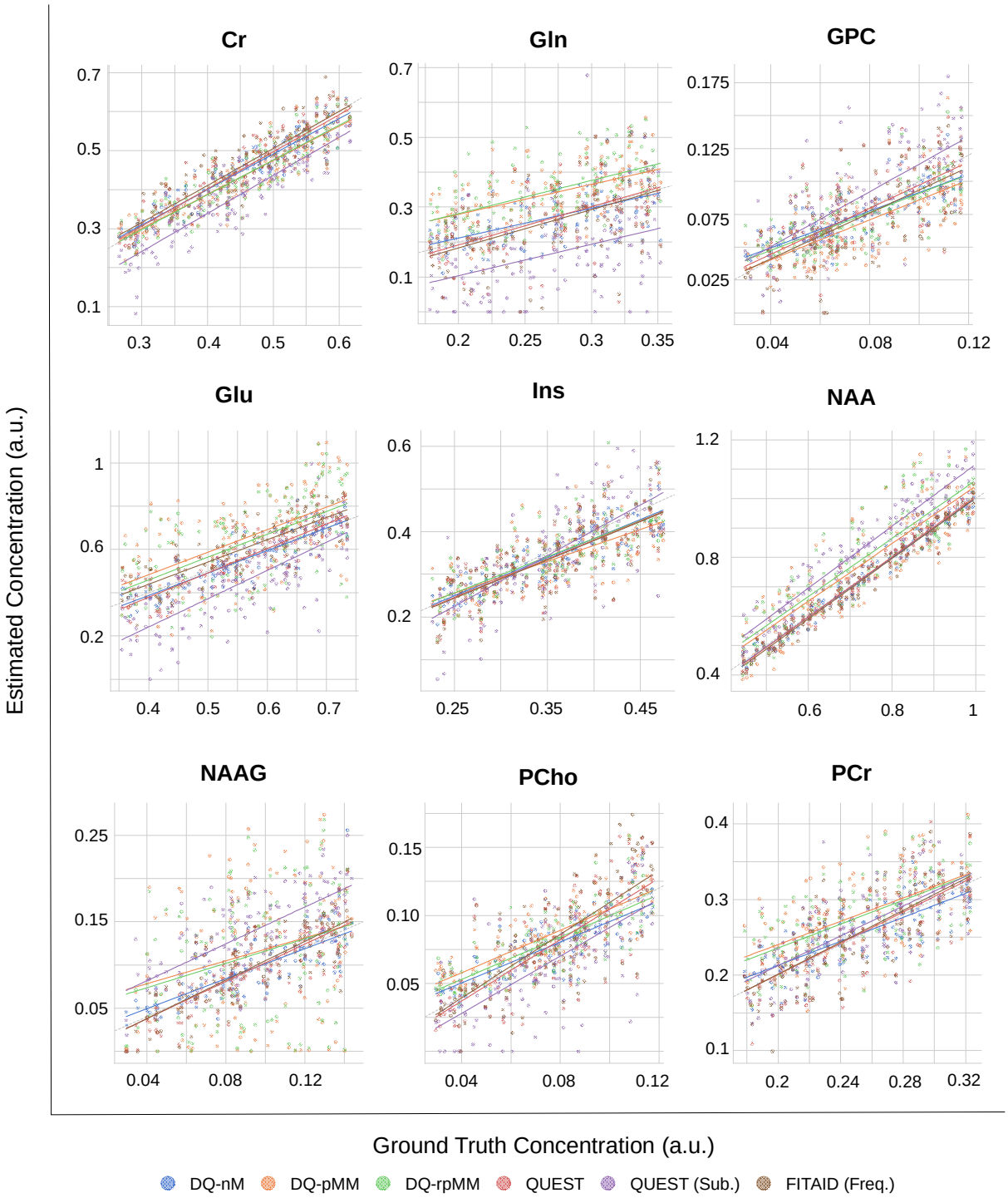

Figure 2. Testing the results of the simulated dataset. Plots of estimated concentrations against the ground truth concentrations in arbitrary units (a.u.). Algorithms are color-coded.

DQ-nMM, deep learning-based quantification using a numeric MM; DQ-pMM, deep

learning-based quantification using parametric MM; DQ-rpMM, deep quantification using parametric MM and a regularization term; Quest (Sub.), Subtract-QUEST; FiTAID (freq.), FiTAID with frequency domain fit.

Figure 3 illustrates example spectra from the test signals of the simulated dataset quantified by DQ-nMM, DQ-pMM, and DQ-rpMM, along with the ground truth spectra. On visual inspection, DQ-pMM and DQ-rpMM showed higher residual compared with DQ-nMM.

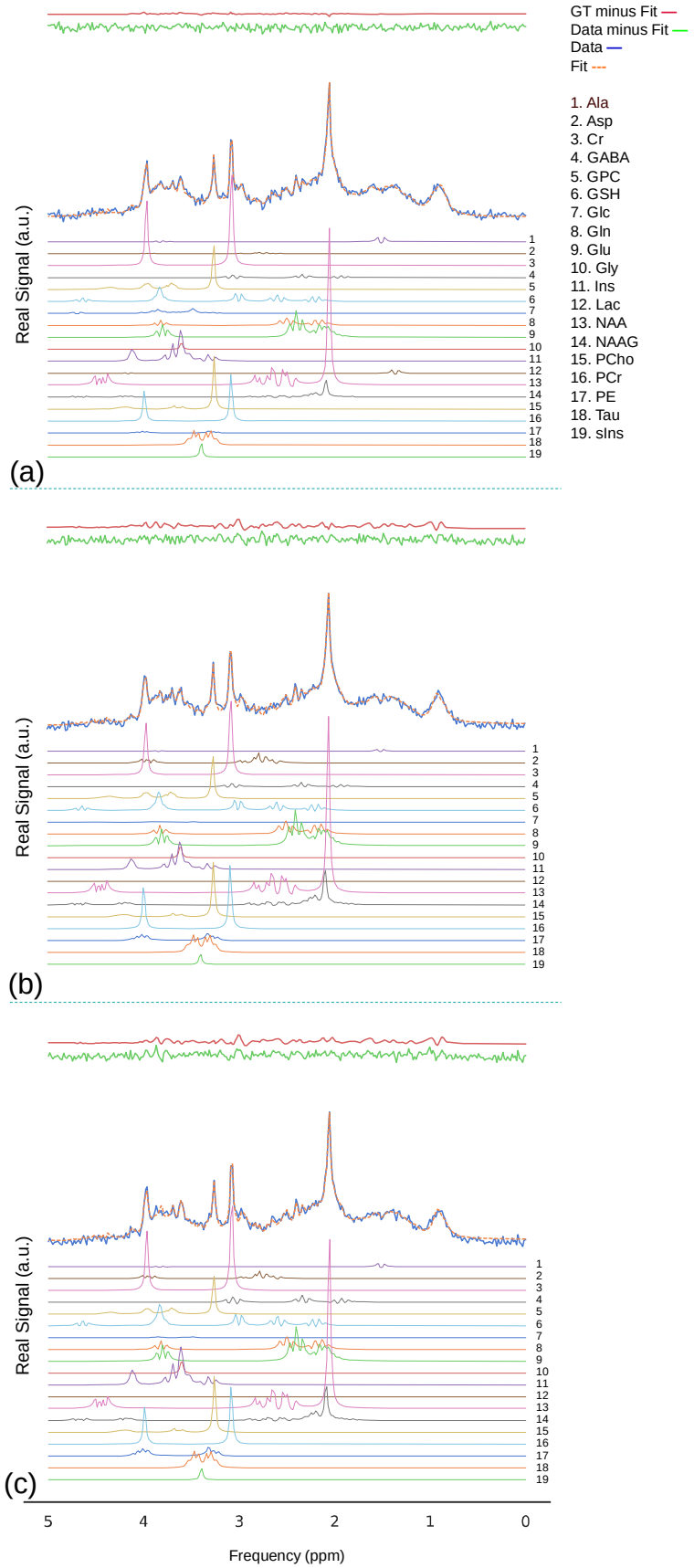
GT minus Fit
Data minus Fit
Data
Fit
1. Ala
2. Asp
3. Cr
4. GABA
5. GPC
6. GSH
7. Glc
8. Gln
9. Glu
10. Gly
11. Ins
12. Lac
13. NAA
14. NAAG
15. PCho
16. PCr
17. PE
18. Tau
19. sIns
Real Signal (a.u.)
(a)
(b)
(c)
5
4
3
2
1
0
Frequency (ppm)

Figure 3. Example spectra from the test subset of the simulated dataset quantified by (a) DQ-nMM, (b) DQ-pMM, and (c) DQ-rpMM. Spectra are color-coded.
GT, ground truth.
DQ-nMM, deep learning-based quantification using a numeric MM; DQ-pMM, deep learning-based quantification using parametric MM; DQ-rpMM, deep quantification using parametric MM and a regularization term.

Figure 4 shows the results of the MC analysis for the DQ-nMM method. Figure 4a shows the distribution of estimated concentrations as histogram plots. In most metabolites, the histogram peak is close to the true value, except that the concentration of NAAG was slightly underestimated. Figure 4b shows a scatter visualization of a joint distribution of estimated frequencies and dampings of spectra, illustrating the bias and variance of the DQ-nMM method for individual metabolites in the test subset of MC analysis. Results revealed robust performance in detecting dampings and frequencies of spectra. Figure 4c shows a heatmap of a correlation matrix representing the correlation between estimated concentrations in the test subset of MC analysis. The heat map highlighted a high correlation between PCr and Cr, and GPC and PCho.

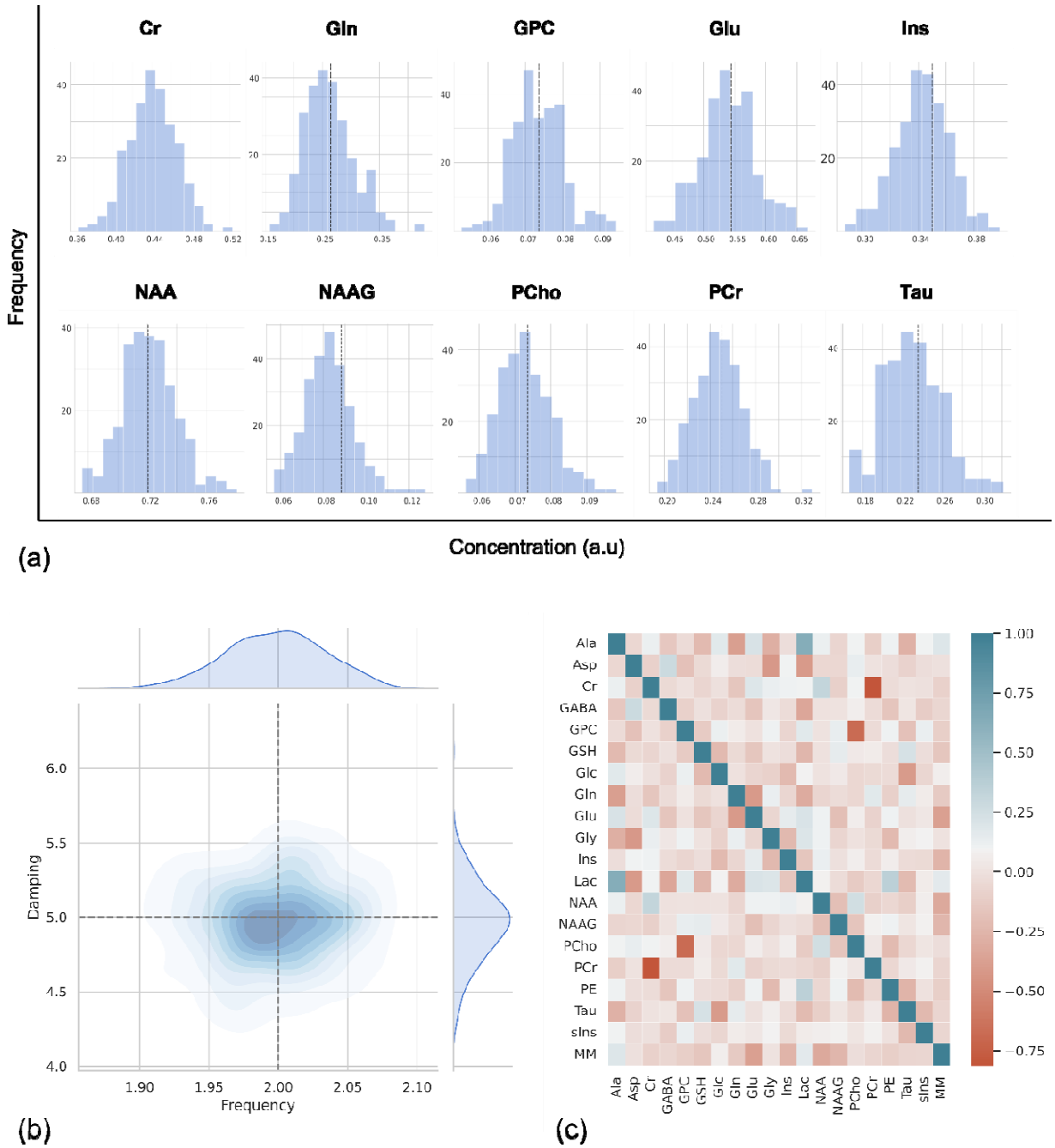

Figure 4. The results of the MC analysis. Spectra are quantified by the DQ-nMM method. (a) Histograms of the estimated concentrations. The vertical lines are true values. (b) A scatter visualization of a joint distribution of estimated frequencies (Hz) and dampings (Hz) of spectra. The dashed vertical and horizontal lines are the true frequency and damping, respectively.

(c) A graphical representation of the correlation matrix representing the correlations

between estimated concentrations. A strip plot of estimated concentration is provided in Supporting Information Figure S2. DQ-nMM, deep learning-based quantification using a numeric MM; MC, Monte Carlo.

Figure 5a shows the effect of dataset size during the training of the DQ-nMM method using the validation subset. The mean R^2 value is calculated across selected metabolites (Cr, GPC, Ins, NAA, PCho, Tau). Figure 5b highlights the effect of λ on the performance of the DQ-rpMM method. The mean R^2 value is calculated across the selected metabolites (Cr, GPC, Ins, NAA, PCho, Tau). Moreover, the trend of t_r in the DQ-rpMM method with seven different λ during training is shown.

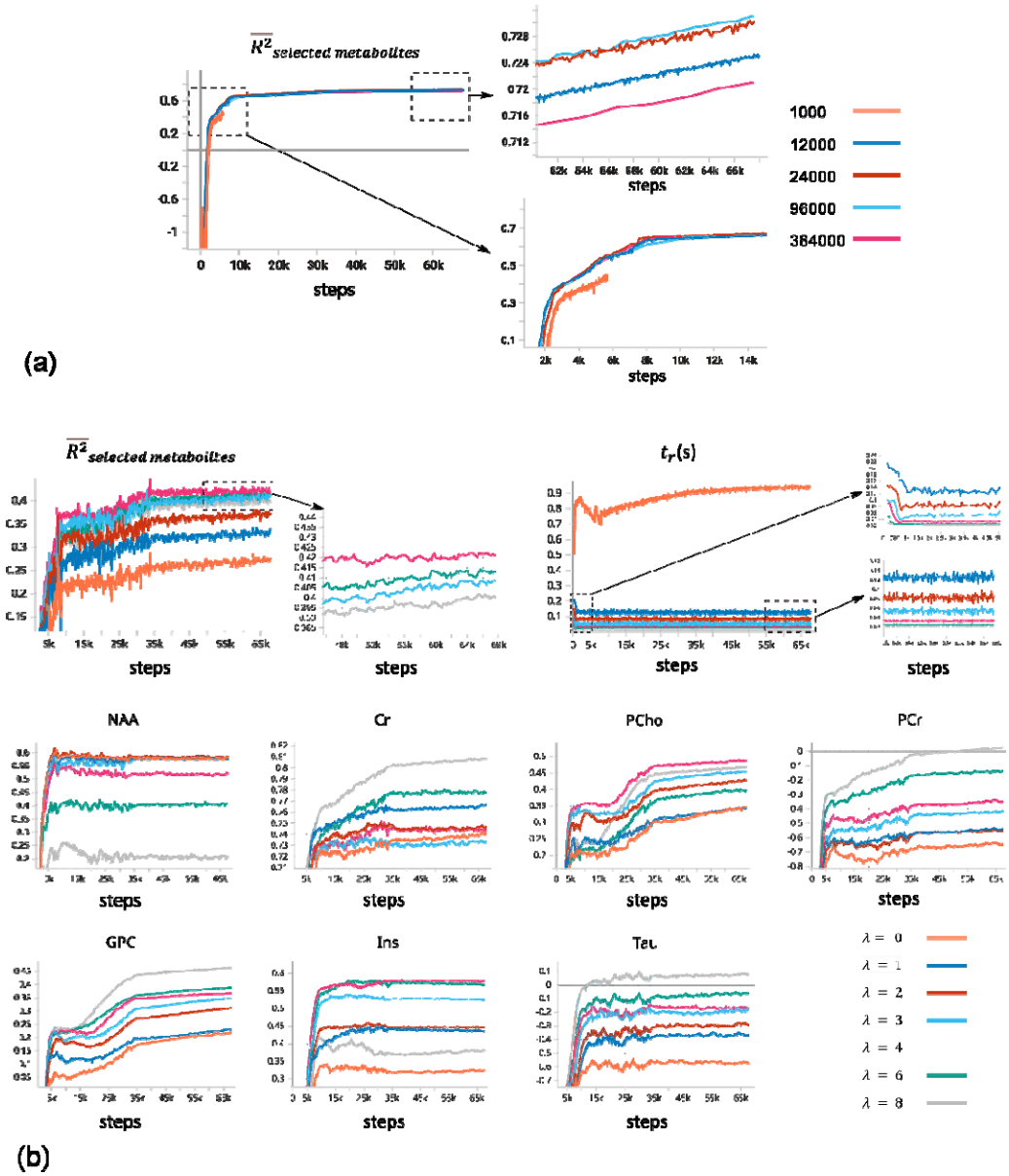

Figure 5. (a) Online monitoring of the mean R^2 value calculated across selected metabolites (Cr, GPC, Ins, NAA, PCho, Tau) for datasets with sizes of 1000 (orange), 12000 (dark blue), 24000 (red), and 96000 (light blue), and 384000 (pink) during training of DQ-nMM. (b) Online monitoring of metrics during training of DQ-rpMM with different . The figure shows that the choice of can affect the performance of quantification of a specific metabolite. For instance, lambda with a high value resulted in a good performance in the quantification of Tau, GPC, PCr, PCho, and Cr, and,

conversely, a poor performance in the quantification of NAA and Ins. The mean R^2 value calculated across the selected metabolites, t_r (in seconds) and the R^2 value for NAA, Cr, PCho, PCr, GPC, Ins, and Tau against training steps are plotted.

DQ-nMM, deep learning-based quantification using a numeric MM; DQ-rpMM, deep learning-based quantification using parametric MM and a regularization term; R^2, coefficient of determination; λ, the significance of the regularization term; t_r, the parameter determined by the network in the regularization term of DQ-rpMM.

The influence of the encoder architecture on the performance of the proposed model (DQ-rpMM) is illustrated in Figure 6. The performance of the proposed networks with convolutional layers was superior to those with fully connected layers.

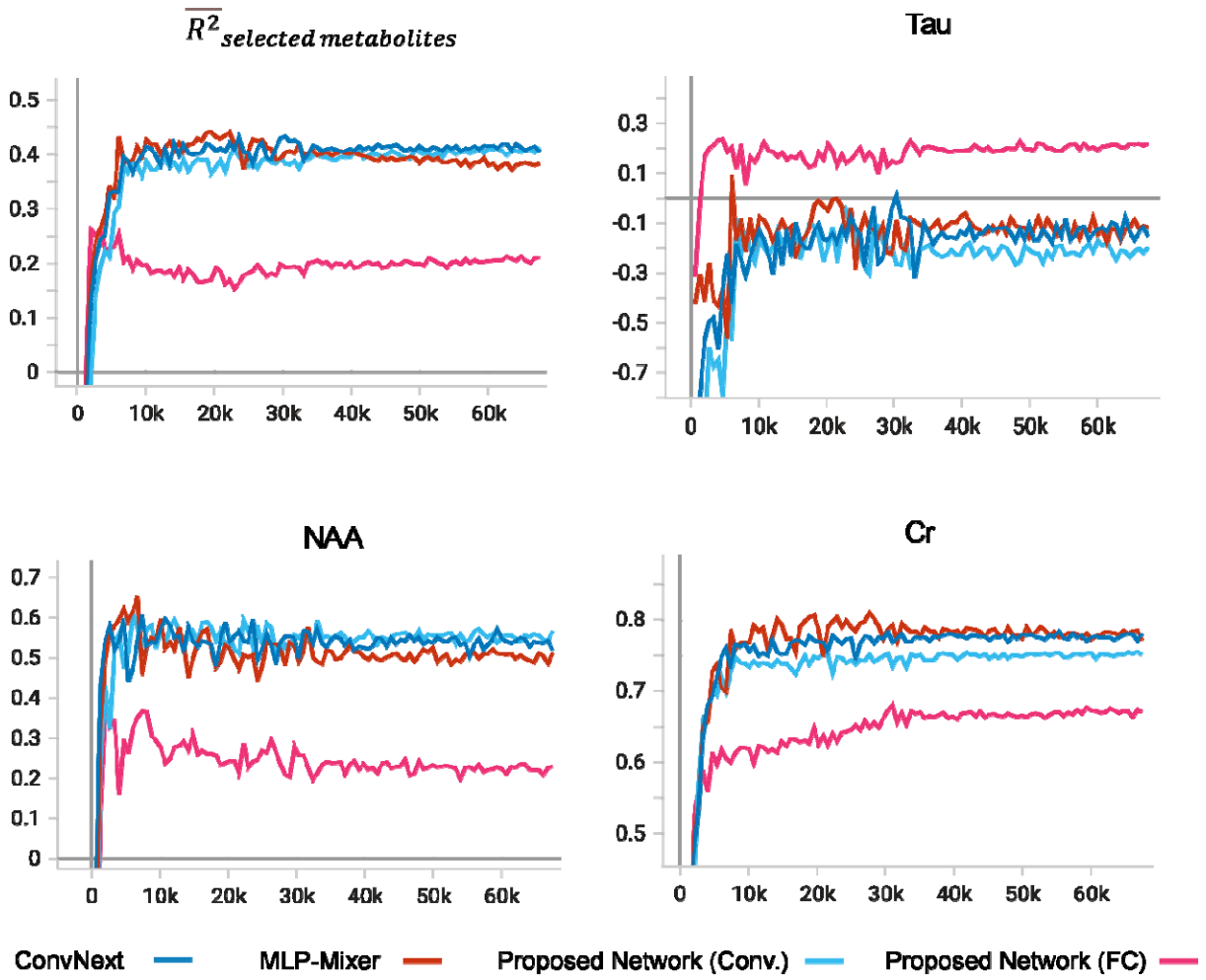

Figure 6. Online monitoring of metrics during training of DQ-rpMM with ConvNext (dark blue), MLP-mixer (red), our proposed convolutional (Conv.) network (light blue), and our proposed fully connected (FC) network (pink). The mean R^2 value calculated across the selected metabolites (Cr, GPC, Ins, NAA, PCho, Tau) and the R^2 values for NAA, Cr, and Tau against training steps are plotted.

DQ-rpMM, deep learning-based quantification using parametric MM components with a regularization term; R^2, coefficient of determination.

Figure 7 illustrates four example spectra from the test subset of the in vivo dataset quantified by DQ-rpMM. The reconstruction capability of our proposed model is clearly shown. The mean, standard deviation (SD), and coefficient of variation (CV) of each metabolite-to-creatine ratio for

five subjects are summarized in Table 3 for in-vivo data. The CV was lowest for Glx (0.9%) and highest for tCho (10.6%) in subject 3. The means of the tNAA-to-Cr ratio were relatively consistent among subjects. Moreover, the SDs of the tNAA-to-Cr ratio were found to be minimal.

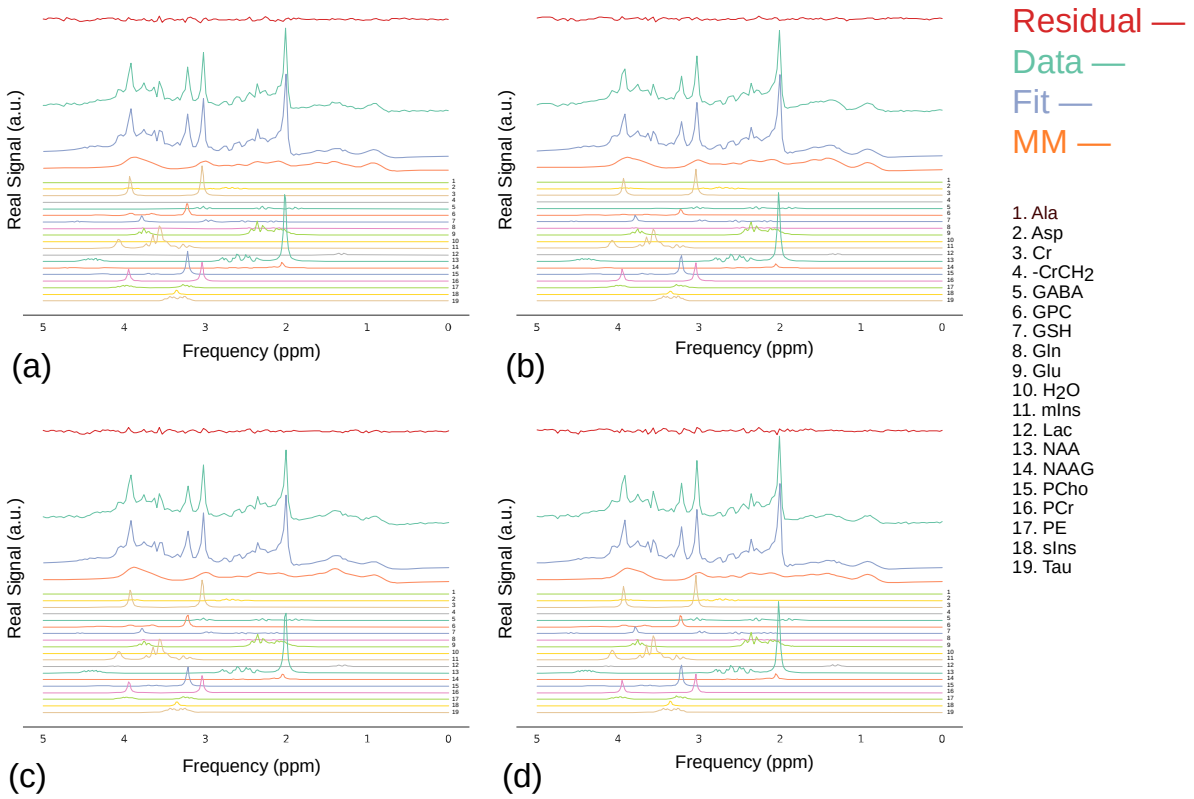

Figure 7. Four example spectra (a,b,c, and d) from the test subset of the in vivo dataset, quantified by DQ-rpMM. Spectra are color-coded. A strip plot of the estimated relative concentration within subjects is provided in Supporting Information Figure S4.

GT, ground truth; DQ-rpMM, deep learning-based quantification using parametric MM and a regularization term.

Table 3. Testing results of the in-vivo dataset using DQ-rpMM. Five metabolites are reported for the five subjects. Each subject has 64 augmented spectra. The mean, standard deviation (SD), and coefficient of variation (CV [SD/mean]) of each metabolite-to-creatine ratio are listed by subjects.

DQ-rpMM, deep quantification using parametric MM and a regularization term.

	[metabolite] / [tCr]														
	mIns			Tau			tNAA			tCho			Glx		
	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV
Subject #1	1.187	0.016	1.3%	0.294	0.01	3.4%	1.69	0.025	1.5%	0.25	0.005	1.8%	1.28	0.016	1.2%
Subject #2	1.078	0.032	3.0%	0.335	0.013	3.9%	1.69	0.017	1.0%	0.259	0.005	1.8%	1.402	0.033	2.4%
Subject #3	0.995	0.033	3.3%	0.377	0.02	5.2%	1.729	0.021	1.2%	0.205	0.022	10.6%	1.424	0.013	0.9%
Subject #4	1.039	0.032	3.1%	0.326	0.032	9.9%	1.616	0.065	4.0%	0.233	0.018	7.9%	1.288	0.073	5.7%
Subject #5	1.1	0.028	2.6%	0.316	0.011	3.4%	1.656	0.043	2.6%	0.237	0.005	2.0%	1.351	0.053	3.9%

4 Discussion

Given recent agreement on using LCM fitting (Near et al., 2020a) and several pioneering research highlighting the possible use of DL in MRS (Gurbani et al., 2019, 2018; Kreis and Kyathanahally, 2018; Lee and Kim, 2019), this work demonstrates that an LCM fitting-based self-supervised DL approach can be used to quantify both simulated and in-vivo MRS data.

No labels or ground truth were required during training since the proposed method maximized the use of prior knowledge in a deep neural network to limit a model solution. The network can identify the contribution of basis metabolite spectra and MM to crowded MRS spectra such as those in Figure 3 and Figure 7. The network can reconstruct the input signal by a linear combination of basis spectra, therefore highlighting the contribution of each basis spectrum to the total spectrum (Figure 3 and Figure 7). Moreover, the difference between the data and fit can be calculated and served as an indicator of whether the fitting operation was successful or not (Stagg and Rothman, 2013).

The network was validated using simulated data in which in-silico ground truth values were available. A comparison between this novel DL-based LCM and traditional LCM fitting algorithms, namely QUEST, QUEST-Subtract, and FiTAID, was performed (Figure 2, Table 2). The results show that the DL-based algorithm can achieve comparable performance with a significantly shorter amount of time for quantification of an MRS signal (0.031 ms) compared with traditional methods (560 ms).

For metabolites such as NAA, Ins, Glu, Tau, and Cr, which have relatively high concentration levels, all methods provided with a numeric MM pattern (DQ-nMM, QUEST, FiTAID) showed good performance (Figures 2 and Table 2). Noteworthy, DQ-nMM showed remarkable result for metabolite such as PCho (R^2 = 0.62, MAPE = 18.2 %) and PCr (R^2 = 0.54, MAPE = 9.12 %) which overlap with GPC (R^2 = 0.67, MAPE = 16.27 %) and Cr (R^2 = 0.88, MAPE = 6.02 %), respectively, while traditional method, such as QUEST struggled with quantification of GPC (R^2 = 0.1, MAPE = 26.06 %), PCr (R^2 = 0.19, MAPE = 11.79 %) and PCho (R^2 = 0.10, MAPE = 25.75 %). Moreover, DQ-nMM showed a good performance in the decomposition of NAA and NAAG, which is notoriously challenging owing to the low concentration of NAAG and spectral overlap (Edden et al., 2007).

DQ-rpMM showed better performance than DQ-pMM and QUEST-Subtract. The use of the regularization terms in DQ-rpMM almost doubled performance compared to DQ-pMM in the quantification of metabolites such as GPC and Ins and PCho in terms of R^2. The reported MAPEs also support our observation (Table 2). Our proposed model DQ-nMM also performed well in an MC analysis where the performance of estimations of metabolites level was found to be very good, and no significant bias was observed (Figure 5).

Along with demonstrating the performance of the proposed approach on simulated data, the method DQ-rpMM was used to conduct quantification in publicly accessible in-vivo MRS data

(Big GABA). Moreover, we demonstrated the generalizability of the proposed model to unseen in-vivo data (Figure 7 and Table 3).

Our results (Table 3) highlighted that CVs for tNAA, mIns, and Glx were consistent among subjects. For tCho and Tau, CVs were less consistent. In general, our proposed model showed stability in the quantification of spectra within subjects.

We observed that the performance of our method could also be affected by the parameters of the network and prior knowledge (Figures 5 and 6). We investigated the influence of deep neural network design by comparing our proposed network to existing architectures for machine vision tasks.

In the field of machine learning, it is typical to estimate the computation cost using FLOPS (floating-point operations per second) (Liu et al., 2022). The computational cost of our CNN was 2.66 M FLOPS. The computational costs (notated with O) of the iterative Levenberg-Marquardt method based on the QR factorization as the linear solver is (Lin et al., 2016): $O(n \times m) + k\left(O(m^3) + O(n \times m) + O(m^3)\right)$ where n is the number of data points and m is the number of model parameters, and k is the number of the Levenberg-Marquardt damping parameters that are being used (Lin et al., 2016). By setting k to 1 for the sake of simplicity, m to 24, and n to 2048, the total cost of Levenberg-Marquardt at every iteration is 1.4 M FLOPS. For example, FiTAID (Daniel G.Q. Chong et al., 2011) requires 100 iterations by default, and the computational cost of a fitting procedure using traditional LCM is above 100 MFLOPS.

As depicted in Figure 6, substituting convolutional layers with FC layers resulted in a substantial drop in performance and an increase in computational cost (4.52 MFLOPs).-The results (Figure 6) indicate that MLP-Mixer with extremely deep designs may perform worse. Furthermore, ConvNext demonstrated the same performance as our proposed convolutional network.

However, ConvNext and MLP-Mixer demanded greater computational power (FLOPs for ConvNext and MLP-Mixer were 710 MFLOPs and 2.4 GFLOPs, respectively).

Because of using a model-decoder as the decoder, our proposed method can calculate the Cramér-Rao lower bounds to estimate uncertainty. However, additional work is required to extend the capability of the proposed method for quantifying MRS data with simultaneous uncertainty estimation.

In this work, we utilized a simplified model for describing a time-domain MRS signal (Eq.1); however, we can expand the capabilities of our proposed DAE to handle more complex models. In the current work, we primarily focused on validating our proposed approaches utilizing in-silico ground truth information and demonstrated its applicability to in-vivo MRS data. We are aware that more examination and comparison with other datasets and methodologies are necessary.

5 Conclusions

In general, our proposed model based on self-supervised DL in combination with LCM fitting provides a unique opportunity to quantify complex-valued MRS data in the time domain. Once trained, the proposed deep neural network can quantify MRS spectra with comparable performance to traditional LCM methods in a significantly shorter amount of time. Thus, our proposed approach could aid in accelerating the quantification of large MRS datasets.

6 List of all Supplementary Information Captions

Supporting Information Figure S1. Scatter plots of the MAPEs in the estimated Cr and NAA concentration versus damping (Hz), frequency (Hz), phase (˚), and SNR of spectra.

Supporting Information Figure S2. The results of MC analysis.

Supporting Information Figure S3. Online monitoring of metrics during training.

Supporting Information Figure S4. Strip plot of estimated relative concentration within subjects.

Supporting Information Table S1. The details of the parametrization of the MM signal.

Supporting Information Table S2. Summary of the proposed network (Convolutional).

Supporting Information Table S3. Summary of the proposed network (MLP-Mixer).

Supporting Information Table S4. Summary of the proposed network (ConvNext).

Supporting Information Table S5. Summary of the proposed network (Fully connected).

Supporting Information Text S1: The detailed model of macromolecules.

Supporting Information Text S2: Bayesian hyper-parameterization.

7 DATA AVAILABILITY STATEMENT

The source code is freely available at [https://github.com/isi-nmr/Deep-MRS-Quantification]. For questions, please contact the authors. The data that support the findings of this study are openly available in Big GABA at https://www.nitrc.org/projects/biggaba/.

8 References

Belkin, M., Hsu, D., Ma, S., Mandal, S., 2019. Reconciling modern machine learning practice and the bias-variance trade-off.

Birch, R., Peet, A.C., Dehghani, H., Wilson, M., 2017. Influence of macromolecule baseline on 1H MR spectroscopic imaging reproducibility. Magn. Reson. Med. 77, 34–43. https://doi.org/10.1002/mrm.26103

Cabanes, E., Fur, Y. Le, Simond, G., Cozzone, P.J., 2001. Optimization of Residual Water Signal Removal by HLSVD on Simulated Short Echo Time Proton MR Spectra of the Human Brain 125, 116–125. https://doi.org/10.1006/jmre.2001.2318

Chandler, M., Jenkins, C., Shermer, S.M., Langbein, F.C., 2019. MRSNet: Metabolite quantification from edited magnetic resonance spectra with convolutional neural networks. arXiv 1–12. https://doi.org/10.48550/arXiv.1909.03836

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, 1–24. https://doi.org/10.7717/PEERJ-CS.623

Chong, D G Q, Kreis, R., Bolliger, C., Boesch, C., Slotboom, J., 2011. Two-dimensional linear-combination model fitting of magnetic resonance spectra to define the macromolecule baseline using FiTAID, a Fitting Tool for Arrays of Interrelated Datasets. Magn Reson Mater Phy 24, 147–164.

Chong, Daniel G.Q., Kreis, R., Bolliger, C.S., Boesch, C., Slotboom, J., 2011. Two-dimensional linear-combination model fitting of magnetic resonance spectra to define the macromolecule baseline using FiTAID, a Fitting Tool for Arrays of Interrelated Datasets. Magn. Reson. Mater. Physics, Biol. Med. 24, 147–164. https://doi.org/10.1007/s10334-011-0246-y

Clarke, W.T., Stagg, C.J., Jbabdi, S., 2021. FSL-MRS: An end-to-end spectroscopy analysis package. Magn. Reson. Med. 85, 2950–2964. https://doi.org/10.1002/MRM.28630

Cudalbu, C., Behar, K.L., Bhattacharyya, P.K., Bogner, W., Borbath, T., de Graaf, R.A., Gruetter, R., Henning, A., Juchem, C., Kreis, R., Lee, P., Lei, H., Marjańska, M., Mekle, R.,

Murali-Manohar, S., Považan, M., Rackayová, V., Simicic, D., Slotboom, J., Soher, B.J., Starčuk, Z., Starčuková, J., Tkáč, I., Williams, S., Wilson, M., Wright, A.M., Xin, L., Mlynárik, V., 2021. Contribution of macromolecules to brain 1H MR spectra: Experts' consensus recommendations. NMR Biomed. 34, 1–24. https://doi.org/10.1002/nbm.4393

Edden, R.A.E., Pomper, M.G., Barker, P.B., 2007. In vivo differentiation of N-acetyl aspartyl glutamate from N-acetyl aspartate at 3 Tesla. Magn. Reson. Med. 57, 977–982. https://doi.org/10.1002/MRM.21234

Falkner, S., Klein, A., Hutter, F., 2018. BOHB: Robust and Efficient Hyperparameter Optimization at Scale.

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36, 193–202. https://doi.org/10.1007/BF00344251

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Govindaraju, V., Young, K., Maudsley, A.A., 2000. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR Biomed. 13, 129–153. https://doi.org/10.1002/1099-1492(200005)13:3<129::AID-NBM619>3.0.CO;2-V

Gurbani, S.S., Schreibmann, E., Maudsley, A.A., Cordova, J.S., Soher, B.J., Poptani, H., Verma, G., Barker, P.B., Shim, H., Cooper, L.A.D., 2018. A convolutional neural network to filter artifacts in spectroscopic MRI. Magn. Reson. Med. 80, 1765–1775. https://doi.org/10.1002/mrm.27166

Gurbani, S.S., Sheriff, S., Maudsley, A.A., Shim, H., Cooper, L.A.D., 2019. Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting. Magn. Reson. Med. 81, 3346–3357. https://doi.org/10.1002/mrm.27641

Hatami, N., Sdika, M., Ratiney, H., 2018. Magnetic resonance spectroscopy quantification using deep learning. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11070 LNCS, 467–475. https://doi.org/10.1007/978-3-030-00928-1_53

Hendrycks, D., Gimpel, K., 2020. GAUSSIAN ERROR LINEAR UNITS (GELUS). arXiv 1–9.

https://doi.org/10.48550/arXiv.1606.08415

Higham, C.F., Higham, D.J., 2019. Deep learning: An introduction for applied mathematicians. SIAM Rev. 61, 860–891. https://doi.org/10.1137/18M1165748

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science (80-.). 313, 504–507. https://doi.org/10.1126/SCIENCE.1127647/SUPPL_FILE/HINTON.SOM.PDF

Iqbal, Z., Nguyen, D., Hangel, G., Motyka, S., Bogner, W., Jiang, S., 2019. Super-Resolution 1H Magnetic Resonance Spectroscopic Imaging Utilizing Deep Learning. Front. Oncol. 9, 1–13. https://doi.org/10.3389/fonc.2019.01010

Iqbal, Z., Nguyen, D., Thomas, M.A., Jiang, S., 2021. Deep learning can accelerate and quantify simulated localized correlated spectroscopy. Sci. Rep. 11, 1–13. https://doi.org/10.1038/s41598-021-88158-y

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations. International Conference on Learning Representations, ICLR.

Klukowski, P., Augoff, M., ZieRba, M., Drwal, M., Gonczarek, A., Walczak, M.J., 2018. NMRNet: A deep learning approach to automated peak picking of protein NMR spectra. Bioinformatics 34, 2590–2597. https://doi.org/10.1093/bioinformatics/bty134

Kreis, R., Cudalbu, C., Graaf, R.A. De, Gasparovic, C., Heerschap, A., Lanz, B., Near, J., Öz, G., Terpstra, M., Tkác, I., Wilson, M., 2021. Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra□: Background and experts ’ consensus recommendations. https://doi.org/10.1002/nbm.4347

Kreis, R., Kyathanahally, S.P., 2018. Deep Learning Approaches for Detection and Removal of Ghosting Artifacts in MR Spectroscopy 863, 851–863. https://doi.org/10.1002/mrm.27096

Lam, F., Li, Y., Peng, X., 2020. Constrained Magnetic Resonance Spectroscopic Imaging by Learning Nonlinear Low-Dimensional Models. IEEE Trans. Med. Imaging 39, 545–555. https://doi.org/10.1109/TMI.2019.2930586

Landheer, K., Swanberg, K.M., Juchem, C., 2019. Magnetic resonance Spectrum simulator (MARSS), a novel software package for fast and computationally efficient basis set simulation 1–13. https://doi.org/10.1002/nbm.4129

Lee, H., Lee, H.H., Kim, H., 2020. Reconstruction of spectra from truncated free induction decays by deep learning in proton magnetic resonance spectroscopy. Magn. Reson. Med. 1–10. https://doi.org/10.1002/mrm.28164

Lee, H.H., Kim, H., 2019. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. Magn. Reson. Med. 82, 33–48. https://doi.org/10.1002/mrm.27727

Li, Y., Wang, Z., Lam, F., 2020. Separation of Metabolite and Macromolecule Signals for 1 H-Mrsi Using Learned Nonlinear Models. Proc. - Int. Symp. Biomed. Imaging 2020-April, 1725–1728. https://doi.org/10.1109/ISBI45749.2020.9098365

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J.E., Stoica, I., 2018. Tune: A Research Platform for Distributed Model Selection and Training.

Lightning, P., Lightning, P., 2020. PyTorch Lightning工具学习 [WWW Document]. GitHub. Note https//github.com/PyTorchLightning/pytorch-lightning. https://doi.org/10.5281/zenodo.3828935

Lin, Y., O'Malley, D., Vesselinov, V. V., 2016. A computationally efficient parallel Levenberg-Marquardt algorithm for highly parameterized inverse model analyses. Water Resour. Res. 52, 6948–6977. https://doi.org/10.1002/2016WR019028

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A ConvNet for the 2020s. arXiv. https://doi.org/10.48550/arxiv.2201.03545

Loshchilov, I., Hutter, F., 2016. SGDR: Stochastic Gradient Descent with Warm Restarts. 5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc. https://doi.org/10.48550/arxiv.1608.03983

Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Z. Med. Phys. 29, 102–127.

https://doi.org/10.1016/J.ZEMEDI.2018.11.002

Marjańska, M., Deelchand, D.K., Kreis, R., Alger, J.R., Bolan, P.J., Borbath, T., Boumezbeur, F., Fernandes, C.C., Coello, E., Nagraja, B.H., Považan, M., Ratiney, H., Sima, D., Starčuková, J., Soher, B.J., Wilson, M., van Asten, J.J.A., 2022. Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM. Magn. Reson. Med. 87, 11–32. https://doi.org/10.1002/mrm.28942

Mikkelsen, M., Barker, P.B., Bhattacharyya, P.K., Brix, M.K., Buur, P.F., Cecil, K.M., Chan, K.L., Chen, D.Y.T., Craven, A.R., Cuypers, K., Dacko, M., Duncan, N.W., Dydak, U., Edmondson, D.A., Ende, G., Ersland, L., Gao, F., Greenhouse, I., Harris, A.D., He, N., Heba, S., Hoggard, N., Hsu, T.W., Jansen, J.F.A., Kangarlu, A., Lange, T., Lebel, R.M., Li, Y., Lin, C.Y.E., Liou, J.K., Lirng, J.F., Liu, F., Ma, R., Maes, C., Moreno-Ortega, M., Murray, S.O., Noah, S., Noeske, R., Noseworthy, M.D., Oeltzschner, G., Prisciandaro, J.J., Puts, N.A.J., Roberts, T.P.L., Sack, M., Sailasuta, N., Saleh, M.G., Schallmo, M.P., Simard, N., Swinnen, S.P., Tegenthoff, M., Truong, P., Wang, G., Wilkinson, I.D., Wittsack, H.J., Xu, H., Yan, F., Zhang, C., Zipunnikov, V., Zöllner, H.J., Edden, R.A.E., 2017. Big GABA: Edited MR spectroscopy at 24 research sites. Neuroimage 159, 32–45. https://doi.org/10.1016/J.NEUROIMAGE.2017.07.021

Mikkelsen, M., Rimbault, D.L., Barker, P.B., Bhattacharyya, P.K., Brix, M.K., Buur, P.F., Cecil, K.M., Chan, K.L., Chen, D.Y.T., Craven, A.R., Cuypers, K., Dacko, M., Duncan, N.W., Dydak, U., Edmondson, D.A., Ende, G., Ersland, L., Forbes, M.A., Gao, F., Greenhouse, I., Harris, A.D., He, N., Heba, S., Hoggard, N., Hsu, T.W., Jansen, J.F.A., Kangarlu, A., Lange, T., Lebel, R.M., Li, Y., Lin, C.Y.E., Liou, J.K., Lirng, J.F., Liu, F., Long, J.R., Ma, R., Maes, C., Moreno-Ortega, M., Murray, S.O., Noah, S., Noeske, R., Noseworthy, M.D., Oeltzschner, G., Porges, E.C., Prisciandaro, J.J., Puts, N.A.J., Roberts, T.P.L., Sack, M., Sailasuta, N., Saleh, M.G., Schallmo, M.P., Simard, N., Stoffers, D., Swinnen, S.P., Tegenthoff, M., Truong, P., Wang, G., Wilkinson, I.D., Wittsack, H.J., Woods, A.J., Xu, H., Yan, F., Zhang, C., Zipunnikov, V., Zöllner, H.J., Edden, R.A.E., 2019. Big GABA II: Water-referenced edited MR spectroscopy at 25 research sites. Neuroimage 191, 537–548. https://doi.org/10.1016/J.NEUROIMAGE.2019.02.059

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., Sutskever, I., 2019. Deep Double

Descent: Where Bigger Models and More Data Hurt. J. Stat. Mech. Theory Exp. 2021, 124003. https://doi.org/10.48550/arxiv.1912.02292

Near, J., Edden, R., Evans, C.J., Harris, A., Jezzard, P., 2015. Frequency and Phase Drift Correction of Magnetic Resonance Spectroscopy Data by Spectral Registration in the Time Domain. Magn. Reson. Med. 50, 44–50. https://doi.org/10.1002/mrm.25094

Near, J., Harris, A.D., Juchem, C., Kreis, R., Marjańska, M., Öz, G., Slotboom, J., Wilson, M., Gasparovic, C., 2020a. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. NMR Biomed. e4257. https://doi.org/10.1002/nbm.4257

Near, J., Harris, A.D., Juchem, C., Kreis, R., Marjańska, M., Öz, G., Slotboom, J., Wilson, M., Gasparovic, C., 2020b. Preprocessing, analysis and quantification in single□voxel magnetic resonance spectroscopy: experts' consensus recommendations. NMR Biomed. 1–23. https://doi.org/10.1002/nbm.4257

Oeltzschner, G., Zöllner, H.J., Hui, S.C.N., Mikkelsen, M., Muhammad, G., 2020. Osprey□: Open-Source Processing , Reconstruction & Estimation of Magnetic Resonance Spectroscopy Data.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.

Poullet, J.B., Sima, D.M., Van Huffel, S., 2008. MRS signal quantitation: A review of time- and frequency-domain methods. J. Magn. Reson. 195, 134–144. https://doi.org/10.1016/j.jmr.2008.09.005

Provencher, S.W., 2001. Automatic quantitation of localized in vivo 1H spectra with LCModel. NMR Biomed. 14, 260–264. https://doi.org/10.1002/NBM.698

Ratiney, H., Sdika, M., Coenradie, Y., Cavassila, S., van Ormondt, D., Graveron-Demilly, D., 2005. Time-domain semi-parametric estimation based on a metabolite basis set. NMR

Biomed. 18, 1–13. https://doi.org/10.1002/nbm.895

Robin A. de Graaf, 2019. In Vivo NMR Spectroscopy. Vivo NMR Spectrosc., Wiley Online Books. https://doi.org/10.1002/9781119382461.ch1

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117. https://doi.org/10.1016/J.NEUNET.2014.09.003

Simpson, R., Devenyi, G.A., Jezzard, P., Hennessy, T.J., Near, J., 2017. Advanced processing and simulation of MRS data using the FID appliance (FID-A)—An open source, MATLAB-based toolkit. Magn. Reson. Med. 77, 23–33. https://doi.org/10.1002/mrm.26091

Soher, B.J., Semanchuk, P., Todd, D., Steinberg, J., Young, K., 1994. VeSPA: Integrated applications for RF pulse design, spectral simulation and MRS data analysis. J. Magn. Reson. Ser. A 12, 75–105.

Stagg, C., Rothman, D.L., 2013. Magnetic resonance spectroscopy□: tools for neuroscience research and emerging clinical applications 500.

Starčuk, Z., Starčuková, J., 2017. Quantum-mechanical simulations for in vivo MR spectroscopy: Principles and possibilities demonstrated with the program NMRScopeB. Anal. Biochem. 529. https://doi.org/10.1016/j.ab.2016.10.007

Stefan, D., Cesare, F. Di, Andrasescu, A., Popa, E., Lazariev, A., Vescovo, E., Strbak, O., Williams, S., Starcuk, Z., Cabanas, M., Van Ormondt, D., Graveron-Demilly, D., 2009. Quantitation of magnetic resonance spectroscopy signals: The jMRUI software package. Meas. Sci. Technol. 20. https://doi.org/10.1088/0957-0233/20/10/104035

Tolstikhin, I., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., Dosovitskiy, A., 2021. MLP-Mixer: An all-MLP Architecture for Vision. arXiv. https://doi.org/10.48550/arXiv.2105.01601

Vanhamme, L., Van Den Boogaart, A., Van Huffel, S., 1997. Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge. J. Magn. Reson. 129, 35–43. https://doi.org/10.1006/jmre.1997.1244

Wilson, M., Reynolds, G., Kauppinen, R.A., Arvanitis, T.N., Peet, A.C., 2011. A constrained

least-squares approach to the automated quantitation of in vivo 1H magnetic resonance spectroscopy data. Magn. Reson. Med. 65, 1–12. https://doi.org/10.1002/mrm.22579

Zöllner, H.J., Považan, M., Hui, S.C.N., Tapper, S., Edden, R.A.E., Oeltzschner, G., 2021. Comparison of different linear-combination modeling algorithms for short-TE proton spectra. NMR Biomed. 34. https://doi.org/10.1002/nbm.4482

9 ACKNOWLEDGEMENTS

This work is part of the project that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 813120 (INSPiRE-MED) and was also supported by institutional support RVO:68081731 - Czech Academy of Sciences, Institute of Scientific Instruments. The authors thank Radim Kořínek, Ph.D. (Czech Academy of Sciences, Institute of Scientific Instruments, Czech Republic) for his valuable technical support.