

BRNO UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering  
and Communication

DOCTORAL THESIS

Brno, 2023

Amirmohammad Shamaei, M.Sc.



# BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

## FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY  
A KOMUNIKAČNÍCH TECHNOLOGIÍ

## DEPARTMENT OF BIOMEDICAL ENGINEERING

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

### DEEP LEARNING FOR SINGLE-VOXEL AND MULTIDIMENSIONAL MR-SPECTROSCOPIC SIGNAL QUANTIFICATION, AND ITS COMPARISON WITH NONLINEAR LEAST-SQUARES FITTING

HLUBOKÉ UČENÍ PRO KVANTIFIKACI JEDNOVOXELOVÝCH A MULTIDIMENZIONÁLNÍCH MR SPEKTROSKOPICKÝCH SIGNÁLŮ A JEHO SROVNÁNÍ S NELINEÁRNÍM FITOVÁNÍM METODOU NEJMEJŠÍCH ČTVERCŮ

### DOCTORAL THESIS

DIZERTAČNÍ PRÁCE

#### AUTHOR

AUTOR PRÁCE

Amir mohammad Shamaei, M.Sc.

#### SUPERVISOR

ŠKOLITEL

Ing. Radovan Jiřík, Ph.D.

#### CO-SUPERVISOR

SPOLUŠKOLITEL

Ing. Jana Starčuková, Dr.

BRNO 2023



## Abstract

Preprocessing, analysis, and quantification of Magnetic resonance spectroscopy (MRS) signals are required for obtaining the metabolite concentrations of the tissue under investigation. However, a fast, accurate, and efficient post-acquisition workflow (preprocessing, analysis, and quantification) of MRS is challenging.

This thesis introduces novel deep learning (DL)-based approaches for preprocessing, analysis, and quantification of MRS data. The proposed methods achieved the objectives of robust data preprocessing, fast and efficient MR spectra quantification, *in-vivo* concentration quantification, and the uncertainty estimation of quantification. The results showed that the proposed approaches significantly improved the speed of MRS signal preprocessing and quantification in a self-supervised manner. Our proposed methods showed comparable results with the traditional methods in terms of accuracy. Furthermore, a standard data format was introduced to facilitate data sharing among research groups for artificial intelligence applications. The findings of this study suggest that the proposed DL-based approaches have the potential to improve the accuracy and efficiency of MRS for medical diagnosis.

The dissertation is structured into four parts: an introduction, a review of state-of-the-art research, a summary of the aims and objectives, and a collection of publications that showcase the author's contribution to the field of DL applications in MRS.

## Keywords

MR spectroscopy, inverse problem, deep learning, machine learning, convolutional neural network, metabolite quantification, frequency and phase correction.

## Abstrakt

Pro získání koncentrace metabolitů ve vyšetřované tkáni ze signálů magnetické rezonanční spektroskopie (MRS) je nezbytné provést předzpracování, analýzu a kvantifikaci MRS signálu. Rychlý, přesný a účinný proces zpracování (předzpracování, analýza a kvantifikace) MRS dat je však náročný.

Tato práce představuje nové přístupy pro předzpracování, analýzu a kvantifikaci MRS dat založené na hlubokém učení (DL). Navržené metody potvrdily schopnost použití DL pro robustní předzpracování dat, rychlou a efektivní kvantifikaci MR spekter, odhad koncentrací metabolitů *in vivo* a odhad nejistoty kvantifikace. Navržené přístupy výrazně zlepšily rychlost předzpracování a kvantifikace MRS signálu a prokázaly možnost použití DL bez učitele. Z hlediska přesnosti byly získány výsledky srovnatelné s tradičními

metodami. Dále byl zaveden standardní formát dat, který usnadňuje sdílení dat mezi výzkumnými skupinami pro aplikace umělé inteligence. Výsledky této studie naznačují, že navrhované přístupy založené na DL mají potenciál zlepšit přesnost a efektivitu zpracování MRS dat pro lékařskou diagnostiku.

Disertační práce je rozdělena do čtyř částí: úvodu, přehledu současného stavu výzkumu, shrnutí cílů a úkolů a souboru publikací, které představují autorův přínos v oblasti aplikací DL v MRS.

## **Klíčová slova**

MR spektroskopie, inverzní problém, hluboké učení, strojové učení, konvoluční neuronová síť, kvantifikace metabolitů, frekvenční a fázová korekce.

## **Bibliographic citation**

### **Printed work citation:**

SHAMAEI, Amir mohammad. Deep learning for single-voxel and multidimensional MR-spectroscopic signal quantification, and its comparison with nonlinear least-squares fitting. Brno, 2023. Available also at: <https://www.vutbr.cz/studenti/zav-prace/detail/135526>. Doctoral Thesis Topic. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Department of Biomedical Engineering. Supervisor Radovan Jiřík, cosupervisor Jana Starčuková.

### **Electronic source citation:**

SHAMAEI, Amir mohammad. Deep learning for single-voxel and multidimensional MR-spectroscopic signal quantification, and its comparison with nonlinear least-squares fitting [online]. Brno, 2023 [cit. 2023-02-16]. Available from: <https://www.vutbr.cz/studenti/zav-prace/detail/135526>. Doctoral Thesis Topic. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Department of Biomedical Engineering. Supervisor Radovan Jiřík, cosupervisor Jana Starčuková.

# DECLARATION

## Author's Declaration

**Author:** *Amirmohammad Shamaei*

**Author's ID:** *223363*

**Paper type:** *Ph.D. Thesis*

**Academic year:** *2022/23*

**Topic:** *DEEP LEARNING FOR SINGLE-VOXEL  
AND MULTIDIMENSIONAL MR-SPECTROSCOPIC SIGNAL QUANTIFICATION,  
AND ITS COMPARISON WITH NONLINEAR LEAST-SQUARES FITTING*

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the project and listed in the comprehensive bibliography at the end of the project.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation S 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll., Section 2, Head VI, Part 4.

Brno, April 26, 2023

-----  
author's signature

# Acknowledgement

Throughout the whole doctoral studies, I have received significant support and assistance. Now, I would like to express my gratitude.


I would like to thank my (co-)supervisors, Ing. Jana Starčuková, Ph.D., Ing. Zenon Starčuk, CSc. and Ing. Radovan Jiřík, Ph.D. for their professional mentoring, inspiration, and motivation. Their advice, feedback, and questions pushed me to sharpen my thinking and brought my work to a higher level.

I would like to thank the Institute of Scientific Instruments of the Czech Academy of Sciences as well as my friends, colleagues, and research team – Ing. Zdeněk Dokoupil, MVDr. Eva Dražanová, Jana Homolová, Ing. Jaroslav Horký, Ing. Radovan Jiřík, Ph.D., Ing. Radim Kořínek (FG), Ph.D., Ing. Lucie Krátká, Ing. Jiří Kratochvíla, Ing. Pavlína Libová, Ing. Ondřej Macíček (CG), Mgr. Aneta Malá, Mgr. Iveta Pavlova (GL), Lucie Pelikánová, Mgr. Tomáš Pšorn (CC) and Jiří Vitouš for their energy, understanding, and help throughout my research and working on my dissertation thesis.

I also thank prof. Ing. Valentine Provazník, Ph.D., the head of the Department of Biomedical Engineering, BUT for providing individual conditions for my studies and research.

Furthermore, I wish to thank my dear friends (Reza, Mammad, Hasti, Omid, Hossein) and colleagues from the INSPIRE-MED consortium for being such an amazing team.

Last but not least, my gratitude belongs also to my family, especially my mother, father, and my lovely sisters, for their encouragement and support throughout my studies.

 This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 813120.

Brno, April 26, 2023

-----  
Author's signature

# CONTENTS

<b>1. INTRODUCTION.....</b>	<b>10</b>
1.1 BASIC PRINCIPLES OF NMR SPECTROSCOPY.....	11
1.2 METABOLITE QUANTIFICATION OF MR SPECTRA.....	14
1.2.1 <i>Data Preprocessing</i> .....	14
1.2.2 <i>Spectral Analysis</i> .....	15
1.2.3 <i>Results of Spectral Analysis</i> .....	18
1.3 INTRODUCTION TO DEEP LEARNING.....	18
1.3.1 <i>Basic Principles of Deep Learning</i> .....	18
<b>2. STATE OF THE ART OF DL IN MRS .....</b>	<b>24</b>
<b>3. AIMS OF THE DOCTORAL THESIS .....</b>	<b>26</b>
<b>4. SELECTED PUBLICATIONS .....</b>	<b>27</b>
4.1 PAPER 1 – MODEL-INFORMED UNSUPERVISED DEEP LEARNING APPROACHES TO FREQUENCY AND PHASE CORRECTION OF MRS SIGNALS .....	30
4.2 PAPER 2 – DEEP LEARNING FOR MAGNETIC RESONANCE SPECTROSCOPY QUANTIFICATION: A TIME-FREQUENCY ANALYSIS APPROACH.....	32
4.3 PAPER 3 – A WAVELET SCATTERING CONVOLUTION NETWORK FOR MAGNETIC RESONANCE SPECTROSCOPY SIGNAL QUANTITATION.....	34
4.4 PAPER 4 – QUANTIFICATION OF MR SPECTRA BY DEEP LEARNING IN AN IDEALIZED SETTING: INVESTIGATION OF FORMS OF INPUT, NETWORK ARCHITECTURES, OPTIMIZATION BY ENSEMBLES OF NETWORKS, AND TRAINING BIAS.....	36
4.5 PAPER 5 – PHYSICS-INFORMED DEEP LEARNING APPROACH TO QUANTIFICATION OF HUMAN BRAIN METABOLITES FROM MAGNETIC RESONANCE SPECTROSCOPY DATA. ....	38
4.6 PAPER 6 – NIFTI-MRS: A STANDARD DATA FORMAT FOR MAGNETIC RESONANCE SPECTROSCOPY 40	
<b>5. CONCLUSION.....</b>	<b>42</b>

# FIGURES

Figure 1 Rotations of the macroscopic magnetization vector in the nonrotating laboratory frame $xyz$ . The primary magnetic field $B_0$ along the $z$ axis and the oscillating magnetic field $B_1$ (RF pulse) in the $xy$ plane, both exert their respective influences on the magnetization, causing it to precess. ....	12
Figure 2 (Left) Time domain signal representation (i.e., FID) and its projection in two planes corresponding to the real (orange) and imaginary (green) components. (Right) The graphic shows the real (orange) and imaginary (green) components of the Fourier Transformed signal .....	13
Figure 3 Short-TE $^1\text{H}$ MRS of brain acquired by PRESS sequence from rat brain <i>in vivo</i> at 9.4 T. ....	14
Figure 4 A linear combination of basis spectra is used to quantify <i>in vivo</i> $^1\text{H}$ MR spectra. The data can be reconstructed by a linear combination of basis spectra. The contribution of each basis spectrum to the total spectrum is plotted.....	17
Figure 5 An MLP with multiple hidden layers. ....	20
Figure 6 LeNet architecture. ....	21
Figure 7 A typical design of an autoencoder. ....	22
Figure 8 A single-layer neural network with a non-linear activation function highlighting the feedforward and backpropagation steps.....	23

# TABLES

Table 1 A summary of related work on MRS signal quantification using DL.....	25
Table 2 A summary of our work on MRS signal quantification using DL.....	28



# 1. INTRODUCTION

Magnetic resonance spectroscopy (MRS) has attracted the magnetic resonance (MR) community over the past seven decades [1]–[3]. A significant part of the interest in biomedical MRS stems from the possibility of non-invasive measurements of metabolites [3]. Information about tissue metabolites can help in clinical diagnostics. For instance, the detection of metabolic pathway changes may facilitate diagnosing disease in earlier stages before anatomy changes can be observed and thus enable more efficient treatment. This is demonstrated in glioma, a decrease of N-acetylaspartate (NAA) and creatine (Cr) concentrations and an increase of choline, lipids, and lactate predicts an increase in the glioma grade [1], [2].

Contrary to other diagnostic techniques such as Computed Tomography (CT) and Radionuclide imaging (e.g., Positron Emission Tomography (PET), Gamma camera), MRS emits no ionizing radiation to the subject, which enables follow-up studies [2].

It is possible to get various information using Spectroscopy (MRS) and Spectroscopic Imaging (MRSI) for atomic nuclei such as Proton ( $^1\text{H}$ ), Phosphorus ( $^{31}\text{P}$ ), Carbon ( $^{13}\text{C}$ ), etc. The proton nucleus has the highest natural abundance (>99.9%) and intrinsic nuclear magnetic resonance (NMR) sensitivity (high gyromagnetic ratio) among all nuclei for MRS. Presently, the primary utilization of MRS in the clinical setting is centered around  $^1\text{H}$ , with other nuclei being predominantly applied in preclinical and fundamental research. Although a number of brain metabolites can be identified with  $^1\text{H}$  MRS, the quantity of substances assessable under *in vivo* conditions does not exceed 15-20, and is typically much lower. Further MR-visible nuclei with biochemical relevance have also demonstrated their value in providing information on tissue physiology and biochemistry [1]–[4]. Notably, the incorporation of non- $^1\text{H}$  MRS in clinical practice has yet to become widespread, potentially due to the prerequisite of nonstandard hardware and specialized MRS techniques [3].

Recently, there has been a resurgence of interest in MRS within the MR community for clinical neuroscience and translational purposes [3]. This is largely due to the increased availability of high and ultrahigh-field scanners, and a better understanding of the role of metabolism in neuroenergetics and neurotransmission. Consequently, there has been a rise in the use of MRS for both typical clinical applications and neurological research, supported by advancements in acquisition and processing techniques [3].

However, MRS has some shortcomings that limit its routine use in clinical practice. Some of these are (i) MRS requires specialized equipment and software that are not widely available or standardized, (ii) MRS is time-consuming and prone to artifacts and noise that affect the quality and reliability of the spectra, (iii) MRS data analysis and interpretation are complex and require expert knowledge and skills [1], [5].

Deep learning (DL) has the potential to enhance the accuracy and efficiency of MRS by automating the data analysis, providing more reliable diagnoses, and enabling the use of MRS in routine clinical practice.

A brief introduction to the basic principles of NMR will be given in the following Introduction chapter, followed by a review of the existing methods and challenges for metabolite quantification of MR spectra. Finally, an overview of DL, including principles, the main architectures, and learning processes, will be presented.

## 1.1 Basic Principles of NMR Spectroscopy

MRS relies on the NMR phenomenon to observe the nuclear magnetism of the atoms in a sample and how they respond to a magnetic field. Several atomic nuclei, e.g. proton ( $^1\text{H}$ ), carbon ( $^{13}\text{C}$ ), fluor ( $^{19}\text{F}$ ), and phosphorus ( $^{31}\text{P}$ ), have a nuclear spin quantum number of  $\frac{1}{2}$ , which is manifested by magnetic moment and angular momentum. Their behavior can be likened to a miniature cylindrical bar magnet that rotates around its axis and exhibits similar magnetic and inertial characteristics. Such characteristics dictate how these nuclei behave in the presence of an external magnetic field, where they tend to align along this external magnetic field. The spins precess with a specific frequency (i.e. the Larmor frequency  $f_0$ ) around the Z-axis (the axis which is parallel to the direction of the magnetic field  $\mathbf{B}_0$ ) which is proportional to the external magnetic field and defined by the Larmor equation

$$f_0 = \gamma \mathbf{B}_0 , \quad (1.1)$$

where  $\gamma$  represents the gyromagnetic ratio (in Hz/T) of the nuclei of interest, and  $\mathbf{B}_0$  is the static external magnetic field (in T) [2], [6].

A macroscopic magnetic moment is created due to an abundance of nuclei with lower-energy orientation. By utilizing another magnetic field, the net longitudinal magnetization in equilibrium ( $M_0$ ) is flipped into the transverse plane (xy-plane) in order to detect the nuclear magnetization of a sample. This is achieved by applying a short powerful RF pulse, so-called  $\mathbf{B}_1$ , along the axis perpendicular to the static magnetic field ( $\mathbf{B}_0$ ). This pulse with a duration of  $t_p$  tips the magnetizations precessing with Larmor frequencies approximately in the range of the RF pulse spectrum through an angle given by

$$\theta = \gamma \mathbf{B}_1 t_p . \quad (1.2)$$

After excitation, the flipped magnetization precesses in the xy-plane, generating an oscillating signal which can be detected by the receiver coil. This signal is called free induction decay (FID). FID has a complicated wave pattern decaying away to zero (see Figure 1). Relaxation refers to a process where molecules move randomly, resulting in energy transfer and random interactions.  $T_1$ , also known as the spin-lattice relaxation time, is the recovery time of the longitudinal magnetization in the same direction as the static magnetic field. The relation between the transverse magnetization and  $T_1$  is given

by

$$M_z(t) = M_0 \left( 1 - e^{-\frac{t}{T_1}} \right), \quad (1.5)$$

where  $M_z$  is the longitudinal net magnetization, and  $M_0$  is the steady-state net magnetization [2].

$T_2$ , the so-called spin-spin relaxation time, reflects the time it takes for the transverse magnetization to decay in the transverse plane. The relation between the transverse magnetization and  $T_2$  is given by

$$M_{xy}(t) = M_{xy0} \left( e^{-\frac{t}{T_2}} \right), \quad (1.6)$$

where  $M_{xy0}$  is the initial net transverse magnetization.

$T_2^*$  (also known as the effective transverse relaxation time) is related to  $T_2$  by  $1/T_2^* = 1/T_2 + 1/T_2'$ , where  $T_2'$  refers to static  $\mathbf{B}_0$  inhomogeneities causing dephasing of the water protons due to small variations in their local magnetic fields.  $T_2^*$  is always either less than or equal to  $T_2$ .

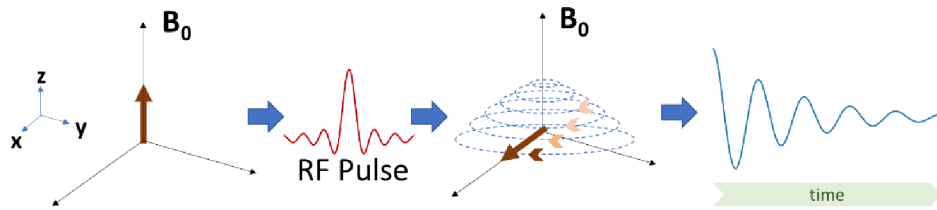


Figure 1 Rotations of the macroscopic magnetization vector in the nonrotating laboratory frame  $xyz$ . The primary magnetic field  $B_0$  along the  $z$  axis and the oscillating magnetic field  $B_1$  (RF pulse) in the  $xy$  plane, both exert their respective influences on the magnetization, causing it to precess.

In an ideal situation, all nuclei of the same isotope ( $^1\text{H}$  in our case) would precess at an identical frequency in a uniform field  $\mathbf{B}_0$ . However, a certain frequency range is detected in the sample in reality. Electrons around the nucleus in a molecule shield the nucleus from the external field and produce nuclear shielding. The amount of shielding experienced by a nucleus determines the net magnetic field given by

$$\mathbf{B} = \mathbf{B}_0(1 - \sigma'), \quad (1.7)$$

where  $\sigma'$  is the shielding constant that depends on the electrical environment of a nucleus [2], [6]. The net magnetic field varies at each nucleus. This phenomenon leads to magnetic-moment precession with a small frequency shift from the reference Larmor frequency  $f_0$ , and therefore the FID contains more components with different frequencies. The transformation of the FID from the time domain into the frequency domain using the Fourier Transform (FT) enables us to observe molecules of metabolites in the form of

peaks corresponding to the excited nuclei (see Figure 2).

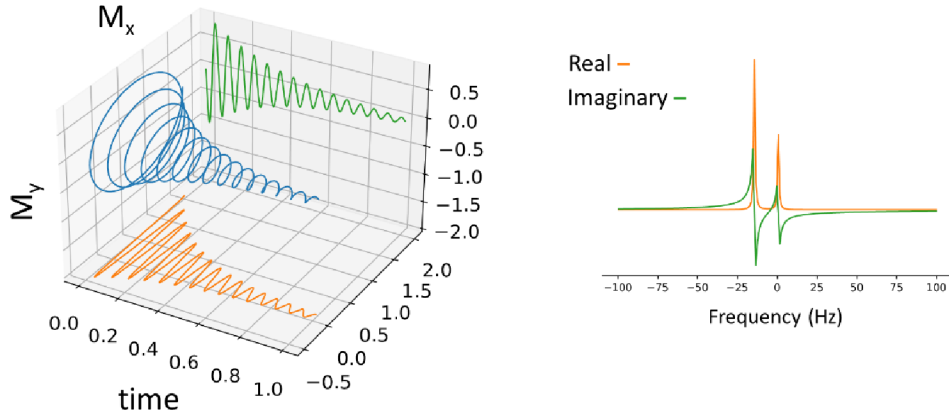


Figure 2 (Left) Time domain signal representation (i.e., FID) and its projection in two planes corresponding to the real (orange) and imaginary (green) components. (Right) The graphic shows the real (orange) and imaginary (green) components of the Fourier-transformed signal

When adjacent nuclei interact with each other's magnetic fields, a phenomenon known as spin-spin coupling or J-coupling occurs. J-coupling arises when multiple spins are within a molecule, producing alterations in the local magnetic field surrounding each nucleus. This is reflected in an NMR spectrum as the splitting of peaks into multiplet structures [2], [6].

In order to increase the SNR of an FID, multiple FIDs are acquired and averaged. The resulting FID signal can be described with a complex function [the real part is the projection of the magnetization vector on the x-axis, while the imaginary part is the projection on the y-axis] as

$$fid(t) = M_x(t) + i M_y(t), \quad (1.3)$$

where  $R(t)$  and  $I(t)$  are the real and imaginary components, respectively. The Fourier transformation calculates a frequency-domain spectrum  $F(\nu)$  from a time-domain signal  $fid(t)$  according to

$$F(\nu) = \int_{-\infty}^{+\infty} fid(t) \cdot e^{-2\pi i \nu t} dt. \quad (1.4)$$

Figure 3 illustrates an example of an *in-vivo* frequency-domain spectrum. Modern scanners are using an analog-to-digital converter (ADC), which measures the instantaneous value of the FID at equal time intervals. The time interval, also known as dwell time, equals  $1/\text{spectral bandwidth (SW)}$  [7].

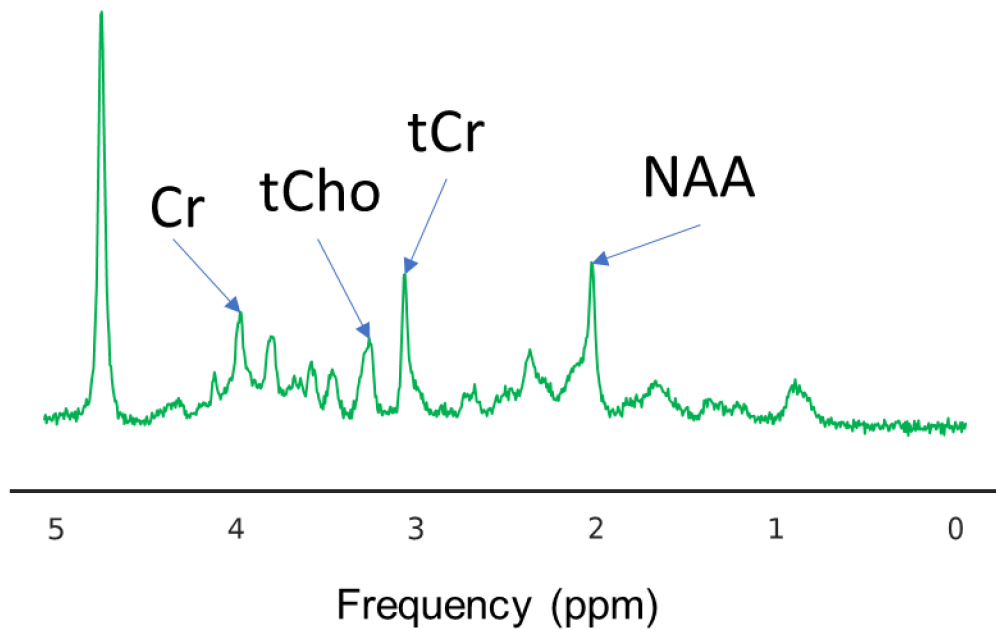


Figure 3 Short-TE  $^1\text{H}$  MRS of brain acquired by PRESS sequence from rat brain *in vivo* at 9.4 T.

## 1.2 Metabolite quantification of MR spectra

### 1.2.1 Data Preprocessing

MRS data must be preprocessed once the time domain signals have been obtained in order to make the signal correspond as closely as possible to the model used in quantification and consequently to get accurate concentration estimations. Data preprocessing in MRS involves several steps, including eddy current correction, motion correction, frequency and phase correction (FPC), nuisance peak removal, RF coil combination, and signal averaging. Some of the preprocessing steps play a crucial role in ensuring the quality of the final data [2], [8]. In the following some of the preprocessing steps will be explained; moreover, readers can find experts' consensus recommendations on preprocessing operations in [8].

In an MRI scanner, rapid switching of gradients can result in short-lived fluctuations in the main magnetic field, known as eddy current effects. These effects can distort the shape of spectral lines and affect the accuracy of further spectral analysis. To correct this, a water-unsuppressed spectrum is collected using the same gradient strengths and timings as the water-suppressed dataset. The phase of the water-unsuppressed signal is then subtracted from both the water-suppressed and water-unsuppressed FID signals to correct for eddy currents [8], [9].

In addition, the main magnetic field of an MRI scanner can experience subtle temporal drift due to the heating and cooling of the ferromagnetic passive shim elements. This can result in frequency drifts during MRS experiments. In addition, physiological or small bulk motion can lead to additional frequency and phase shifts during an MRS scan. Several FPC approaches have been developed [10]–[16]. If not corrected, these frequency and phase shifts can lead to the broadening of spectral peaks, reduction in signal-to-noise ratio (SNR), and line shape distortion [8]. A commonly used approach for FPC is to use the water peak and estimate the phase and frequency from it [12], [14], [17]. Another approach is to fit a certain metabolite peak to a model [18] and then estimate the frequency and phase shifts from the model. One approach that has been proposed and evolved recently is spectral registration (SR) [10], [11], [15], [16]. It has been shown that DL can also be employed for FPC [13], [19], [20] and could speed up FPC once it has been successfully trained.

Nuisance peak removal is performed to minimize signal contamination, such as water, lipid, or macromolecule signals. It is important to ensure that the baseline is accurately corrected, as it can interfere with the detection of the metabolite signals. The nuisance peaks can be removed by fitting a set of line shape functions to peaks and subtracting the resulting fit from the spectrum. Alternatively, the nuisance peak can be left in place and analysis can be performed using a fitting model that incorporates the nuisance peaks [8], [21].

In conclusion, MRS data preprocessing is an important step in the analysis of MRS data. The preprocessing steps are performed to enhance the quality of data. The success of MRS studies largely depends on the quality of acquired data and the preprocessing steps, which should be performed carefully and accurately to obtain reliable results.

### **1.2.2 Spectral Analysis**

The aim of spectral analysis is to estimate concentrations (often only relative) of metabolites [8]. Spectral analysis can be performed in several ways, depending on the type of information that is needed. One way is to use the area under the curve of each metabolite peak in the spectrum, which is proportional to the concentration of that metabolite. This method is simple and straightforward, but it is only recommended for MRS signals that have non-overlapping metabolites, well-phased peaks, and no baseline or MM contribution [8], [22].

More advanced techniques have been developed for MRS quantification, such as spectral fitting, which involves the use of mathematical models to simulate the complex spectra of multiple metabolites and their interactions [23], [24].

The most commonly used methods of spectral fitting are (1) linear combination model fitting (LCM), and (2) peak fitting. In LCM fitting, each metabolite in the spectrum is represented by a "basis spectrum," which is a description of the spectral shape of individual metabolite components (see Figure 4) [2], [8], [22], [23] and 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

$$S(t) = \left( \sum_{m=1}^M A_m X_m(t) + A_b B(t) \right) e^{(\Delta\alpha + 2i\pi\Delta f)t} e^{i\Delta\theta} + \xi(t), \quad (1.11)$$

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.  $\Delta\alpha$ ,  $\Delta f$ , 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.

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  $^1\text{H}$ -MRS of the brain due to the excessive number of metabolites and peaks per metabolite [2], [8], [22], [23].

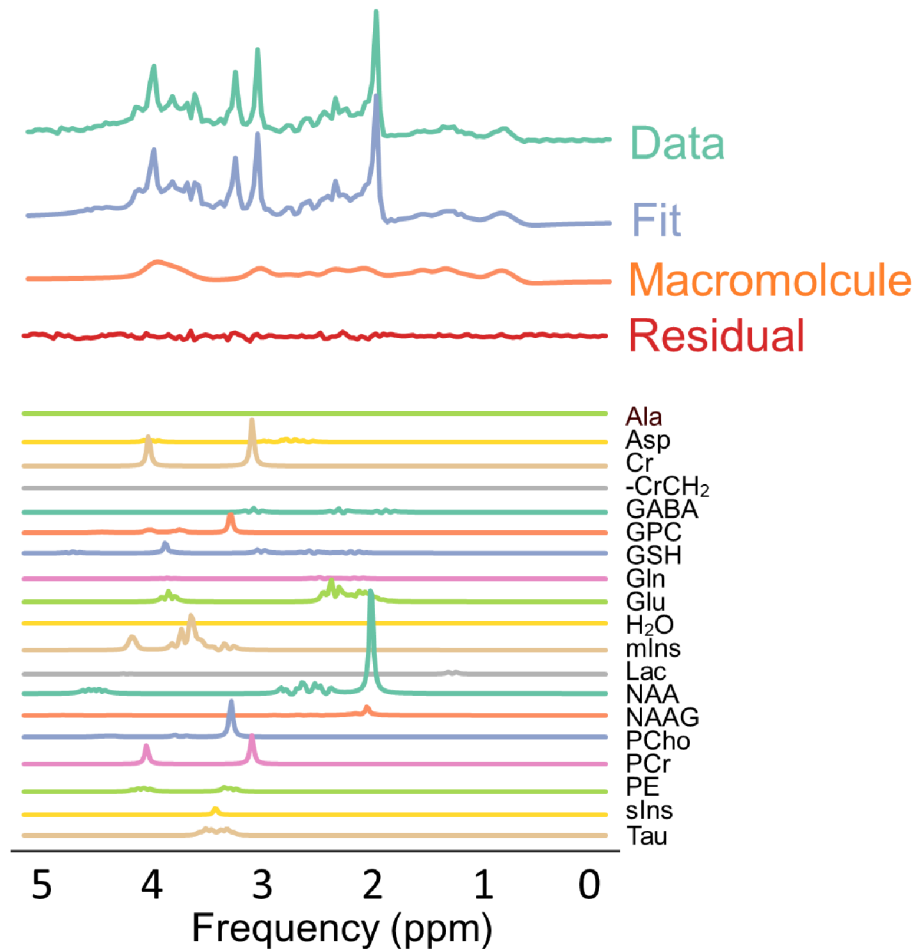


Figure 4 A linear combination of basis spectra is used to quantify *in vivo*  $^1\text{H}$  MR spectra. The data can be reconstructed by a linear combination of basis spectra. The contribution of each basis spectrum to the total spectrum is plotted.

In addition, LCM fitting is the preferred method according to experts [24] owing to its shown efficacy, flexibility, and relative simplicity of usage. Using basis spectra instead of individual peak components (peak fitting method) reduces the number of model functions needed to accurately represent the spectrum, resulting in fewer fitting parameters. Basis spectra are realistic as they are obtained from phantom experiments or numerical simulations [2], [8], [22], [23], [25], [26].

A long-standing interest in LCM fitting has resulted in the development of numerous fitting methods, including time-domain and frequency-domain algorithms [23], such as LCMModel [27], QUEST [28], TARQUIN [29], FiTAID [30], FSL-MRS [31], and Osprey [32]. Even though several software packages and techniques have been introduced over the past two decades for spectral analysis [29], [33]–[35], reliable quantification of MRS signals is still an issue because of several challenges, such as (i) strongly overlapping metabolite peaks, (ii) low signal-to-noise ratio, (iii) peak shape variations at high magnetic field strengths, and (iiii) broad and only partially known background from macromolecules and lipids, and further improvements are necessary [33], [36].



Despite advancements in MRS spectral analysis, it has been demonstrated that the various methods could show different quantification results [37]. Recent attempts to compare and verify these tools using gold standard phantoms have shown weak accuracy and significant differences between the concentrations recorded by different methods [37], [38]. In addition, traditional methods are time-consuming and a computational bottleneck in the quantification of large MRS datasets.

### 1.2.3 Results of Spectral Analysis

Results obtained with fitting algorithms are amplitudes or areas representing the relative proportion of the metabolite signal in the MRS signals. Therefore results of spectral analysis are often expressed as ratios of estimated amplitudes (areas) of metabolites to the estimated amplitude (area) of a selected reference metabolite or of water [21].

In practice, a reference compound of known concentration [*Ref*] has to be utilized to detect absolute concentration. The metabolite concentration can be calculated according to

$$\text{Concentration} = [\textit{Ref}] \cdot \frac{S_M}{S_R} \cdot \text{CF}, \quad (1.12)$$

where  $S_M$ ,  $S_R$ , and CF are the estimated amplitude of the interested compound, the estimated amplitude of the reference compound, and a correction factor, respectively [2]. However, it is hard and time-consuming to calculate a reliable CF; Thus, it is common to use the metabolites ratio rather than the absolute concentration [2].

## 1.3 Introduction to Deep learning

The field of DL has come a long way since its inception. As the world progresses toward automation, understanding how computers learn from data has become essential. In this chapter, the basic principles of DL will be explored, followed by DL architectures and some of the most popular and effective learning processes and algorithms.

### 1.3.1 Basic Principles of Deep Learning

DL is a subset of machine learning that is revolutionizing the field of artificial intelligence [39]. It is a powerful approach to building intelligent systems that can learn and make decisions on their own. DL algorithms are based on neural networks, which are modeled after the structure and function of the human brain [39], [40].

DL has proven to be highly effective at solving complex problems that were previously thought to be insurmountable. It has been used to achieve state-of-the-art results in a wide range of applications, including image and speech recognition, natural language processing, autonomous vehicles, and medical diagnosis [39], [41]–[43].

One of the key advantages of DL is its ability to automatically learn features from raw data. This means that DL algorithms can learn to recognize patterns and relationships in

data without the need for explicit feature engineering. This makes DL particularly well-suited to problems where the underlying structure of the data is complex and difficult to capture with traditional approaches [39].

However, DL is not a silver bullet and requires large amounts of data and computing power to be effective. It also requires careful tuning of hyperparameters and model architectures to achieve optimal performance. Despite these challenges, DL has already made significant contributions to the field of artificial intelligence and is expected to continue to drive innovation and progress in the coming years [44], [45].

### **Multilayer perceptron**

MLPs are artificial neural networks that consist of input and output layers, and one or more hidden layers of interconnected processing units (see Figure 5). The basic idea of an MLP is to use a series of nonlinear transformations to map an input vector to an output vector. Each layer of an MLP has multiple nodes, which receive an input vector and perform a set of mathematical operations to produce an output vector [39].

The input layer of an MLP takes the raw input data and passes it on to the next layer. The output of the input layer is fed into a series of hidden layers, which perform nonlinear transformations on the input data. The output of each hidden layer is then passed on to the next layer, until the final output layer is reached. The output layer produces the final output of the network. An MLP can be trained utilizing gradient descent optimization and backpropagation algorithms.

MLPs are commonly used in a wide range of applications, including image recognition, speech recognition, natural language processing, and predictive analytics [39], [46]. They are popular due to their ability to learn complex non-linear relationships between input and output data, and their ability to generalize to new data [39], [47].

One of the challenges of using MLPs is selecting the appropriate architecture, such as the number of layers and nodes in each layer, as well as the activation functions (nonlinear transformations) used in each layer. The choice of these parameters can greatly affect the performance of the network, and a trial-and-error approach is often used to find the optimal configuration [39].

# Multilayer Perceptrons (MLPs)

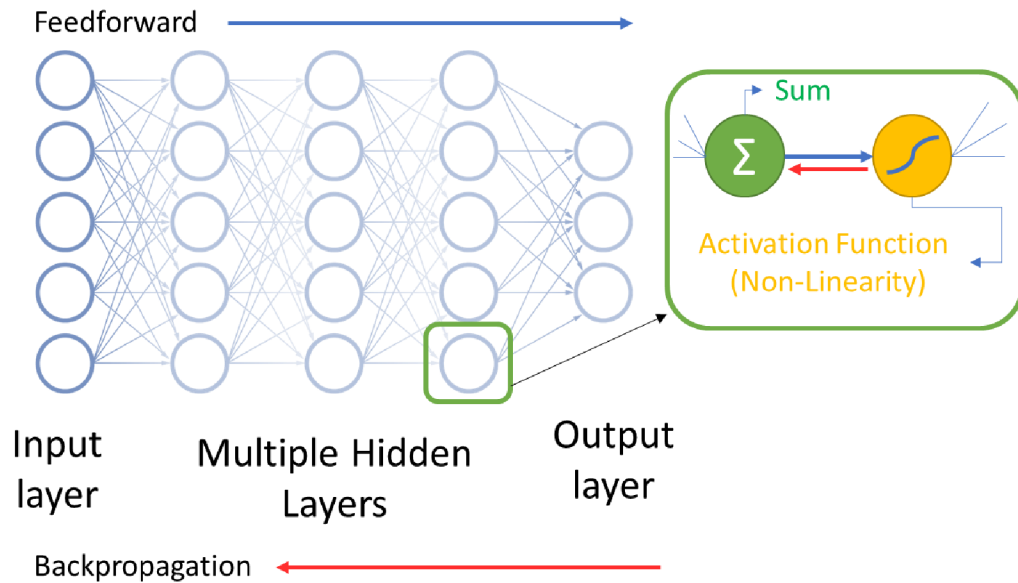


Figure 5 An MLP with multiple hidden layers.

## Convolutional neural networks

CNNs are a special type of neural network that have been widely used in recent years for various applications, including medical signal analysis. CNNs are designed to automatically learn and extract relevant features from input data, by convolving convolution kernels with the input signal or image and learning weights that optimize the kernels for specific tasks [39], [48], [49].

CNNs have shown great success in several medical signal analysis tasks, including electroencephalogram (EEG) analysis [50], electromyography (EMG) analysis [51], and medical image analysis [45], [52]. In medical image analysis, CNNs can segment and classify images of different organs, tumors, and pathologies [45], [52].

The formulation of a CNN consists of a series of layers, including convolutional layers (1D and 2D convolutional kernels for 1D data (signals) and 2D data (images), respectively), pooling layers, and fully connected layers. In the convolutional layer, the input signal is convolved with a set of kernels, which are learned by the network during training. The output of the convolutional layer is a set of feature maps that represent the learned features. In the pooling layer, the feature maps are down-sampled to reduce the size and complexity of the data. Finally, the fully connected layer takes the flattened feature maps and uses them to make a prediction about the input data [39].

LeNet is a CNN architecture (see Figure 6) developed by Yann LeCun in the late 1990s for handwritten digit recognition [53]. It was one of the first successful applications of DL in computer vision and paved the way for many modern DL models.

The LeNet architecture consists of seven layers, namely two convolutional layers, two

subsampling (pooling) layers, and three fully connected layers. The input to the network is a 32×32 grayscale image, and the output is a probability distribution over the 10 possible digit classes.

The convolutional layers learn a set of filters that detect local features in the input image, such as edges and corners. The subsampling layers reduce the size of the feature maps by taking the maximum value over a small window. The fully connected layers combine the local features into global features and output the final classification.

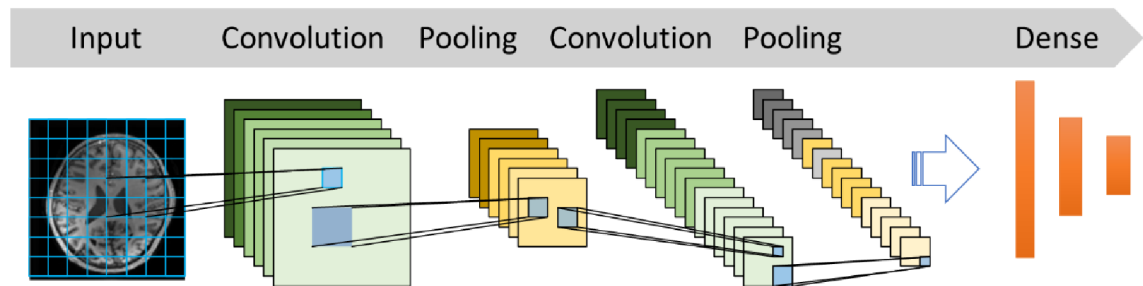


Figure 6 LeNet architecture.

### Autoencoders

Autoencoders are neural networks that learn to encode data into a low-dimensional representation and then decode it back to its original form [39], [54]. They are widely used in unsupervised and self-supervised learning tasks as they can learn to represent data without explicit labels. The idea behind autoencoders is that by encoding the data into a low-dimensional space, they can capture the most important features of the data. Figure 7 shows a typical design of an autoencoder.

The autoencoder is trained on the input data, and its objective is to minimize the difference between the input and the output. This forces the autoencoder to learn a compressed representation of the data that can be used for other tasks. Once the autoencoder has learned a good representation of the data, it can be used to train other models for a variety of tasks, such as image classification, object detection, and natural language processing [39], [54].

In self-supervised learning, the task is to learn a useful representation of the data without using any explicit labels. This can be done by using an autoencoder to learn a representation of the data that captures its most important features.

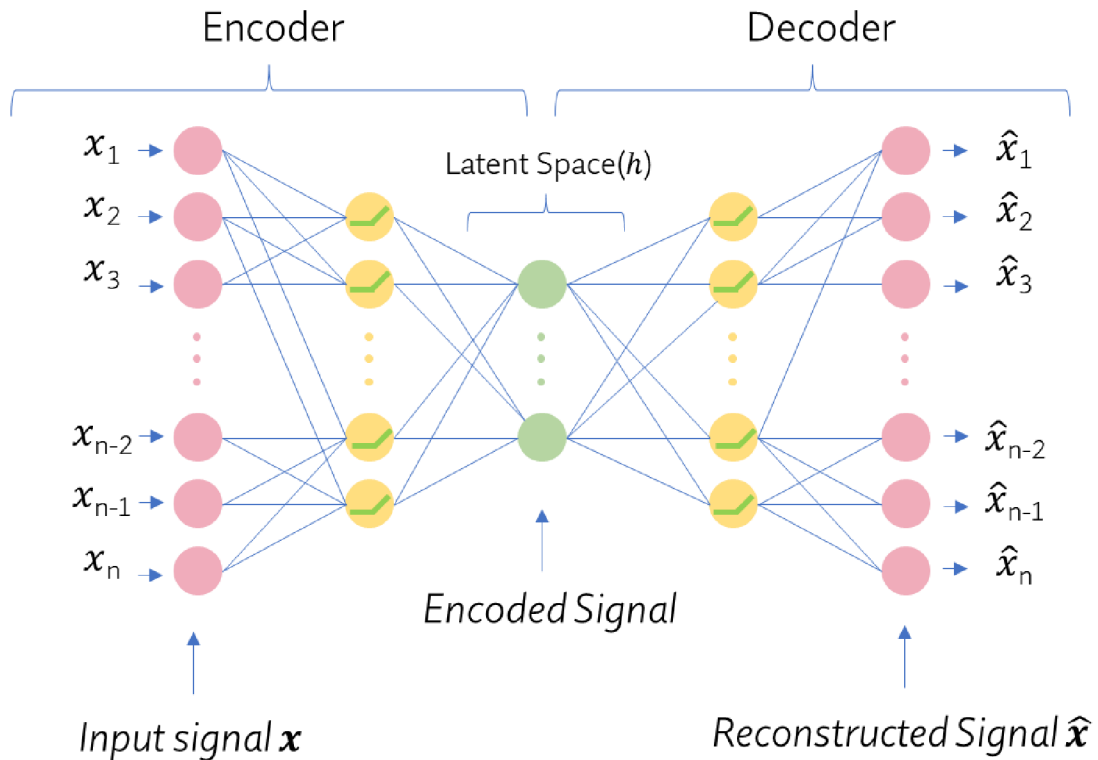


Figure 7 A typical design of an autoencoder.

### Learning processes

There are various types of machine learning problem formulations, but the most commonly used ones are supervised learning, unsupervised learning, and self-supervised learning.

Supervised learning is a type of machine learning where the model is trained on labeled data, which means the training data has already been labeled with the correct answers. The goal of supervised learning is to predict the labels of new, unseen data. For example, a supervised learning algorithm could be used to predict the price of a house based on its features, such as location, number of bedrooms, and square footage [39], [43].

Unsupervised learning is a type of machine learning where the model is trained on unlabeled data. The goal of unsupervised learning is to find patterns or structures in the data. For example, an unsupervised learning algorithm could be used to group similar customers together based on their purchasing habits [39], [55].

Self-supervised learning is a type of machine learning where the model is trained on data that has been labeled by itself. This is often achieved by using a technique called auto-encoding, where the model learns to encode and decode the input data. The goal of self-

supervised learning is to learn a representation of the data that can be used for other tasks. For example, self-supervised learning could be used to learn a representation of images that can be used for object detection [56].

### Gradient Descent Optimization and Backpropagation Algorithm

Gradient descent optimization and backpropagation are two critical components of DL algorithms that are used for training neural networks. Gradient descent optimization is used to minimize the error between the predicted output and the actual output. The error is represented by a cost function that is defined by the choice of the neural network model. The goal of gradient descent is to find the optimal set of weights that minimize the cost function [39].

The algorithm works by calculating the gradient of the cost function with respect to each weight in the network. This gradient provides the direction in which the weights should be updated (see Figure 8). The learning rate ( $\delta$ ) is a hyperparameter that determines how much the weights should be changed in each iteration. A higher learning rate can result in faster convergence, but it can also lead to overshooting the optimal weights and oscillation around the minimum [39].

Backpropagation is the process of computing the gradient of the cost function with respect to the weights of the neural network. It is an efficient algorithm that allows us to calculate the gradient using the chain rule of calculus. The errors are propagated backward through the network, and the gradient of the cost function is calculated at each layer [39]. The weights are then updated using the gradient descent algorithm, which involves subtracting a fraction of the gradient from the current weight value (see Figure 8). The process is repeated for many iterations until the cost function converges to a minimum value.

Gradient descent optimization and backpropagation are critical components of DL algorithms that are used for training neural networks. These techniques allow us to minimize the error between the predicted output and the actual output and update the weights of the network to improve its performance.

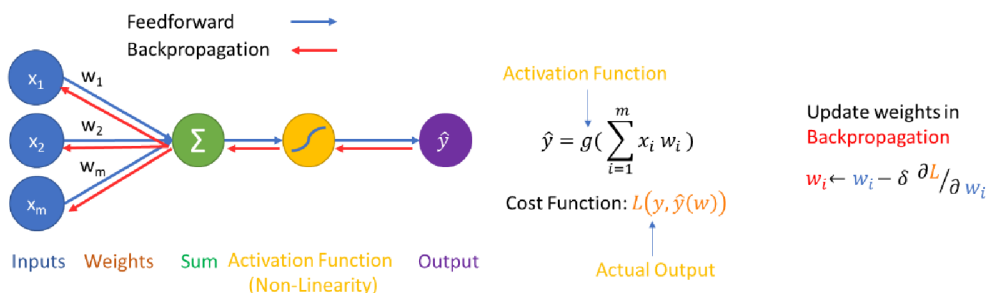


Figure 8 A single-layer neural network with a non-linear activation function highlighting the feedforward and backpropagation steps

## 2. STATE OF THE ART OF DL IN MRS

The recent success of DL, one of the latest machine learning approaches, in a variety of tasks, including applications with a low signal-to-noise ratio (SNR) [39], [57], suggests that it might also handle the spectral analysis of an MRS signal. Supervised DL-based approaches have been used for ghosting artifacts detection and removal [58], spectral reconstruction [59], automatic peak picking [60], MRSI spatial resolution enhancement [61], localized correlated spectroscopy acceleration [62], metabolites and MM separation in MRS signals, the quantification and noise removal of MRSI signals [63], [64], and poor-quality spectra identification [65].

It has been shown that supervised DL can also be employed for FPC [13], [19] and could speed up FPC once it has been successfully trained. This supervised approach, using two separate networks in sequence to estimate frequency and phase, showed encouraging results. The first network was trained for frequency shift estimation using the magnitude of frequency- and phase-shifted spectrum as the input and the known frequency shift as the output. Subsequently, the second network was trained for phase shift estimation using real parts of the frequency-corrected spectrum as the input and phase shift as the output. In this approach, any error in the first step (frequency correction) may bias the phase shift estimation. Training two networks is a computationally expensive task. Moreover, the networks were trained in a supervised manner using simulated data. Any discrepancy between the *in-vivo* and the simulated spectra may result in errors in frequency and phase shift estimation. The true output values are unknown in MRS data, and obtaining hundreds of spectra with labeled frequency and phase shifts is almost infeasible. This makes it challenging to use supervised DL methods that rely on labeled transients. Self-supervised or unsupervised learning may eliminate the drawbacks of supervised learning. Moreover, several studies demonstrated the potential of artificial neural networks for quantifying MR spectra. Hiltunen et al. [66] have demonstrated the feasibility of constructing a quantifying analyzer for long echo time (TE) *in vivo* proton MRS (<sup>1</sup>H NMR) spectra using artificial neural networks with magnitude spectra. Hatami et al. [67] and Lee et al. [68] applied supervised DL approaches to metabolite quantification and presented results comparable to conventional LCM approaches. Chandler et al. [69] also applied a supervised DL approach to study metabolite quantification in edited human brain MRS spectra. These studies utilized supervised learning approaches, in which the input and the output were simulated spectra and known values, respectively. The true output values are however unknown in *in-vivo* MRS data. Moreover, a network trained in a supervised manner using simulated data might be prone to overfit training data [70]; 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. Bhat et al. [71] investigated the application of a radial

basis function neural network (RBFNN) for the automatic quantification of short echo time, multi-voxel, phased spectral data. Gurbani et al. [72] presented a self-supervised DL architecture that integrates a CNN with peak fitting for quantifying MR spectra. In their approach, a deep autoencoder is used as a framework for self-supervised or unsupervised learning. However, their method does not utilize the advantages of LCM such as fewer fitting parameters and realistic basis spectra. Table 1 provides a comparison of related work on MRS signal quantification using DL.

Even though DL algorithms have demonstrated equivalent quantitation performance to traditional methods, concerns have been raised about their robustness. Moreover, the effects of DL architectures, spectroscopic input types, and learning designs for optimal quantification in MRS of pathological spectra have not been investigated in previous studies and warrant further investigation.

Table 1 A summary of related work on MRS signal quantification using DL.

	<b>Model architectures</b>	<b>Input types</b>	<b>Learning process</b>	<b>Data type (training)</b>	<b>Prior metabolite resonances model</b>
<b>Hiltunen et al. [66]</b>	Shallow Neural Networks	Magnitude spectra (1D)	Supervised	Simulated	–
<b>Bhat et al. [71]</b>	Radial basis function neural network	Real part of spectra (1D)	Self-supervised	<i>In-vivo</i>	Lorentzian-Gaussian lineshape
<b>Hatami et al. [67]</b>	Convolutional neural network	Complex spectra (1D)	Supervised	Simulated	–
<b>Lee et al. [68]</b>	Convolutional neural network	Complex spectra (1D)	Supervised	Simulated	–
<b>Chandler et al. [69]</b>	Convolutional neural network	Real, imaginary and magnitude component of spectra (2D)	Supervised	Simulated	–
<b>Gurbani et al. [72]</b>	Convolutional neural network	Real part of spectra (1D)	Self-supervised	<i>In-vivo</i>	Lorentzian-Gaussian lineshape



### **3. AIMS OF THE DOCTORAL THESIS**

The primary objective of this dissertation is to introduce and verify an advanced approach for rapid, efficient, and accurate quantification of metabolites using DL techniques in MRS. Moreover, this dissertation aims to construct a deep neural network that extracts features from FID signals, i.e., MRS signals in the time domain, for preprocessing and analyzing MRS data. The objectives of this research can be summarized as follows:

1. To develop a robust DL method for preprocessing MR data, including frequency and phase correction,
2. To design and implement a fast and appropriate DL-based approach for quantifying MR spectra,
3. To propose a self-supervised approach for quantifying concentrations in *in-vivo* spectra,
4. To propose a method for estimating the uncertainty of the concentration's estimation in DL-based methods.
5. To compare the effectiveness and reliability of DL-based methods with nonlinear least-squares (NLLS) fitting methods for quantification,
6. To propose a standard data format to facilitate data sharing among research groups for artificial intelligence applications.

## 4. SELECTED PUBLICATIONS

The present thesis is composed of six research papers (4 IF-journal publications, and 2 conference extended papers), which are interconnected in their scope and theme. The central focus of this work is the development of DL approaches for MRS data analysis, with a particular emphasis on frequency and phase correction, metabolite quantification, and data standardization.

### **MRS data preprocessing:**

The first paper [20] presents a model-informed unsupervised DL approach to the frequency and phase correction of MRS signals. The feasibility and efficiency of physics-informed DL-based signal processing of MR spectroscopy data in an unsupervised manner were investigated. Simulated, phantom, and *in vivo* MEGA-edited MRS data were used in the study. This work aimed to solve the challenge of obtaining spectra with labels (i.e. spectra with labeled frequency and phase shifts) for the supervised DL approach to FPC.

### **Metabolite quantification (supervised DL approaches):**

The next three papers focus on the development of supervised DL approaches to MRS signal quantification. The second paper [73] proposes a time-frequency analysis approach that leverages DL to produce highly accurate and robust MRS signal quantitation. This paper verified the hypothesis that DL in combination with time-frequency analysis can be used for metabolite quantification and yielded results more robust than DL trained with MR signals in the frequency domain [67]. This paper utilizes a CNN, but the training of CNNs is computationally intensive and its optimal architecture and hyper-parameters are not well understood. To address this issue, the third paper [74] utilizes a Wavelet Scattering Convolution Network (WSCN), which is a well-understood and computationally cheap. The WSCN approach achieved better quantification accuracy and computational efficiency compared to the CNN-based approach proposed in the second paper. The fourth paper [75] investigates the quantification of MR spectra by a supervised DL approach in an idealized setting, examining various forms of input, network architectures, optimization by ensembles of networks, and training bias. This work aimed to address concerns about the robustness of DL for MR spectra quantification.

### **Metabolite quantification (self-supervised DL approaches):**

At the beginning of this Ph.D. study, the author was captivated by the idea of using supervised learning to solve complex problems in the quantification of MR spectra. The author spent several months exploring different supervised learning techniques and experimenting with various datasets. However, as the author delved deeper into the subject, the author began to realize the limitations of supervised learning. At first, the

author was hesitant to try self-supervised learning, as the author noticed that developing a neural network and training it in a self-supervised manner were much more challenging than supervised learning. However, the author quickly discovered that self-supervised learning offered a wealth of opportunities to explore new ideas and solve problems in innovative ways. With self-supervised learning, the author proposed a physics-informed DL approach to quantifying human brain metabolites from MRS data in the fifth paper [76]. The author ventured to develop an approach in which neural networks can learn in a self-supervised manner to solve an inverse problem for human brain metabolite concentration estimation.

The fourth and fifth papers shed light on a comparison between DL-based methods (supervised and self-supervised) and traditional methods (nonlinear least-squares fitting methods) for MRS data quantification. Additionally, the fifth paper's proposed method utilizes a physics-informed DL approach, allowing it to compute the Cramer-Rao lower bounds (CRLB) to estimate the uncertainty in concentration estimation. Table 2 provides a comparison of the author's work on MRS signal quantification using DL.

Table 2 A summary of our work on MRS signal quantification using DL.

	<b>Model architectures</b>	<b>Input types</b>	<b>Learning process</b>	<b>Data type (training)</b>	<b>Prior metabolite resonances model</b>	<b>Output of network</b>
<b>Shamaei et al.</b> [73]	Convolutional neural network	Complex time-frequency domain scalogram (wavelet coefficients) (2D)	Supervised	Simulated <sup>a</sup>	–	Amplitudes
<b>Shamaei et al.</b> [77]	Wavelet scattering network	Complex time-domain signal (FID) (1D)	Supervised	Simulated <sup>b</sup>	–	Amplitudes
<b>Rizzo et al.</b> [78]	Convolutional neural network	Complex spectra, time-frequency domain spectrograms (2D)	Supervised	Simulated <sup>c</sup>	–	Amplitudes
<b>Shamaei et al.</b> [76]	Physics-informed convolutional neural network	Complex time-domain signal (FID) (1D)	Self-supervised	Simulated <sup>d</sup> and <i>in-vivo</i>	Quantum-mechanics simulated metabolite responses	Parameters of a complex model

<sup>a</sup> a combination of amplitude-scaled frequency-shifted damped metabolite basis set signals, the baseline (without noise).

<sup>b</sup> a combination of amplitude-scaled frequency-shifted phase-shifted damped metabolite basis set signals, the baseline and white noise.

<sup>c</sup> a combination of amplitude-scaled damped metabolite basis set signals, the baseline and white noise (all basis set signals share the same damping factor).

<sup>d</sup> a combination of amplitude-scaled frequency-shifted phase-shifted damped metabolite basis set signals, the baseline and white noise (all basis set signals share the same damping factor, frequency shift, and phase shift).

**Standard data format to facilitate data sharing among research groups:**

Finally, the sixth paper [79] advocates for a standard data format for MRS, which could facilitate data sharing and comparability across research groups for artificial intelligence applications.

Taken together, these papers present an overview of the state-of-the-art in DL approaches to MRS data analysis. They offer insights into the potential of these methods to improve the efficiency, reproducibility, and reliability of MRS data analysis, and highlight some of the challenges and opportunities that lie ahead in this rapidly evolving field. The thesis concludes with a synthesis of the key findings and recommendations for future research.

## 4.1 Paper 1 – Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals

### Citation

- [1] Shamaei, A, Starcukova, J, Pavlova, I, Starcuk, Z. Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals. *Magn Reson Med.* 2023; 89: 1221– 1236. doi:10.1002/mrm.29498

### Paper contribution

The paper proposes and investigates the feasibility and efficiency of two novel unsupervised deep learning-based methods for frequency and phase correction (FPC) of magnetic resonance spectroscopy (MRS) data. These proposed methods utilize a priori physics domain knowledge to improve the performance of FPC. The paper presents the training, validation, and evaluation of these methods using simulated, phantom, and *in vivo* MEGA-edited MRS data. This study proposes a new measure to evaluate the FPC method performance and compares the performance of the proposed methods with other commonly used FPC methods. This study also evaluates the ability of the proposed methods to perform FPC at varying signal-to-noise ratios (SNR). Additionally, a Monte Carlo study is conducted to investigate the performance of the proposed methods. The contribution of the paper lies in the development and evaluation of two novel unsupervised deep learning-based FPC methods for MRS data, which can improve the accuracy and efficiency of FPC in various MRS applications.

### Author's contribution

**Conceptualization:** *The overall research direction and framework were developed by the author.*

**Methodology:** *A novel methodology for frequency and phase correction (FPC) of magnetic resonance spectroscopy (MRS) data was designed and implemented by the author.*

**Software:** *The software in Python programming language used for the research was developed and maintained by the author.*

**Formal Analysis:** *The data were analyzed by the author using statistical and computational methods.*

**Writing - Original Draft:** *The initial manuscript was written by the author.*

**Writing - Review & Editing:** *The manuscript was reviewed and edited for clarity and coherence by the author.*

**Visualization:** *The results were visualized and presented by the author.*

## **About Journal**

*Magnetic Resonance in Medicine, Q1[Radiology, Nuclear Medicine and Imaging, Impact factor: 4.108.*

## **Copyright notice**

© 2022 Ústav přístrojové techniky AV ČR, v. v. i. Magnetic Resonance in Medicine published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. doi:10.1002/mrm.29498. The author of the thesis is the main author of the article.

## 4.2 Paper 2 – Deep Learning For Magnetic Resonance Spectroscopy Quantification: A Time-Frequency Analysis Approach

### Citation

- [2] Shamaei, Amirmohammad. Deep Learning For Magnetic Resonance Spectroscopy Quantification: A Time-Frequency Analysis Approach. In: *Proceedings II of the 26st Conference STUDENT EEICT 2020: Selected papers* [online]. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2020, s. 131-135 [cit. 2023-02-16]. ISBN 978-80-214-5868-0. Dostupné z: <http://hdl.handle.net/11012/200638>

### Paper contribution

This paper explores the combination of deep learning and time-frequency analysis for more reliable metabolite quantification in magnetic resonance spectroscopy (MRS). This study verifies the hypothesis that this combination can produce more robust results than deep learning trained on MR signals in the frequency domain. The study uses the complex matrix of absolute wavelet coefficients for the time-frequency representation of the signal and implements convolutional neural networks (CNN) for deep learning. The paper also presents a comparison with DL used for the quantification of data in the frequency domain. Overall, the paper's contribution lies in providing an innovative approach to MRS quantification and advancing the understanding of the potential of deep learning and time-frequency analysis in this field.

### Author's contribution

**Conceptualization:** *The overall research direction and framework were developed by the author.*

**Methodology:** *A novel methodology for the combination of deep learning and time-frequency analysis for quantification of magnetic resonance spectroscopy data was designed and implemented by the author.*

**Software:** *The software in MATLAB used for the research was developed and maintained by the author.*

**Formal Analysis:** *The data were analyzed by the author using statistical and computational methods.*

**Writing - Original Draft:** *The initial manuscript was written by the author.*

*Writing - Review & Editing:* The manuscript was reviewed and edited for clarity and coherence by the author.

*Visualization:* The results were visualized and presented by the author.

## **Copyright notice**

This is the published version of the article published in Proceedings II of the 26st Conference STUDENT EEICT 2020: Selected Papers. s. 131-135. ISBN 978-80-214-5868-0, <https://conf.feec.vutbr.cz/eeict/EEICT2020>. The author of the thesis is the main author of the article.



## 4.3 Paper 3 – A Wavelet Scattering Convolution Network for Magnetic Resonance Spectroscopy Signal Quantitation

### Citation

- [3] Shamaei A., Starčuková J. and Starčuk Jr. Z. (2021). A Wavelet Scattering Convolutional Network for Magnetic Resonance Spectroscopy Signal Quantitation. In Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOSIGNALS, ISBN 978-989-758-490-9, pages 268-275. DOI: 10.5220/0010318502680275

### Paper contribution

The paper's contribution is to explore the use of a Wavelet Scattering Convolutional Network (WSCN) for magnetic resonance spectroscopy (MRS) signal quantification. The paper highlights that the most widely used network for MRS signal quantification is the Convolutional Neural Network (CNN), but that its optimal architecture and hyper-parameters for MRS are not well understood. The paper shows that a WSCN, which is well-understood and computationally cheap, could yield more robust results for metabolite quantification than one of the quantitation methods based on model fitting (QUEST) and equivalent results to a CNN while being faster. The study investigates the effects of phase, noise, and macromolecule variation on the WSCN estimation accuracy. Overall, the paper's contribution is to propose an alternative deep learning approach to MRS signal quantification that is more computationally efficient and potentially more accurate than existing methods.

### Author's contribution

**Conceptualization:** *The overall research direction and framework were developed by the author.*

**Methodology:** *A novel methodology for the use of a Wavelet Scattering Convolutional Network (WSCN) for quantification of magnetic resonance spectroscopy data was designed and implemented by the author.*

**Software:** *The software in MATLAB used for the research was developed and maintained by the author.*

**Formal Analysis:** *The data were analyzed by the author using statistical and computational methods.*

**Writing - Original Draft:** *The initial manuscript was written by the author.*

*Writing - Review & Editing:* The manuscript was reviewed and edited for clarity and coherence by the author.

*Visualization:* The results were visualized and presented by the author.

## **Copyright notice**

This is the published version of the article published in Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOSIGNALS, ISBN 978-989-758-490-9. The author of the thesis is the main author of the article. The author of the thesis is the main author of the article.

## **4.4 Paper 4 – Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias.**

### **Citation**

- [4] Rizzo, R, Dziadosz, M, Kyathanahally, SP, Shamaei, A, Kreis, R. Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias. *Magn Reson Med.* 2022; 1- 21. doi:10.1002/mrm.29561

### **Paper contribution**

The paper explores the application of deep learning (DL) architectures, spectroscopic input types, and learning designs for optimal quantification in magnetic resonance spectroscopy (MRS) of simulated pathological spectra. The study investigates 24 different DL architectures, with active learning through altered training and testing data distributions to optimize quantification performance. Ensembles of networks are explored to improve DL robustness and reduce the variance of estimates. The paper compares the performance of DL predictions and traditional model fitting (MF) using a set of scores. The results show that ensembles of heterogeneous networks that combine 1D frequency-domain and 2D time-frequency domain spectrograms as input perform best, and dataset augmentation with active learning can improve performance but gains are limited. MF is more accurate, although DL appears to be more precise at low signal-to-noise ratios (SNRs). However, the overall improved precision of DL predictions originates from a strong bias for cases with high uncertainty

toward the dataset the network has been trained with, tending toward its average value. The paper also highlights potential intrinsic biases on training sets, which are dangerous in a clinical context that requires the algorithm to be unbiased to outliers (i.e., pathological data). The contribution of the paper lies in providing a comprehensive evaluation of DL architectures and learning designs for MRS quantification, and highlighting the importance of unbiased and robust MRS quantification algorithms in a clinical context.

### **Author's contribution**

*Conceptualization:* The overall research direction and framework were developed by the author.

*Formal Analysis: The data were analyzed by the author using statistical and computational methods.*

*Writing - Review & Editing: The manuscript was reviewed and edited for clarity and coherence by the author.*

## **About Journal**

*Magnetic Resonance in Medicine, Q1[Radiology, Nuclear Medicine and Imaging, Impact factor: 4.108.*

## **Copyright notice**

© 2022 The Authors. Magnetic Resonance in Medicine published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## 4.5 Paper 5 – Physics-informed Deep Learning Approach to Quantification of Human Brain Metabolites from Magnetic Resonance Spectroscopy Data.

### Citation

- [5] Shamaei, A, Starcukova, J, Starcuk, Z. Physics-informed deep learning approach to quantification of human brain metabolites from magnetic resonance spectroscopy data. *Computers in Biology and Medicine*. 2023; 158: 106837. doi: 10.1016/j.combiomed.2023.106837

### Paper contribution

The paper presents a novel, self-supervised deep learning (DL) method for the quantification of magnetic resonance spectroscopy (MRS) and magnetic resonance spectroscopic imaging (MRSI) data. This method is based on a linear combination model (LCM) and uses quantum-mechanics simulated metabolite responses and neural networks for the quantification of relative metabolite concentrations. The proposed DL-based method is evaluated and compared to traditional methods using simulated and publicly accessible *in-vivo* human brain MRS data. The paper also includes a novel adaptive macromolecule fitting algorithm. The performance of the proposed methods is investigated in a Monte Carlo study. The contribution of this paper lies in the development and evaluation of a self-supervised DL-based method for MRS data analysis that does not require ground truth fitted spectra, which is not always practical. This method can improve the accuracy and efficiency of MRS data analysis in various applications. To our knowledge, this is the first report showing the feasibility of the physics-informed self-supervised DL method for the quantification of MRS data.

### Author's contribution

**Conceptualization:** *The overall research direction and framework were developed by the author.*

**Methodology:** *A novel methodology for the quantification of magnetic resonance spectroscopy (MRS) utilizing self-supervised deep learning (DL) method was designed and implemented by the author.*

**Software:** *The software in Python programming language used for the research was developed and maintained by the author.*

***Formal Analysis:** The data were analyzed by the author using statistical and computational methods.*

***Writing - Original Draft:** The initial manuscript was written by the author.*

***Writing - Review & Editing:** The manuscript was reviewed and edited for clarity and coherence by the author.*

***Visualization:** The results were visualized and presented by the author.*

## **About Journal**

*Computers in Biology and Medicine, Q1[Computer Science Applications], Impact factor: 6.698.*

## **Copyright notice**

© 2023 The Authors. Published by Elsevier Ltd.

This is an open access article distributed under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. You are not required to obtain permission to reuse this article. The author of the thesis is the main author of the article.

## 4.6 Paper 6 – NIfTI-MRS: A standard data format for magnetic resonance spectroscopy

### Citation

- [6] Clarke, W, Mikkelsen, M, Oeltzschner, G, Bell T.K., Shamaei, A, Soher, B.J., Emir, U, Wilson, W. NIfTI-MRS: A standard data format for magnetic resonance spectroscopy. *Magn Reson Med.* 2022; 88: 2358- 2370. doi:10.1002/mrm.29418

### Paper contribution

The contribution of this paper is the proposal of a standardized format, NIfTI-MRS, for magnetic resonance spectroscopy (MRS) data that incorporates essential spectroscopic metadata and additional encoding dimensions, and its implementation as an extension to the Neuroimaging informatics technology initiative (NIfTI) format. The standard format allows for easy data sharing, algorithm development, and integration of MRS analysis with other imaging modalities. The paper provides a detailed description of the NIfTI-MRS format specification, an open-source command-line conversion program to convert MRS data to NIfTI-MRS, and a dedicated plugin for FSLeves, the FMRIB Software Library (FSL) image viewer for visualization of data in the proposed format. The paper also includes online documentation, 10 example datasets in the proposed format, and code examples of NIfTI-MRS readers implemented in common programming languages.

### Author's contribution

**Conceptualization:** *The overall research direction and framework were developed by the author.*

**Methodology:** *The research methodology was designed and implemented by the author.*

**Software:** *The software in JAVA programming language used for the research was developed and maintained by the author.*

**Writing - Review & Editing:** *The manuscript was reviewed and edited for clarity and coherence by the author.*

**Visualization:** *The results were visualized and presented by the author.*

### About Journal

*Magnetic Resonance in Medicine, Q1[Radiology, Nuclear Medicine and Imaging, Impact factor: 4.108.*

## **Copyright notice**

© 2022 The Authors. Magnetic Resonance in Medicine published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



## 5. CONCLUSION

In summary, this dissertation aimed to develop a novel DL-based approach for rapid and accurate quantification of metabolites in magnetic resonance spectroscopy.

The proposed methods achieved the objectives of developing a robust DL method for preprocessing MR data [20], designing and implementing a fast and efficient DL-based approaches for quantifying MR spectra [73]–[76], investigating the applicability of the proposed approaches for quantifying concentrations in *in-vivo* spectra [76], and comparing the effectiveness and reliability of DL-based methods with NLLS fitting methods for quantification [75], [76].

The results of this study demonstrate that the proposed DL-based approaches can improve the speed and accuracy of MRS signal preprocessing and quantification in a self-supervised manner [76]. The performance of our methods on synthetic data is comparable with the traditional methods in terms of accuracy in a shorter amount of processing time. Furthermore, a standard data format was proposed to facilitate data sharing among research groups for artificial intelligence applications [79].

The results achieved in this research, corresponding to the objectives outlined in the Aims of the doctoral thesis section, can be summarized as follows:

1. Two novel unsupervised DL-based FPC methods for MRS data have been developed, which can improve the accuracy and efficiency of FPC in various MRS applications. The results have been published in an IF journal [Paper 1].
2. Fast and efficient DL-based solutions for quantifying MR spectra have been developed and tested on simulated and *in-vivo* data, and the results have been published in conferences and IF journals [Papers 2, 3, 4, and 5].
3. The applicability of the proposed self-supervised approach for quantifying relative concentrations in *in-vivo* spectra has been investigated [Paper 5].
4. The functionality of the proposed self-supervised approach has been extended to estimate the uncertainty of the concentration's estimation [Paper 5 and Item 10 in Appendix A -(An abstract based on the findings of this thesis has been accepted for presentation at International Society for Magnetic Resonance in Medicine conference in 2023, in Toronto. Additionally, a manuscript based on this research is currently under preparation for submission to a high-impact factor journal.)].
5. A comprehensive comparison between DL-based methods and traditional NLLS fitting methods for MRS data quantification has been made, and the results have been published in two IF journals [Papers 4 and 5].

6. In collaboration with an international group of experts, a standard data format has been proposed to facilitate data sharing among research groups, and the results have been published in an IF journal [Paper 6]. The proposed format has been used in the present study for data sharing.

This research opens the door to further exploration of the applications of DL techniques in MR spectroscopy and spectroscopic imaging signal processing, potentially leading to significant advancements in medical diagnosis. Some potential applications of DL in MR spectroscopy and spectroscopic imaging for future work are:

**Automated analysis:** DL algorithms can be trained to automatically analyze MRS and MRSI data, reducing the time and effort required for manual analysis,

**Improved signal-to-noise ratio:** DL algorithms can be used to denoise MR signals [Item 7 in Appendix A (An abstract was presented by the author at the European Society of Magnetic Resonance in Medicine and Biology (ESMRMB) conference 2021)], improving the signal-to-noise ratio and enabling higher-quality data,

**Quantitative analysis:** DL algorithms can be trained to perform quantitative analysis of MR spectroscopic data, allowing for the automated calculation of metabolite concentrations and other important parameters,

**Image segmentation:** DL algorithms can be used for automatic image segmentation, allowing for the separation of different tissues and structures within MR spectroscopic images,

**Signal classification:** DL algorithms can be trained to classify MRS signals based on specific features, allowing for improved diagnosis and treatment planning.

Overall, DL has the potential to significantly enhance the capabilities of MR spectroscopy and spectroscopic imaging in medical imaging. It is important to note that while DL can greatly aid in the analysis and interpretation of MR data, it should be used in conjunction with traditional methods and the expertise of experienced radiologists to ensure accurate and reliable results.

## LITERATURE

- [1] M. Van Der Graaf, "In vivo magnetic resonance spectroscopy: Basic methodology and clinical applications," *Eur. Biophys. J.*, vol. 39, no. 4, pp. 527–540, 2010.
- [2] Robin A. de Graaf, "In Vivo NMR Spectroscopy," *In Vivo NMR Spectroscopy*. p. 560, 13-Feb-2019.
- [3] C. Stagg and D. Rothman, *Magnetic Resonance Spectroscopy: Tools for Neuroscience Research*. 2014.
- [4] J.-B. Pouillet, D.-M. Sima, and S. va. Huffel, "MRS signal quantification: a review of time- and frequency-domain methods," *J. Magn. Reson.*, vol. 195, no. September 2008, pp. 134–144, 2008.
- [5] "Magnetic Resonance Spectroscopy (MRS) - Medical Clinical Policy Bulletins | Aetna." [Online]. Available: [https://www.aetna.com/cpb/medical/data/200\\_299/0202.html#dummyLink1](https://www.aetna.com/cpb/medical/data/200_299/0202.html#dummyLink1). [Accessed: 14-Mar-2023].
- [6] J. Keeler, *Understanding NMR Spectroscopy*. Wiley, 2002.
- [7] T. Rachman, *Robin A. de Graaf - In Vivo NMR Spectroscopy Principles and Techniques (2018, Wiley)*. 2018.
- [8] J. Near *et al.*, "Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations.," *NMR Biomed.*, no. July 2019, p. e4257, 2020.
- [9] G. Öz *et al.*, "Advanced single voxel 1H magnetic resonance spectroscopy techniques in humans: Experts' consensus recommendations," *NMR Biomed.*, vol. 34, no. 5, pp. 1–18, 2021.
- [10] J. Near, R. Edden, C. J. Evans, A. Harris, and P. Jezzard, "Frequency and Phase Drift Correction of Magnetic Resonance Spectroscopy Data by Spectral Registration in the Time Domain," *Magn. Reson. Med.*, vol. 50, pp. 44–50, 2015.
- [11] E. C. Wieggers, B. W. J. Philips, A. Heerschap, and M. van der Graaf, "Automatic frequency and phase alignment of in vivo J-difference-edited MR spectra by frequency domain correlation," *MAGMA*, vol. 30, no. 6, p. 537, Dec. 2017.
- [12] B. Keating *et al.*, "Prospective motion correction for single-voxel 1H MR spectroscopy," *Magn. Reson. Med.*, vol. 64, no. 3, pp. 672–679, 2010.
- [13] S. Tapper *et al.*, "Frequency and phase correction of J-difference edited MR spectra using deep learning," *Magn. Reson. Med.*, vol. 85, no. 4, pp. 1755–1765, 2021.
- [14] G. Helms and A. Piringer, "Restoration of motion-related signal loss and line-shape deterioration of proton MR spectra using the residual water as intrinsic reference," *Magn. Reson. Med.*, vol. 46, no. 2, pp. 395–400, 2001.
- [15] M. Wilson, "Robust retrospective frequency and phase correction for single-voxel MR spectroscopy," *Magn. Reson. Med.*, vol. 81, no. 5.
- [16] M. Mikkelsen, S. Tapper, J. Near, S. H. Mostofsky, and R. A. E. Edden, "Correcting frequency and phase offsets in MRS data using robust spectral registration," no. June, pp. 1–12, 2020.
- [17] T. Ernst and J. Li, "A novel phase and frequency navigator for proton magnetic resonance spectroscopy using water-suppression cycling," *Magn. Reson. Med.*, vol. 65, no. 1, pp. 13–17, Jan. 2011.
- [18] K. W. Waddell, M. J. Avison, J. M. Joers, and J. C. Gore, "A practical guide to

- robust detection of GABA in human brain by J-difference spectroscopy at 3 T using a standard volume coil,” *Magn. Reson. Imaging*, vol. 25, no. 7, pp. 1032–1038, Sep. 2007.
- [19] D. J. Ma *et al.*, “MR spectroscopy frequency and phase correction using convolutional neural networks,” *Magn. Reson. Med.*, vol. 87, no. 4, pp. 1700–1710, Apr. 2022.
- [20] A. Shamaei, J. Starcukova, I. Pavlova, and Z. Starcuk, “Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals,” *Magn. Reson. Med.*, vol. 89, no. 3, pp. 1221–1236, 2023.
- [21] R. Kreis *et al.*, *Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: Background and experts’ consensus recommendations*, no. September 2019. 2020.
- [22] C. Stagg and D. L. Rothman, *Magnetic resonance spectroscopy: tools for neuroscience research and emerging clinical applications*. Academic Press, 2013.
- [23] J. B. Pouillet, D. M. Sima, and S. Van Huffel, “MRS signal quantitation: A review of time- and frequency-domain methods,” *J. Magn. Reson.*, vol. 195, no. 2, pp. 134–144, 2008.
- [24] J. Near *et al.*, “Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts’ consensus recommendations,” *NMR Biomed.*, no. July 2019, pp. 1–23, 2020.
- [25] Z. Starčuk and J. Starčuková, “Quantum-mechanical simulations for in vivo MR spectroscopy: Principles and possibilities demonstrated with the program NMRScopeB,” *Anal. Biochem.*, vol. 529, 2017.
- [26] K. Landheer, K. M. Swanberg, and C. Juchem, “Magnetic resonance Spectrum simulator ( MARSS ), a novel software package for fast and computationally efficient basis set simulation,” no. May, pp. 1–13, 2019.
- [27] S. W. Provencher, “Automatic quantitation of localized in vivo <sup>1</sup>H spectra with LCModel,” *NMR Biomed.*, vol. 14, no. 4, pp. 260–264, 2001.
- [28] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly, “Time-domain semi-parametric estimation based on a metabolite basis set,” *NMR Biomed.*, vol. 18, no. 1, pp. 1–13, 2005.
- [29] M. Wilson, G. Reynolds, R. A. Kauppinen, T. N. Arvanitis, and A. C. Peet, “A constrained least-squares approach to the automated quantitation of in vivo <sup>1</sup>H magnetic resonance spectroscopy data.,” *Magn. Reson. Med.*, vol. 65, no. 1, pp. 1–12, 2011.
- [30] D. G. Q. Chong, R. Kreis, C. S. Bolliger, C. Boesch, and J. Slotboom, “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.*, vol. 24, no. 3, pp. 147–164, 2011.
- [31] W. T. Clarke, C. J. Stagg, and S. Jbabdi, “FSL-MRS: An end-to-end spectroscopy analysis package,” *Magn. Reson. Med.*, vol. 85, no. 6, pp. 2950–2964, Jun. 2021.
- [32] G. Oeltzschner *et al.*, “Osprey: Open-source processing, reconstruction & estimation of magnetic resonance spectroscopy data.,” *J. Neurosci. Methods*, vol. 343, p. 108827, Sep. 2020.
- [33] D. Graveron-Demilly, “Quantification in magnetic resonance spectroscopy based

- on semi-parametric approaches,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 27, no. 2. Springer Verlag, pp. 113–130, 28-Jul-2014.
- [34] G. Oeltzschner, H. J. Zöllner, S. C. N. Hui, M. Mikkelsen, and G. Muhammad, “Osprey : Open-Source Processing , Reconstruction & Estimation of Magnetic Resonance Spectroscopy Data,” 2020.
- [35] A. Fuchs, P. Boesiger, R. F. Schulte, and A. Henning, “ProFit revisited,” *Magn. Reson. Med.*, vol. 71, no. 2, pp. 458–468, 2014.
- [36] M. I. Osorio-Garcia *et al.*, “Quantification of in vivo 1H magnetic resonance spectroscopy signals with baseline and lineshape estimation,” *Meas. Sci. Technol.*, vol. 22, no. 11, pp. 1–17, 2011.
- [37] C. Jenkins, M. Chandler, F. Langbein, and S. Shermer, “Quantification of edited magnetic resonance spectroscopy: a comparative phantom based study of analysis methods,” May 2019.
- [38] M. Marjańska *et al.*, “Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM,” *Magn. Reson. Med.*, vol. 87, no. 1, pp. 11–32, 2022.
- [39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [40] I. G. and Y. B. and A. Courville, “Deep learning by Ian Goodfellow, Yoshua Bengio, Aaron Courville,” *Nature*, vol. 29, no. 7553, pp. 1–73, 2016.
- [41] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42. Elsevier B.V., pp. 60–88, 01-Dec-2017.
- [42] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.
- [43] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering*. Cambridge University Press, 2019.
- [44] J. Schmidhuber, “Deep Learning in Neural Networks: An Overview,” 2014.
- [45] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [46] I. Tolstikhin *et al.*, “MLP-Mixer: An all-MLP Architecture for Vision,” *arXiv*, 2021.
- [47] N. Lange, C. M. Bishop, and B. D. Ripley, “Neural Networks for Pattern Recognition,” *J. Am. Stat. Assoc.*, vol. 92, no. 440, 1997.
- [48] D. Learning, “Deep Learning - Goodfellow,” *Nature*, vol. 26, no. 7553, 2016.
- [49] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nat. 2015 5217553*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [50] S. Liu *et al.*, “Deep learning for electroencephalogram (EEG) classification tasks: a review,” *J. Neural Eng.*, vol. 16, no. 3, p. 031001, Apr. 2019.
- [51] A. Phinyomark and E. Scheme, “EMG Pattern Recognition in the Era of Big Data and Deep Learning,” *Big Data Cogn. Comput. 2018, Vol. 2, Page 21*, vol. 2, no. 3, p. 21, Aug. 2018.
- [52] D. Shen, G. Wu, and H. Il Suk, “Deep Learning in Medical Image Analysis,” <https://doi.org/10.1146/annurev-bioeng-071516-044442>, vol. 19, pp. 221–248, Jun. 2017.
- [53] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

- [54] W. H. Lopez Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, "Autoencoders," *Mach. Learn. Methods Appl. to Brain Disord.*, pp. 193–208, Mar. 2020.
- [55] G. H. Yann LeCun, Yoshua Bengio, "Deep learning (2015), Y. LeCun, Y. Bengio and G. Hinton," *Nature*, 2015.
- [56] A. Jaiswal, A. Ramesh Babu, M. Zaki Zadeh, D. Banerjee, and F. Makedon, "A Survey on Contrastive Self-Supervised Learning," 2020.
- [57] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Z. Med. Phys.*, vol. 29, no. 2, pp. 102–127, May 2019.
- [58] R. Kreis and S. P. Kyathanahally, "Deep Learning Approaches for Detection and Removal of Ghosting Artifacts in MR Spectroscopy," vol. 863, pp. 851–863, 2018.
- [59] H. Lee, H. H. Lee, and H. Kim, "Reconstruction of spectra from truncated free induction decays by deep learning in proton magnetic resonance spectroscopy," *Magn. Reson. Med.*, no. September 2019, pp. 1–10, 2020.
- [60] P. Klukowski, M. Augoff, M. ZieRba, M. Drwal, A. Gonczarek, and M. J. Walczak, "NMRNet: A deep learning approach to automated peak picking of protein NMR spectra," *Bioinformatics*, vol. 34, no. 15, pp. 2590–2597, 2018.
- [61] Z. Iqbal, D. Nguyen, G. Hangel, S. Motyka, W. Bogner, and S. Jiang, "Super-Resolution 1H Magnetic Resonance Spectroscopic Imaging Utilizing Deep Learning," *Front. Oncol.*, vol. 9, no. October, pp. 1–13, 2019.
- [62] Z. Iqbal, D. Nguyen, M. A. Thomas, and S. Jiang, "Deep learning can accelerate and quantify simulated localized correlated spectroscopy," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021.
- [63] Y. Li, Z. Wang, and F. Lam, "Separation of Metabolite and Macromolecule Signals for 1 H-Mrsi Using Learned Nonlinear Models," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2020-April, pp. 1725–1728, 2020.
- [64] F. Lam, Y. Li, and X. Peng, "Constrained Magnetic Resonance Spectroscopic Imaging by Learning Nonlinear Low-Dimensional Models," *IEEE Trans. Med. Imaging*, vol. 39, no. 3, pp. 545–555, Mar. 2020.
- [65] S. S. Gurbani *et al.*, "A convolutional neural network to filter artifacts in spectroscopic MRI," *Magn. Reson. Med.*, vol. 80, no. 5, pp. 1765–1775, 2018.
- [66] Y. Hiltunen, J. Kaartinen, J. Pulkkinen, A. M. Häkkinen, N. Lundbom, and R. A. Kauppinen, "Quantification of human brain metabolites from in vivo 1H NMR magnitude spectra using automated artificial neural network analysis," *J. Magn. Reson.*, vol. 154, no. 1, pp. 1–5, 2002.
- [67] N. Hatami, M. Sdika, and H. Ratiney, "Magnetic resonance spectroscopy quantification using deep learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11070 LNCS, pp. 467–475, 2018.
- [68] H. H. Lee and H. Kim, "Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain," *Magn. Reson. Med.*, vol. 82, no. 1, pp. 33–48, 2019.
- [69] M. Chandler, C. Jenkins, S. M. Shermer, and F. C. Langbein, "MRSNet: Metabolite quantification from edited magnetic resonance spectra with convolutional neural networks," *arXiv*, pp. 1–12, 2019.
- [70] C. F. Higham and D. J. Higham, "Deep learning: An introduction for applied

- mathematicians,” *SIAM Rev.*, vol. 61, no. 4, pp. 860–891, 2019.
- [71] H. Bhat, B. R. Sajja, and P. A. Narayana, “Fast quantification of proton magnetic resonance spectroscopic imaging with artificial neural networks,” *J. Magn. Reson.*, vol. 183, no. 1, pp. 110–122, 2006.
- [72] S. S. Gurbani, S. Sheriff, A. A. Maudsley, H. Shim, and L. A. D. Cooper, “Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting,” *Magn. Reson. Med.*, vol. 81, no. 5, pp. 3346–3357, 2019.
- [73] A. Shamaei, “Deep Learning For Magnetic Resonance Spectroscopy Quantification: A Time-Frequency Analysis Approach,” in *Proceedings II of the 26st Conference STUDENT EEICT 2020: Selected papers [online]*, 2020, pp. 131–135.
- [74] A. Shamaei, J. Starčuková, and Z. S. Zenon Starčuk, “A wavelet scattering convolutional network for magnetic resonance spectroscopy signal quantitation,” *BIOSIGNALS 2021 - 14th Int. Conf. Bio-Inspired Syst. Signal Process. Part 14th Int. Jt. Conf. Biomed. Eng. Syst. Technol. BIOSTEC 2021*, pp. 268–275, 2021.
- [75] R. Rizzo *et al.*, “Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias,” 2022.
- [76] A. Shamaei, J. Starcukova, and Z. Starcuk, “Physics-informed deep learning approach to quantification of human brain metabolites from magnetic resonance spectroscopy data,” *Comput. Biol. Med.*, vol. 158, p. 106837, May 2023.
- [77] A. Shamaei, J. Starčuková, and Z. S. Zenon Starčuk, “A wavelet scattering convolutional network for magnetic resonance spectroscopy signal quantitation,” in *BIOSIGNALS 2021 - 14th International Conference on Bio-Inspired Systems and Signal Processing; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021*, 2021, pp. 268–275.
- [78] R. Rizzo, M. Dziadosz, S. P. Kyathanahally, A. Shamaei, and R. Kreis, “Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias,” *Magn. Reson. Med.*, 2022.
- [79] W. T. Clarke *et al.*, “NIfTI-MRS: A standard data format for magnetic resonance spectroscopy,” *Magn. Reson. Med.*, vol. 88, no. 6, pp. 2358–2370, Dec. 2022.

## SYMBOLS AND ABBREVIATIONS

### Abbreviations:

FEEC	Faculty of Electrical Engineering and Communications
BUT	Brno University of Technology
MR	Magnetic Resonance
MRS	Magnetic Resonance Spectroscopy
NMR	Nuclear Magnetic Resonance
RF	Radio Frequency
FID	Free Induction Decay
ppm	parts per million
ADC	Analog-to-Digital Converter
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
PET	Positron Emission Tomography
DL	Deep Learning
FPC	Frequency and phase correction
MM	Macromolecules
ECG	Electrocardiogram
EMG	Electromyography
AI	Artificial Intelligence
MLP	Multilayer Perceptron
CNN	Convolutional Neural Network
FT	Fourier Transform

### Symbols:

$\mathbf{B}_0$	static external magnetic field
$\gamma$	gyromagnetic ratio
$M$	magnetic moments
$h$	Planck constant
$k$	Boltzmann's constant
$T$	temperature
$M_0$	net longitudinal magnetization
$\mathbf{B}_1$	RF pulse
$t_p$	duration of an RF pulse
$\sigma'$	shielding constant
TE	echo time
TR	repetition time
$T_1$	spin-lattice relaxation time



$T_2$	spin-spin relaxation time
$M_{xy0}$	net transversal magnetization
$T_2^*$	effective transverse relaxation time
$S_M$	estimated amplitude of interested compound
$SR$	estimated amplitude of the reference compound
$CF$	correction factor

# Appendix A - Author's Publications List

## IF Journal Publications

1. Shamaei, A, Starcukova, J, Pavlova, I, Starcuk, Z. Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals. *Magn Reson Med.* 2023; 89: 1221– 1236. doi:10.1002/mrm.29498.
2. Shamaei, A, Starcukova, J, Starcuk, Z. Physics-informed deep learning approach to quantification of human brain metabolites from magnetic resonance spectroscopy data. *Computers in Biology and Medicine.* 2023; 158: 106837. doi:10.1016/j.combiomed.2023.106837.
3. Clarke, W, Mikkelsen, M, Oeltzschner, G, Bell T.K., Shamaei, A, Soher, B.J., Emir, U, Wilson, W. A standard data format for magnetic resonance spectroscopy. *Magnetic Resonance in Medicine* 2022; 1- 13. doi:10.1002/mrm.29418.
4. Rizzo, R, Dziadosz, M, Kyathanahally, SP, Shamaei, A, Kreis, R. Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias. *Magn Reson Med.* 2022; 1- 21. doi:10.1002/mrm.29561.

## Peer-reviewed extended papers at international conferences (in WOS):






5. Shamaei, A, Starcukova, J, Starcuk, Z. A wavelet scattering convolutional network for magnetic resonance spectroscopy signal quantitation. *BIOSIGNALS 2021 - 14th Int. Conf. Bio-Inspired Syst. Signal Process. Part 14th Int. Jt. Conf. Biomed. Eng. Syst. Technol. BIOSTEC 2021*, pp. 268–275, 2021.
6. Shamaei, A. Deep Learning For Magnetic Resonance Spectroscopy Quantification: A Time-Frequency Analysis Approach. in *Proceedings II of the 26st Conference STUDENT EEICT 2020: Selected papers [online]*, 2020, pp. 131–135.

## Conference abstracts at international conferences:

7. Shamaei, A.M., Starčuková J., Radim Kořínek, and Starcuk Jr. Z., Magnetic Resonance Spectroscopic Imaging Data Denoising by Manifold Learning: An Unsupervised Deep Learning Approach, Poster presentation delivered in person at ISMRMB 2022, May, 2022. London., UK.
8. Shamaei, A.M., Starčuková J. and Starcuk Jr. Z., Frequency and Phase Shift Correction of MR Spectra Using Deep Learning in Time Domain, Poster presentation delivered virtually at ESMRMB 2021, October, 2021. Online.

9. Clarke, W., Bell, T., Emir, U., Mikkelsen, M., Oeltzschner, G., Rowland, B., Shamaei, A.M, Soher, B, Tapper, S., and Wilson, M, NIfTI MRS: A standard format for spectroscopic data, Poster presentation delivered in person at ISMRMB 2021, May, 2021. Online.
10. Amir M Shamaei, Rudy Rizzo, Physics-Informed Deep Learning Approach to Quantifying Magnetic Resonance Spectroscopy Data with Simultaneous Uncertainty Estimation, Power pitch (oral and poster) presentation will be delivered in person at ISMRMB 2023, June, 2023. Toronto, Canada. Accepted
11. Amir M Shamaei, Jana Starcukova, Jedrek Burakiewicz, Zenon Starcuk Jr, Water removal in MR spectroscopic imaging with Casorati Singular Value Decomposition, a Poster presentation will be delivered in person at ISMRMB 2023, June, 2023. Toronto, Canada. Accepted
12. Shamaei, A.M., Starčuková J., and Starcuk Jr. Z., EigenMRS: A computationally cheap data-driven approach to MR spectroscopic imaging denoising, Poster presentation will be delivered in person at ISMRMB 2023, June, 2023. Toronto, Canada. Accepted

# Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals

Amirmohammad Shamaei<sup>1,2</sup>   | Jana Starcukova<sup>1</sup>  | Iveta Pavlova<sup>1</sup>  |  
Zenon Starcuk Jr.<sup>1</sup> 

<sup>1</sup>Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

<sup>2</sup>Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic

## Correspondence

Amirmohammad 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](mailto:amirshamaei@isibrno.cz)

## Funding information

European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie, Grant/Award Number: 813120; Czech Academy of Sciences, Institute of Scientific Instruments, RVO, Grant/Award Number: 68081731

**Purpose:** A supervised deep learning (DL) approach for frequency and phase correction (FPC) of MRS data recently showed encouraging results, but obtaining transients with labels for supervised learning is challenging. This work investigates the feasibility and efficiency of unsupervised deep learning-based FPC.

**Methods:** Two novel deep learning-based FPC methods (deep learning-based Cr referencing and deep learning-based spectral registration), which use a priori physics domain knowledge, are presented. The proposed networks were trained, validated, and evaluated using simulated, phantom, and publicly accessible in vivo MEGA-edited MRS data. The performance of our proposed FPC methods was compared with other generally used FPC methods, in terms of precision and time efficiency. A new measure was proposed in this study to evaluate the FPC method performance. The ability of each of our methods to carry out FPC at varying SNR levels was evaluated. A Monte Carlo study was carried out to investigate the performance of our proposed methods.

**Results:** The validation using low-SNR manipulated simulated data demonstrated that the proposed methods could perform FPC comparably with other methods. The evaluation showed that the deep learning-based spectral registration over a limited frequency range method achieved the highest performance in phantom data. The applicability of the proposed method for FPC of GABA-edited in vivo MRS data was demonstrated. Our proposed networks have the potential to reduce computation time significantly.

**Conclusions:** The proposed physics-informed deep neural networks trained in an unsupervised manner with complex data can offer efficient FPC of large MRS data in a shorter time.

## KEYWORDS

deep learning, edited MRS, frequency correction, MR spectroscopy, phase correction

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 Ústav přístrojové techniky AV ČR, v. v. i. *Magnetic Resonance in Medicine* published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

## 1 | INTRODUCTION

In MRS, typically more transients are acquired and averaged to increase the low SNR.<sup>1</sup> However, individual transients might have different frequency and phase shifts because of hardware imperfections, physiologic processes, or other instabilities.<sup>2,3</sup> Averaging transients without frequency and phase correction (FPC) would result in line-broadening and lineshape imperfection of the combined MRS signal. Thus, FPC should be performed for each transient before averaging. It is even more critical to use accurate FPC while using spectral-edited MRS<sup>1</sup> methods to prevent artifacts caused by subtraction. Thus, FPC is a consensus-recommended and effective step<sup>4</sup> in MRS signal processing.

Several FPC approaches have been developed.<sup>3,5–10</sup> The FPC methods can be classified into absolute and relative methods.<sup>7</sup> Absolute approaches correct each individual transient absolutely, whereas relative methods align the transients to a reference signal. A commonly used approach for FPC is to use the water peak and read the phase and frequency from it.<sup>6,8,11</sup> Another approach is to fit a certain metabolite peak to a model<sup>12</sup> and then estimate the frequency and phase shifts from the model. One approach that has been proposed and evolved recently is spectral registration (SR).<sup>3,5,9,10</sup> Spectral registration fits each signal to a reference signal in the time domain through the adjustment of frequency and phase terms. Even though SR works very well for small shifts, it struggles with larger shifts and signals with low SNR. Modified versions of SR successfully addressed some of the mentioned problems.<sup>5,9,10</sup> Most of these approaches are time-consuming for large data sets, such as high-resolution MRSI data sets, which may have thousands of spectra.

The recent success of deep learning (DL), one of the latest machine learning (ML) approaches, in a wide range of tasks, including the MR field,<sup>13,14</sup> suggests that it could also handle FPC. Recently, DL-based solutions have been proposed for metabolite quantification in the frequency domain,<sup>15,16</sup> detecting and removing ghosting artifacts,<sup>17</sup> FID reconstruction,<sup>18</sup> automatic peak picking,<sup>19</sup> enhancement of MRSI spatial resolution,<sup>20</sup> and identifying and filtering out poor-quality spectra.<sup>21</sup> It has been shown that DL can also be used for FPC<sup>7,22</sup> and could speed up FPC once it has been successfully trained. This approach, using two separate networks in sequence to estimate frequency and phase, showed encouraging results. The first network was trained for frequency-shift estimation using the magnitude of frequency-shifted and phase-shifted spectrum as the input and the known frequency shift as the output. Subsequently, the second network was trained for phase

shift estimation using real parts of the frequency-corrected spectrum as the input and phase shift as the output. In this approach, any error in the first step (frequency correction) may bias the phase shift estimation. Training two networks is a computationally expensive task. Moreover, the networks were trained in a supervised manner using simulated data. Any discrepancy between the *in vivo* and the simulated spectra may result in errors in frequency and phase shift estimation. The true output values are unknown in MRS data, and obtaining hundreds of spectra with labeled frequency and phase shifts is almost infeasible. Unsupervised learning may eliminate the drawbacks of supervised learning.

Frequency and phase correction is traditionally described by adjusting two parameters (frequency and phase). Therefore, it is natural to expect that the variability of all the signals in the set acquired for SNR improvement should have a very low-dimensional representation. One of the methods for nonlinear dimensionality reduction is manifold learning, which assumes that the available high-dimensional data vectors are embedded in low-dimensional manifolds.<sup>23,24</sup> These low-dimensional manifolds can be learned by deep autoencoders (DAEs),<sup>25</sup> which automate feature extraction by merging all relevant data into a cohesive framework. A DAE with a common architecture<sup>13,25</sup> is not able to learn to estimate the frequency and the phase shift of a transient, as the features in a low-dimensional space might not be readily interpretable. Therefore, a DAE can be redesigned to have two functions: a function for nonlinear mapping between the input and certain features (frequency and phase shifts) in a two-dimensional space, and another function, for reconstructing the input from those features. Accordingly, we designed a DAE network that can learn in an unsupervised manner to estimate the frequency and phase shifts of MRS data. The proposed method takes advantage of the parametric analytical approach and embeds it into the DAE to estimate the frequency and phase shifts of a transient.

The proposed network was trained and validated using a simulated data set in which ground-truth knowledge was available and evaluated using phantom and *in vivo* MEGA-edited MRS data obtained from the publicly accessible Big GABA repository.<sup>26,27</sup> The FPC performance of our proposed network was compared with the commonly used FPC methods, namely, SR,<sup>3</sup> spectral registration over a limited frequency range (SRF),<sup>3</sup> frequency domain correlation,<sup>5</sup> frequency domain correlation over a limited frequency range,<sup>5</sup> creatine referencing (CrR),<sup>12,28</sup> as well as supervised deep-learning approaches<sup>7,22</sup> in terms of precision and time efficiency.

## 2 | METHODS

### 2.1 | Data normalization

In contrast to the application of DL in machine vision or speech recognition, where the input data can be normalized by a nonlinear transform, MRS signals must be normalized by a linear transform. In this study, each complex signal  $S(t)$  was rescaled as

$$S(t)_{\text{normalized}} = \frac{S(t)}{\max_{\tau} (|S(\tau)|)}. \quad (1)$$

A similar approach could be dividing the signal by the absolute value of the first point of the signal, but it would be less generally applicable because the initial point can be influenced by the filtering processes in the receive chain,<sup>29</sup> or the maximum may occur later in an echo or with coupled resonances.

### 2.2 | Data augmentation

It is known that the sufficient size and diversity of data are important factors for the effectiveness of most DL models.<sup>13</sup> However, having rich and sufficient data sets is rare<sup>30</sup> in the field of MRS and MRSI. Data augmentation is a viable option, which simulates credible data by minor alterations to data in a small existing data set. Data augmentation used in computer vision applications to reduce the generalization error of models<sup>13</sup> applies to flips, translations, and rotations,<sup>13</sup> which would be meaningless in spectroscopy. To generate credibly varied FID signals, we chose a set of physics-informed alterations, simulating the practical data variability:

1. Frequency shift
2. Phase shift
3. Apodization (line broadening)
4. Amplitude change
5. Adding noise
6. Adding a nuisance peak (residual water and lipids)

### 2.3 | Data sets

#### 2.3.1 | Simulated data set

A simulated data set was used in *in silico* ground-truth information for evaluating the performance of our proposed networks and for comparison with the commonly used FPC methods. The simulated data set was obtained by alternating a single MR signal acquired from a rat brain as described subsequently.

A single-voxel spectroscopy (SVS) MR *in vivo* signal was acquired from a rat's right hippocampus (256

transients, voxel size =  $1.5 \times 1.5 \times 4 \text{ mm}^3$ ) in a 9.4T small animal MR system (Bruker BioSpin MRI, Ettlingen, Germany) using a PRESS sequence (spectral width = 4400 Hz, 2048 points, TE = 16.5 ms, TR = 2500 ms) with water and outer-volume suppression by VAPOR.<sup>31</sup> The signal (further referred to as the basis signal  $S_{\text{basis}}(t)$ ) was created after transients were corrected for  $B_0$  instability due to eddy currents as well as  $B_0$  drift, and averaged using Bruker proprietary software, Paravision.

All experiments were approved by the Czech Governmental Animal Care Committee, in compliance with Czech Animal Protection.

The simulated data set, containing 24 000 artificial signals, was generated from  $S_{\text{basis}}(t)$  by an augmentation procedure. The basis signal was multiplied by factors drawn from a normal distribution with a mean of 1 and a SD of 0.1. Then a set of lipid nuisance peak and a set of unstable residual water nuisance peak, generated using Equation 2 (parameters are listed in Supporting Information Table S1), were added to signals, randomly and independently:

$$S(t)_{\text{nuisance peak}} = A_a e^{-d_a t} e^{-i(2\pi f_a t + \varphi_a)}, \quad (2)$$

where  $i = \sqrt{-1}$ ;  $t$  is a vector of time points; and  $A_a$ ,  $d_a$ ,  $f_a$ , and  $\varphi_a$  are the amplitude, the damping factor, the precession frequency, and the phases of the nuisance peak, respectively. Signals containing the lipid peak, the unstable residual water peak, and both peaks were labeled as LC, UW, and UW & LC, respectively.

All artificial signals were further apodized by normally distributed random dampings corresponding to Lorentzian linewidths with a mean of 2 Hz and a SD of 0.2 Hz. Then, uniformly distributed artificial frequency and phase offsets in the range of  $-20$  to  $20$  Hz and  $-90^\circ$  to  $90^\circ$ , respectively, were applied to signals. Before the normalizing step, the SNR of signals (time origin magnitude to noise SD) was set in the range of approximately 9 to 27 by introducing random complex Gaussian white noise. Signals in the simulated data set were shuffled randomly, and 90% of the data set was allocated to the training subset: 9% for the validation subset, and the remaining 1% for the test subset.

#### 2.3.2 | Phantom data set

Phantom data were used to assess the performance of our methods in the absence of a large number of training data measured during temperature-dependent changes in the  $B_0$  field.

An SVS MR signal was acquired in a phantom of known metabolite concentrations (N-acetylaspartate



17.5 mmol/L, glutamate 26.2 mmol/L, myo-Inositol 17.5 mmol/L, creatine [CR] 12.7 mmol/L, taurine 10.1 mmol/L; 2048 transients, voxel size =  $3 \times 3 \times 3 \text{ mm}^3$ ) in a 9.4T small animal MR system (Bruker BioSpin MRI, Ettlingen, Germany), while the temperature of the phantom was altered between 35°C and 40°C and the frequency adjustment of the scanner was switched off, using PRESS sequence (spectral width = 4400 Hz, 2048 points, TE = 16.5 ms, TR = 2500 ms) with water and outer-volume suppression by VAPOR.<sup>31</sup>

The phantom data set, containing 24 000 artificial signals, was generated as follows: We selected 400 (out of 2048) of the acquired transients as basis signals, randomly; then, a subset of 60 signals was generated from each basis signal by the same augmentation procedure as used for the simulated data set except that the basis signal was not multiplied by factors, the nuisance peaks were not added, and the SNR of signals was set in the range of about 7 to 70. Finally, all subsets were stacked together to create the final training data set. The rest of the 1648 transients were used as an unseen test subset.

### 2.3.3 | Big GABA data set

Data from the public repository Big GABA<sup>26,27</sup> were used to demonstrate the applicability of the proposed method on FPC of edited in vivo signals.

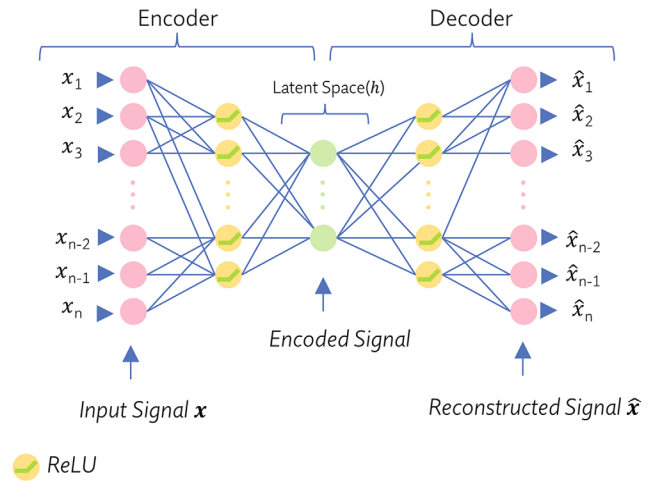
We selected 48 GABA-edited MEGA-PRESS subsets (subjects) acquired on Siemens scanners from four different sites (S1, S5, S6, and S8 [all that we found public and readable]; 3T field strength; spectral width = 4000 Hz, 4096 points, TE = 68 ms; ON/OFF editing pulses = 1.9/7.46 ppm; editing pulse duration = 15 ms, TR = 2000 ms; 320 averages;  $30 \times 30 \times 30 \text{ mm}^3$ ; and medial parietal lobe voxel) because Siemens data<sup>3</sup> had greater median within-participant SD of estimated phase offsets and larger variance of lipid contamination than data from Philips and GE, and relatively high average frequency offsets, which are undesirable conditions for conventional FPC methods.<sup>5</sup>

We allocated 40 of 48 selected in vivo subsets (15 360 transients) to the training subset (12 800 transients), and the rest of the subsets (2560 transients) were used as an unseen test subset.

## 2.4 | Deep model

### 2.4.1 | The DAE proposed for deep learning-based peak referencing

The DAE is a type of deep artificial neural network that is created to learn the coding of data in an unsupervised



**FIGURE 1** Illustration of a common deep autoencoder (DAE) architecture with multiple nonlinear hidden layers composed of rectified linear units (ReLU) and fully connected (FC) layers to elicit nonlinear features of the data

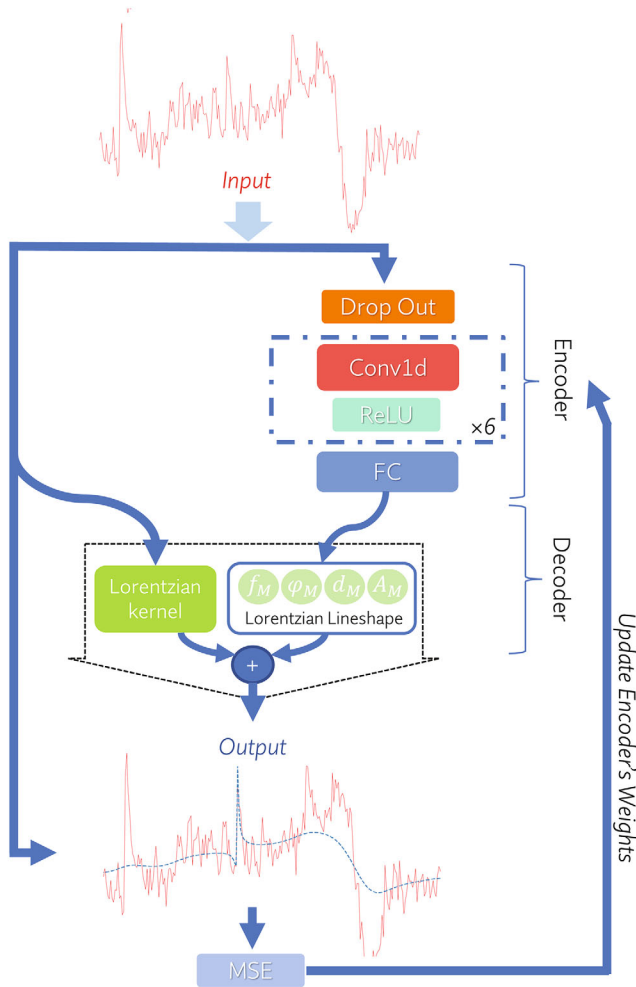
manner. The fundamental underlying concept of autoencoders is to use the input data as the target (ie, attempting to reconstruct the input data in the output layer).<sup>13</sup> Typically, a DAE consists of two parts: an encoder and a decoder.

Figure 1 illustrates the most common architecture of a DAE. The encoder function  $\mathbf{h} = f(\mathbf{x})$  maps the  $n$ -dimensional input vector ( $\mathbf{x} \in R^n$ ) to the  $n'$ -dimensional latent vector ( $\mathbf{h} \in R^{n'}$ ), while the decoder function  $\hat{\mathbf{x}} = g(\mathbf{h})$  aims to reconstruct the  $n$ -dimensional output vector ( $\hat{\mathbf{x}} \in R^n$ ) from the latent space representation. The mathematical expression of a DAE can be written as follows:

$$\hat{\mathbf{x}} = g(f(\mathbf{x}; \theta_e); \theta_d), \quad (3)$$

where  $\theta_e$  and  $\theta_d$  are the parameters set of encoder and decoder layers, respectively.

For FPC, the latent space representation must be interpretable parameters, such as frequency and phase shifts. Frequency and phase shifts can be estimated by fitting a certain metabolite peak to a model, such as a Lorentzian lineshape. Then, the frequency and phase shifts can be read from the model. To this end, we proposed a convolutional encoder/model-decoder<sup>15</sup> architecture. Our proposed DAE has a conventional encoder consisting of a pipeline of a dropout layer,<sup>32</sup> convolutional layers,<sup>33,34</sup> fully connected layers,<sup>13,34</sup> and rectified linear unit layers,<sup>34</sup> which encode a complex input signal into a latent space, and a decoder that reconstructs a Lorentzian lineshape of a certain peak in the input signal using the latent space parameters. The proposed DAE architecture is depicted in Figure 2. Because it has been shown that using complex-valued data improves the robustness and



**FIGURE 2** Illustration of the proposed convolutional encoder/model-decoder for the deep learning-based Creatine referencing (dCrR) method. The network's input is a complex signal in the time domain, which is fed to the encoder. The encoder consists of a dropout layer, six convolutional blocks, and an FC layer (see details in Supporting Information Table S2). A convolutional block (dashed square) consists of a one-dimensional convolution (Conv1d) layer followed by a ReLU layer. The model-decoder of the dCrR (Equation 4) reconstructs the output signal. The DAE was trained to encode the input vector in the time domain into parameters that can be used to reconstruct the output vector in the time domain. The proposed network is trained by minimizing the mean square error (MSE) between the input and the output. The input and output signals are depicted in the frequency domain for the sake of easier understanding. Abbreviations:  $A_M$ , amplitude;  $d_M$ , damping factor;  $f_M$ , resonance frequency;  $\varphi_M$ , zero-order phase of the Lorentzian lineshape

efficiency of fitting MRS data,<sup>35</sup> the input and output of the proposed DAE were set to be complex signals in the time domain.

Time-domain fitting of a single Lorentzian lineshape to a signal with several peaks using our proposed network is a challenging task in which the optimization algorithm

aims to increase the linewidth to decrease the error. Previous studies<sup>3,12,28</sup> addressed these problems by fitting a signal in the frequency domain over a limited range and including a linear baseline in their model. We found that adding a rough estimate of a baseline, obtained by apodizing the input signal with Lorentzian kernel with a large linewidth, into the reconstruction function improves our fitting. Hence, the decoder part combines a mathematical model (Lorentzian lineshape) and the input signal,  $x$ , apodized with the Lorentzian kernel. The mathematical expression of the decoder can be written as follows:

$$g(x; A_M, d_M, f_M, \varphi_M, t) = A_M e^{-d_M t} e^{-i(2\pi f_M t + \varphi_M)} + x(t) e^{-d_L t}, \quad (4)$$

where  $A_M$ ,  $d_M$ ,  $f_M$ , and  $\varphi_M$  are the amplitude, the damping factor, the resonance frequency, and the zero-order phase of the selected Lorentzian lineshape, respectively, and  $d_L$  is the linewidth of the Lorentzian kernel. Experimentally,  $d_L$  was set to 500 Hz in this study.

Training our proposed network is an unsupervised learning task that does not require ground-truth frequency and phase shifts and can be done by minimizing the differences between the original input and the consequent reconstruction. In each iteration step of training, the parameters of the encoders are adjusted according to the gradient of the loss function with respect to the given parameters of the Lorentzian lineshape ( $A_M$ ,  $d_M$ ,  $f_M$ , and  $\varphi_M$ ).

In this study, Cr peak at 3.027 ppm was selected to be fitted by a Lorentzian line shape in the decoder. The FIDs were truncated to the initial 512 points for limiting the contribution of noise, which typically predominates in the later part of FIDs. Then, the truncated FIDs were used as inputs to the network. After training, the encoder of the proposed DAE (deep learning-based creatine referencing [dCrR]) was detached from the network and used to estimate the frequency and phase of the Cr peak in a test transient. Then the estimates were used for frequency and phase correction of the test transient. The pipeline of FPC of GABA-edited MEGA-PRESS transients is provided in Supporting Information Figure S4.

#### 2.4.2 | The DAE proposed for deep learning-based SR

With a simple modification, the proposed approach could estimate the relative frequency and phase shifts by fitting each signal to a reference signal and be applicable to various MRS experiments. In other words, the SR method could be used in our proposed encoder/model-decoder network. This modification is referred to as deep learning-based spectral registration (dSR). For dSR, the



mathematical expression of the decoder can be written as follows:

$$g(f_M, \varphi_M, t) = R(t)e^{-i(2\pi f_M t + \varphi_M)}, \quad (5)$$

where  $f_M$  and  $\varphi_M$  are the frequency and the phase shifts of the input signal with respect to the reference scan  $R(t)$ . Thus, in the dSR, the encoder estimates two parameters (relative phase and frequency shifts) instead of estimating the four parameters of a lineshape. The proposed DAE architecture for dSR is depicted in Supporting Information Figure S5. We proposed a new ML-based algorithm for finding the reference signal in which the k-means algorithm<sup>36</sup> was used to cluster magnitude-valued signals in the training set. The number of clusters was set to two: one for samples with nuisance peaks and another for samples without nuisance peaks. k-Means algorithm is an unsupervised method; thus, the clusters are not identified. Based on prior testing, samples containing nuisance peaks have a higher intensity than samples without nuisance peaks in the initial points of their mean of the samples in the cluster. Therefore, the first 10 points of each mean of samples in the cluster were averaged, and the cluster with a lower average was identified as the cluster containing samples without nuisance peaks; then, the signal with the highest SNR from this cluster was selected as the reference scan (see details in Supporting Information Text S2 and Figure S6).

### 2.4.3 | Training the proposed DAEs over a limited frequency range

The MRS signals may have unstable frequency components. Because all frequency components are present at all time points, unstable frequency components may bring errors into our proposed method. To avert this situation, the proposed architectures (dCrR and dSR) were trained over a limited frequency range (2.5 to 3.5 ppm) (referred to as dCrRF [deep learning-based creatine referencing over a limited frequency range] and dSRF [deep learning-based spectral registration over a limited frequency range]). Note that the input of the encoder and the output of the model-decoder were still in the time domain, and the discrepancy between fast Fourier transformation of the input and output was calculated over a limited frequency range.

## 2.5 | Implementation details and training

All steps were run on a computer with a dual EPYC 7742 ( $2 \times 64$  cores) processor and one graphics processing

unit (NVIDIA A100 40 GB). Moreover, all steps were run on Google Colaboratory (free-to-use hardware).<sup>37</sup> The DAE was implemented in Python with the help of the Pytorch lightning interface.<sup>38,39</sup> The architecture of the network and training parameters were optimized using the Bayesian Optimization HyperBand algorithm<sup>40</sup> with the help of the Tune framework.<sup>41</sup> The details of the optimization are given in Supporting Information Text S1. All training was performed using the mean-squared error loss and an Adam optimizer<sup>42</sup> with a batch size of 16, a learning rate of  $4 \times 10^{-5}$ , and 150 epochs. The training progress for the simulated data set is provided in Supporting Information Figures S2 and S3.

An early-stopping strategy<sup>38</sup> was performed by monitoring the average error of the validation subset at the end of every epoch and stopping the training when no improvement was observed in 10 epochs. The SR and the SRF methods were tested using the FID-A toolbox,<sup>43</sup> the CrR and the CrRF methods with the Gannet toolbox,<sup>28</sup> the frequency domain correlation and frequency domain correlation over a limited frequency range methods,<sup>5</sup> and previous studies (Tapper et al<sup>7</sup> and Ma et al<sup>22</sup>) with our in-house code.

All proposed methods (dCrR, dCrRF, dSR, and dSRF) were trained, validated, and tested using the simulated and the phantom data set, and the dCrR method was trained and tested using the big GABA data sets.

## 2.6 | Statistics and quality evaluation

### 2.6.1 | Performance analysis

For the simulated data set, in which the true shifts from the basis signal were known, the error was defined as the difference between the estimated and the true shifts. The accuracy and precision of an FPC method were established as the average and the SD of the error, respectively. In addition, the performance of the dCrR method trained with the simulated data set was investigated beyond the trained range of frequency and phase ( $-40$  to  $40$  Hz and  $-180^\circ$  to  $180^\circ$ , respectively).

For the phantom and the Big GABA data set, in which true shifts were not known, the quality of alignment was measured by comparing the similarity index (SI), which is the sum of all elements of a similarity matrix. Each element of the similarity matrix is the normalized scalar product (the equation as implemented is provided in Supporting Information Equation S1) for each pair of spectra (fast Fourier transformation of a FID) over a limited frequency range (from 2.5 to 3.5 ppm). Note that higher SI denotes superior performance.

## 2.6.2 | Performance against noise

The stability of the dCrR, dSR, and SR method against noise was evaluated. A set of transients was generated using the following procedure. First, a frequency shift of 5 Hz and a phase shift of  $45^\circ$  were added to  $S_{\text{basis}}(t)$ . Second, 20 realizations of a random complex Gaussian noise with a linearly increasing SD were introduced to the shifted signal such that SNR was in the range of approximately 8 to 110. The frequency and phase shifts of the generated set were estimated using the networks trained with the simulated data set. The FPC performance was evaluated as a function of SNR.

## 2.6.3 | Monte Carlo analysis

Monte Carlo studies were carried out to investigate the performance of the dCrR, dSR, and SR methods. A set

of transients was generated using the following procedure. First, a frequency shift of 5 Hz and a phase shift of  $45^\circ$  were added to  $S_{\text{basis}}(t)$ . Second, 256 realizations of a random complex Gaussian noise with the same SD were introduced to the shifted signal such that SNR was approximately 15. The frequency and phase shifts of the generated set were estimated using the networks trained with the simulated data set. The FPC performance of dCrR, dSR, and SR methods was compared.

## 3 | RESULTS

The training and processing time of each proposed network for the simulated data set is listed in Table 1. Approximately, frequency and phase estimation of one transient requires approximately 3 ms. The evaluation of the performance of different methods (precision, SI, linewidth, the processing time of FPC, and training time

**TABLE 1** Comparison of our proposed method with existing commonly used FPC methods for the simulated data set

		dSR	dSRF	SR	SRF	dCr	dCrF	Cr	Corr	CorrF	Tapper et al	Ma et al
Precision of frequency estimation (Hz)	All samples	1.02	0.73	1.57	11.46	0.94	0.95	7.92	10.70	11.33	<b>0.71</b>	0.99
	Samples without nuisance peaks	1.00	0.69	1.01	11.02	0.93	0.90	9.62	9.98	9.89	<b>0.56</b>	0.85
	Samples with nuisance peaks	0.97	<b>0.73</b>	1.57	11.68	0.94	0.97	6.51	11.07	11.96	0.78	1.06
Precision of phase estimation ( $^\circ$ )	All samples	5.87	4.89	7.44	15.92	6.62	6.79	32.34	16.87	17.46	6.78	<b>4.29</b>
	Samples without nuisance peaks	4.16	3.34	<b>2.61</b>	16.75	6.89	7.00	38.80	16.62	16.46	4.64	4.24
	Samples with nuisance peaks	5.75	5.52	8.50	15.34	6.35	6.52	26.68	17.02	18.07	5.89	<b>4.32</b>
SI	All samples	0.44	<b>0.44</b>	0.44	0.32	0.42	0.42	0.36	0.34	0.32	0.44	0.44
Cr <sub>linewidth</sub> (Hz)	Averaged signal	10.38	10.22	10.77	15.34	<b>10.14</b>	10.17	10.71	19.14	20.55	10.45	10.53
Training time (min) CPU	All samples	<b>127</b>	138	—	—	328	354	—	—	—	—	—
Training time (min) GPU	All samples	<b>2.26</b>	7.8	—	—	8.4	7.3	—	—	—	20.8	29.64
Training time (min) Google Colaboratory	All samples	<b>4.53</b>	10.6	—	—	11.1	11.2	—	—	—	—	—
Processing time (ms) CPU	Per signal	<b>3</b>	<b>3</b>	57	71	<b>3</b>	<b>3</b>	53	33	29	15	15
Processing time (ms) GPU	Per signal	<b>0.1</b>	<b>0.1</b>	—	—	0.2	0.2	—	—	—	0.6	0.6

Note: Bold text indicates the best performance in each metric. The SI of the test subset without FPC was 0.14. Precision: SD of the difference between the estimated and the true shift.

Abbreviations: Corr, frequency domain correlation; CorrF, frequency domain correlation over a limited frequency range; CPU, central processing unit; Cr<sub>linewidth</sub>, linewidth of Cr peak at 3 ppm; CrR, creatine referencing; GPU, graphics processing unit; SI, similarity index; SRF, SR over a limited frequency range.

**TABLE 2** Comparison of our proposed method with existing commonly used FPC methods for the phantom data set

	SI	Cr <sub>linewidth</sub> (Hz)
dCrR	0.66	6.02
dCrRF	0.66	<b>6.00</b>
dSR	0.60	6.80
dSRF	<b>0.67</b>	6.03
SR	0.58	6.75
SRF	0.60	6.46
CrR	0.58	7.03

Note: Bold text indicates the best performance in each metric. The SI of the test subset without FPC was 0.50.

for DAE) for the test signals of the simulated data set and of the phantom are summarized in Tables 1 and 2, respectively.

The Bland–Altman plots in Figure 3A,B show the accuracy and biases in the estimation of frequency and phase shifts using dCrR, dCrRF, and dCr. The highest precision (0.94 Hz and 6.62°) among the absolute FPC methods was achieved with the proposed dCrR. Although dCrRF performed similarly to dCrR, CrR showed the lowest performance (7.92 Hz and 32.34°). All of the proposed methods performed well in spectra with and without the simulated nuisance peaks. The agreement, estimated by  $R^2$  value, was high for dCrR ( $R^2_{\text{frequency}} = 0.99$  and  $R^2_{\text{phase}} = 0.99$ ) and dCrRF ( $R^2_{\text{frequency}} = 0.99$  and  $R^2_{\text{phase}} = 0.99$ ) and moderate for CrR ( $R^2_{\text{frequency}} = 0.61$  and  $R^2_{\text{phase}} = 0.75$ ). Figure 3C,D illustrates the results of dCrR on unseen test signals in which offsets are beyond trained bound (−40 to 40 Hz and −180° to 180°). The network showed poor performance beyond its trained bound (precisions for frequency and phase estimation were 8.69 Hz and 60.04°, respectively). Figure 3E,F shows the spectra and the similarity matrix heatmaps obtained before and after the FPC tested, respectively. The dCrR method increased the SI in the visualized test signals from 0.14 to 0.42.

Figure 4 shows a pairwise correlational comparison of the relative FPC methods (our proposed dSR and dSRF with SR and SRF). The  $R^2$  value of each method pair and the true value of shifts are reported in the corresponding axes. The  $R^2$  indicates that there is a high degree of agreement between the frequency and phase estimations between methods except for SRF. The agreement between the estimations of methods and true values ( $R^2$ ) was high for dSR, dSRF, and SR and was low for SRF ( $R^2_{\text{frequency}} = 0.03$  and  $R^2_{\text{phase}} = 0.94$ ).

Figure 5 shows spectra from the test signals of the phantom data set and the corresponding heatmaps of similarity matrices before and after correction using the dCrR method. The dCrR method achieved the highest performance and increased the similarity among transients from 0.50 to 0.66, and decreased the linewidth of Cr peak at 3 ppm in the averaged spectrum by 1.5 Hz (Table 2). We observed that changing temperature during measurement altered the amplitude and linewidth of peaks, which resulted in a decreased similarity index. The dSR method failed to increase the SI and reduce the Cr linewidth because dSR tends to correct frequency and phase shifts of large frequency components, such as the residual water. However, dSRF overcame this issue and demonstrated superior performance by limiting the frequency range.

The link between the absolute error in estimates and the SNR is shown in Figure 6A,B using a scatter plot. The comparison of the proposed dSR and dCrR with SR methods is presented. In low SNR, dSR and SR methods outperformed dCrR in terms of the precision of phase shift estimation, but dCrR demonstrated more resilience in terms of the precision of frequency estimation.

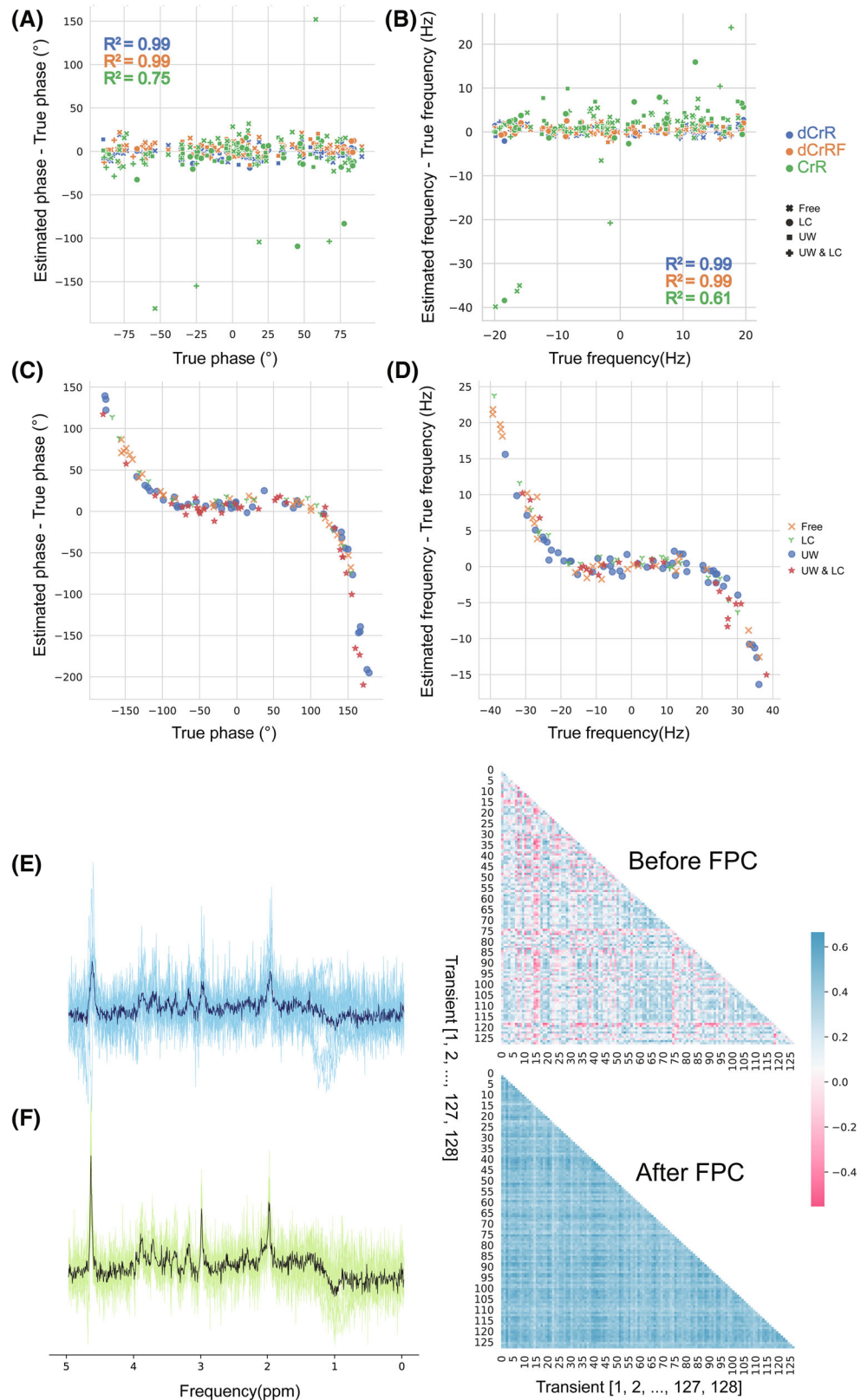
Figure 6C shows a comparison of the dSR, dCrR, and SR methods in the Monte Carlo analysis using a scatter-plot visualization of the joint distribution of frequency and phase. For the simulated data set, the mean error of dSR ( $1.36 \pm 0.69$  Hz and  $0.51 \pm 2.33^\circ$ ) and SR ( $-0.67 \pm 0.81$  and  $3.15 \pm 2.53^\circ$ ) showed similar performance, whereas dCrR performed less precise phase-shift estimation ( $4.69 \pm 0.77$  Hz and  $-41.27^\circ \pm 5.827^\circ$ ). The true values of the frequency and the phase shifts were 5 Hz and 45°.

Figure 7 illustrates an unseen test subset of a GABA-edited in vivo data set (site = 1, subject = 3; 160 edited transients; 160 unedited transients) and a heatmap of their similarity matrix before and after FPC by the dCrR method. Our method increased the SI in the visualized test subset from 0.80 and 0.82 to 0.92 for edited (ON) and unedited (OFF) spectra, respectively. In all test subsets, the SI was increased from  $0.88 \pm 0.05$  to  $0.93 \pm 0.02$ .

Figure 8 shows the comparison of the results of dCrR-based and SR-based correction of ON and OFF transients of the test subsets. On average, dCrR performed better than SR, as indicated by the mean SI (the mean SI was increased from  $0.88 \pm 0.05$  [ON:  $0.89 \pm 0.04$ , OFF:  $0.87 \pm 0.05$ ] to  $0.93 \pm 0.02$  [ON:  $0.93 \pm 0.02$ , OFF:  $0.93 \pm 0.02$ ] and  $0.90 \pm 0.04$  [ON:  $0.91 \pm 0.03$  OFF:  $0.89 \pm 0.04$ ] by dCrR-based correction and SR-based correction, respectively).



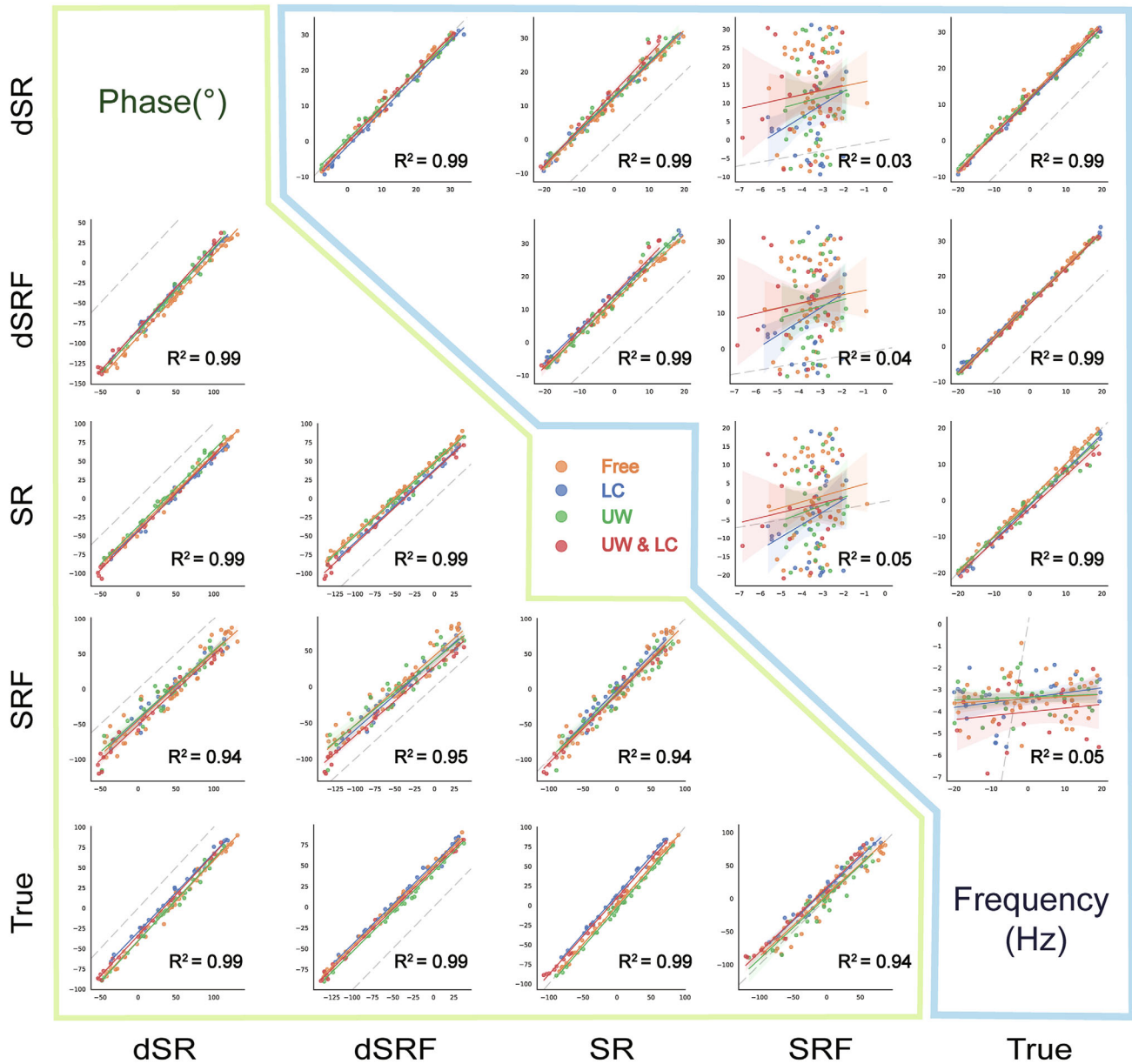
**FIGURE 3** A, B, Testing results of the simulated data set. A, B, Plots of the error in deep learning (DL)-estimated phase (A) and frequency (B) shifts against the actual shifts, respectively. The  $R^2$  values of the method types are color-coded. C, D, Testing the dCrR method beyond the trained range of phase and frequency ( $-180^\circ$  to  $180^\circ$  and  $-40$  to  $40$  Hz, respectively). Plots of the error in DL-estimated phase (C) and frequency (D) shifts against the actual shifts, respectively, using dCrR. E, Uncorrected spectra and their similarity matrix. F, The same spectra after frequency and phase correction (FPC) using dCrR and their similarity matrix. Dark blue and green spectra show the average uncorrected and corrected spectra, respectively. Abbreviations: CrR, creatine referencing; dCrRF, dCrR over a limited frequency range; Free, without any nuisance peak; LC, with lipid peak; UW, with unstable water peak



## 4 | DISCUSSION

In this study, the combination of DL and mathematical modeling was demonstrated to be able to provide FPC in

simulated, phantom, and in vivo MRS data. We compared our method with five previously published methods: SR, SRF, frequency domain correlation, frequency domain correlation over a limited frequency range, and CrR, as well

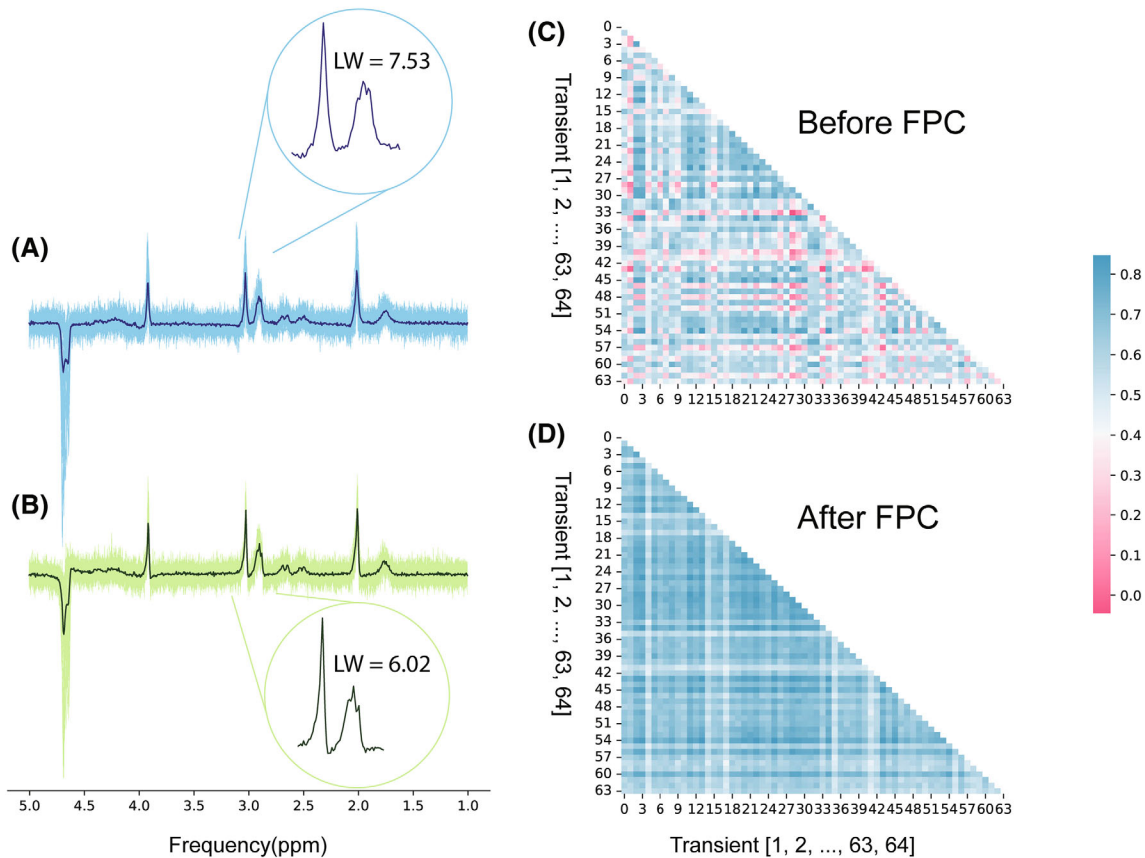


**FIGURE 4** Pairwise comparison of the results of the spectral registration (SR)-based FPC methods for the test subset of the simulated data set. The upper triangular shows the correlations of frequency estimations, and the lower triangular shows the correlations of phase estimations. The fitted lines represent the linear regression model and the 95% confidence interval. The type of simulated nuisance peak is color-coded. The total  $R^2$  values are calculated along the corresponding axes. The dashed gray lines are identity lines. Note that the reference signals of relative FPC methods were different, which resulted in shifting their scatter plot from the identity line. Abbreviations: dSR, deep learning-based SR; dSRF, dSR over a limited frequency range; SRF, SR over a limited frequency range

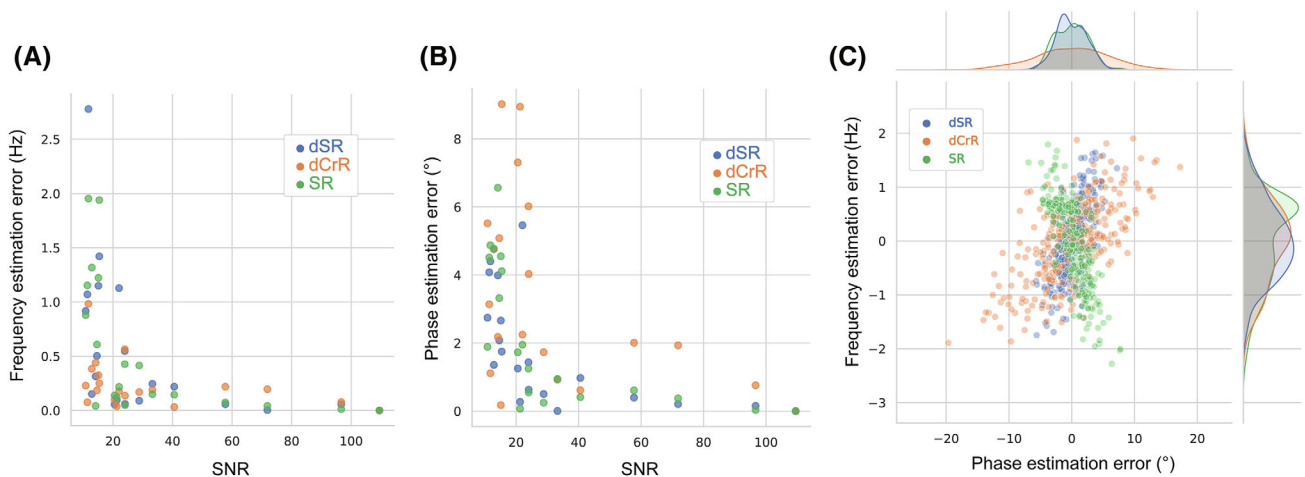
as supervised DL-based methods. We evaluated the ability of each of these methods to estimate the frequency and phase shifts in the simulated MRS data set with known shifts and at varying SNR levels. Our results (Table 1) indicated an improvement in performance in terms of precision and the SI, the latter of which is a new measure proposed in this study for evaluating the FPC performance. Additionally, we compared our results with those obtained using other methodologies in terms of the linewidth of

the Cr peak of the averaged signal obtained by summing corrected transients, and we discovered that the unsupervised DL-based methodology performed comparably to others.

We found that the traditional FPC over a limited frequency range performed poorly in spectra with low SNR (Table 1), whereas our proposed FPC methods over a limited frequency range (dSRF and dCrRF) can perform equally (Table 1) to our proposed



**FIGURE 5** Frequency and phase correction of the phantom test subset using the dCrR method. Uncorrected (A) and corrected (B) spectra from the test subset. Dark blue and green spectra show the average uncorrected and corrected spectra, respectively. The circled inset shows the zoomed creatine (Cr) peak at 3 ppm. The similarity matrix of 64 samples of the test subset before (C) and after (D) FPC



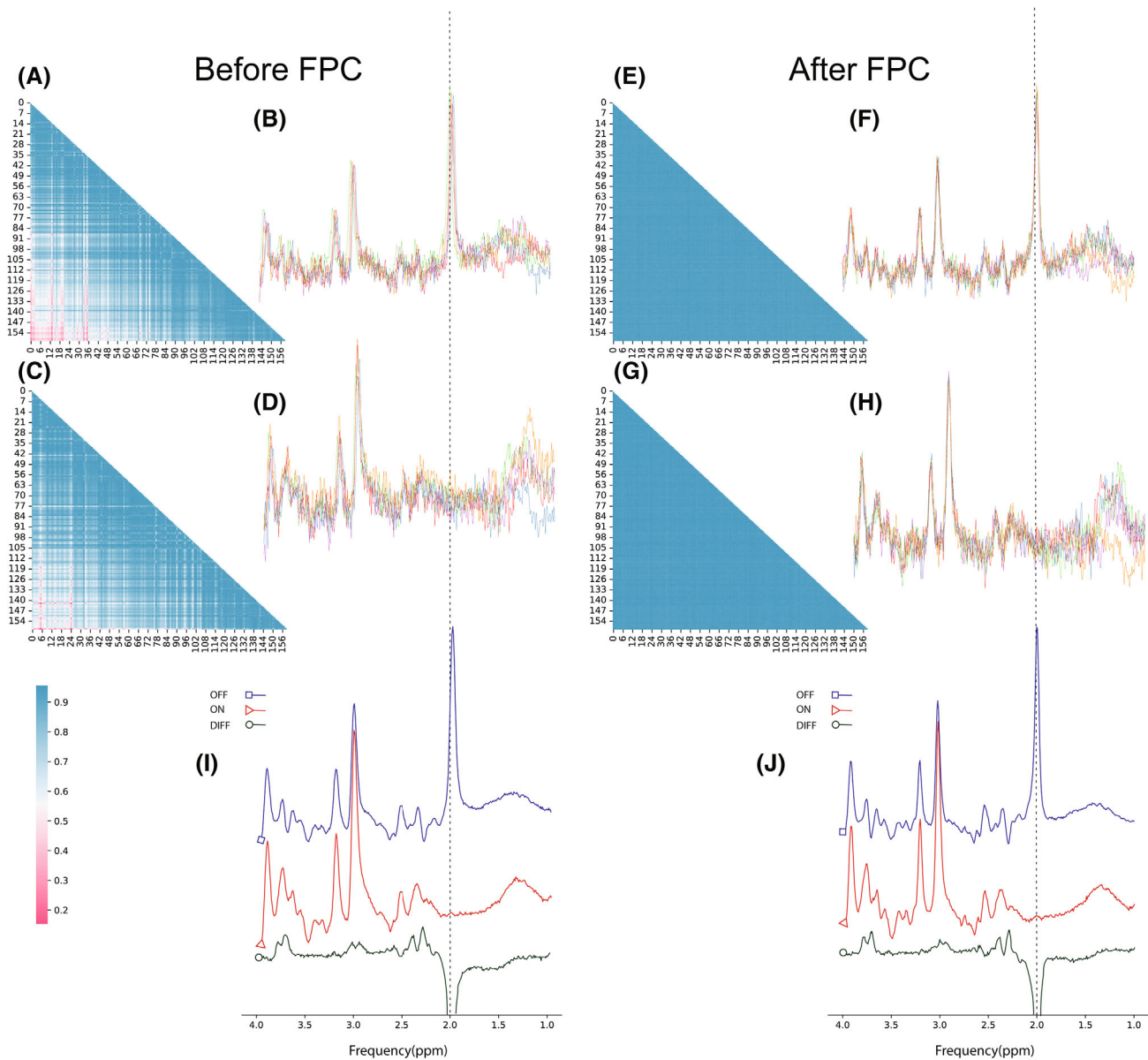
**FIGURE 6** A, B, Comparison of frequency (A) and phase (B) correction precision of the dSR, dCrR, and SR methods over various SNR levels. C, The results of Monte Carlo (MC) analysis. Comparison of dSR, dCrR, and SR methods. For the sake of visualization of absolute and relative methods alike, the estimations from each method were subtracted from their average value. The results of the dCrR method without subtracting from its average value can be found in Supporting Information Figure S1

methods operating in the time domain (dCrR and dSR). Our methods also performed well in a Monte Carlo study, in which the phase-shift and frequency-shift estimate precisions were found to be very good

and in general accordance with the SR method (Figure 6C).

While the performance of conventional SR-based FPC methods can depend on the presence of nuisance peaks





**FIGURE 7** An example of FPC using dCrR for a test set in the GABA-edited in vivo data set. Unedited spectra (A) and their similarity matrix (B) before FPC. Edited spectra (C) and their similarity matrix (D) before FPC. Unedited spectra (E) and their similarity matrix (F) after FPC. Edited spectra (G) and their similarity matrix (H) after FPC. (I) Average uncorrected spectra (blue, unedited; red, edited) and their difference (dark green). (J) Average corrected spectra using dCrR (blue, unedited; red, edited) and their difference (dark green)

(Table 1 and Refs 3,7), our result demonstrated that our methods functioned effectively in the simulated data set regardless of the presence of nuisance peaks (Table 1).

Our proposed method can be trained in a few minutes (Table 1) because of using a one-dimensional signal as the input and a relatively tiny network. In the case of large MRS data sets or repeated measurements with the same conditions, the training time will be compensated by the FPC processing time, which is significantly shorter than in methods based on nonlinear least squares (Table 1). Moreover, the training time can be reduced by using well-established methods such as few-shot transfer

learning,<sup>13,44</sup> in which a network pretrained on large simulated or in vivo data sets can be trained in a few iterations using a few samples.

When the SNR of the test signals was lowered, the performance of our proposed methods (dCrR and dSR) and SR was reduced. When the SNR was decreased below the SNR of the training set, the performance deteriorated even further, as the signal is dominated by noise. We observed that in low-SNR signals, the dSR and SR methods performed better in phase-shift estimation, whereas dCrR worked much better in frequency estimation (Figure 6).

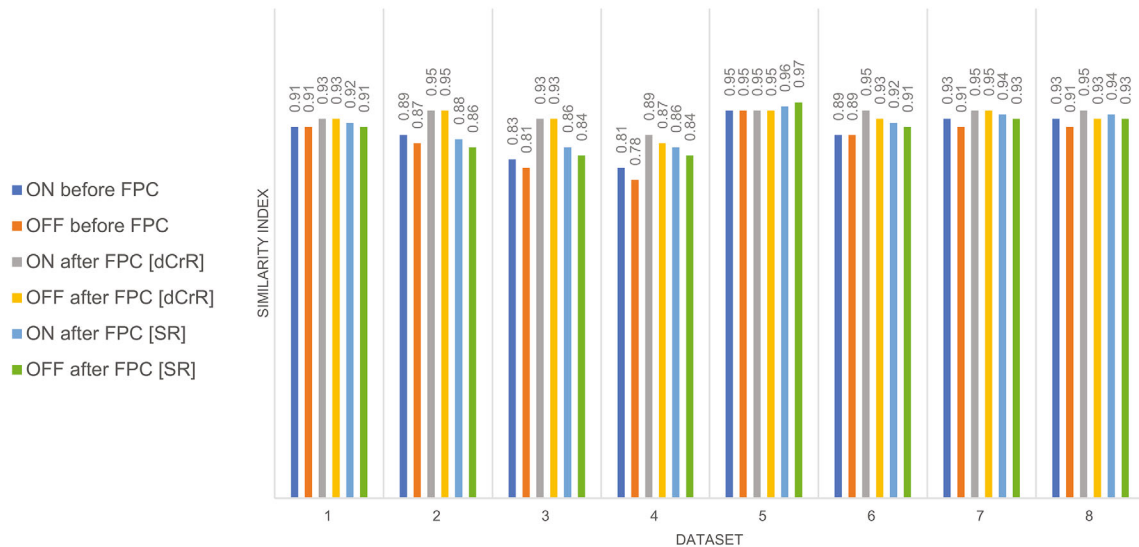


FIGURE 8 Similarity index comparing dCrR to SR for test signals of the Big GABA data set

Regardless of performance, the encoder/model-decoder provides a unique flexibility advantage. It leverages the underlying prior knowledge, which can be beneficial for estimating frequency and phase, independent of the kind of MRS data. Therefore, this approach may be used immediately to almost any sort of MRS data with little or no modification.

Contrary to the previous applications<sup>7,22</sup> of DL in FPC that used a supervised way utilizing simulated data, our proposed network was trained in an unsupervised way. This is advantageous, as most of the MRS data are unlabeled, and simulated data sets may not accurately reflect all in vivo circumstances, such as macromolecules and artifacts. Additionally, a single network was trained in this work to deliver both frequency and phase estimations by including prior knowledge into the decoder, whereas earlier work failed to train a single network.

It has been demonstrated that traditional FPC methods can benefit from additional information in a data set,<sup>10</sup> addressing the problem of selecting a reference signal in SR by using a weighted average reference determined by mutual information in data. The encoder/model-decoder may assist in extracting patterns and information from data by introducing more complex models.

Overfitting is a common pitfall in DL.<sup>13</sup> We implemented a dropout layer in the input to remove a part of the input randomly in every training step, which is a computationally inexpensive and very effective regularization strategy for decreasing overfitting and increasing the generalization of the network.<sup>32</sup>

Training neural networks for regression problems necessitates a well-calibrated estimation. The results revealed a significant linear link between the true and estimated values (Figures 3 and 4), indicating a well-calibrated

estimation, although additional examination of the results is necessary.

Along with demonstrating the performance of the proposed approach on simulated and phantom data, the method was used to carry out FPC and enhance the similarity of signals in publicly accessible GABA-edited in vivo MRS data. It should be emphasized that the proposed network was fed with both edited and unedited signals and trained simultaneously. The result shows the same performance for the edited and the unedited input.

In general, DL-based techniques are restricted in terms of generalizability,<sup>13</sup> especially DL algorithms in the MR domain due to the fact that their training data might be confined to a single scanner, a single sequence, and/or a single vendor.<sup>45</sup>

In this study, only phantom data and in vivo data sets gathered from four distinct locations utilizing a single vendor and a single sequence were used to demonstrate the applicability of our method. A crucial step toward the generalizability and clinical use of our method is training and testing using multicenter and multivendor data.

One caveat in this study is that comparing processing time per transient between algorithms might not be widely valid, as it might be affected by the parameters and conditions of algorithms. Using FLOPS (floating-point operations per second) to assess the computing cost<sup>46</sup> can help for a better comparison.

In the present study, we focused primarily on the validation of our proposed methods using in silico ground-truth knowledge and showed its application in GABA-edited in vivo MRS data. We are aware that further evaluation and comparison with a more robust method<sup>10</sup> are needed. In addition, Cr referencing-based approaches, such as dCrR, are not the optimal way for GABA-edited



in vivo MRS data alignment,<sup>47</sup> and dSR or a more sophisticated DL-based method should be investigated in future researches.

## 5 | CONCLUSIONS

In general, our proposed time-domain FPC method, which is based on DL networks trained in an unsupervised way with complex data, may yield results comparable to previous FPC methods. The proposed approaches can perform absolute and relative FPC on extensively manipulated data in a shorter amount of time once the network is trained. Thus, our proposed approach could aid in the acceleration of analyzing large MRS data sets. Further study is needed to evaluate the generalizability of the proposed methods for multivendor data.

## ACKNOWLEDGMENT

The authors thank Radim Kořinek, Ph.D. (Czech Academy of Sciences, Institute of Scientific Instruments, Czech Republic), for his valuable technical support.

## FUNDING INFORMATION


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 Skłodowska-Curie grant agreement No 813120 (INSPIRE-MED) and has been also supported with the institutional support RVO:68081731 - Czech Academy of Sciences, Institute of Scientific Instruments.

## DATA AVAILABILITY STATEMENT


The source code is freely available at <https://github.com/isi-nmr/DeepFPC>. For questions, please contact the authors. For testing the frequency domain correlation and frequency domain correlation over a limited frequency range, we developed a *MATLAB* script, which is publicly accessible at <https://github.com/isi-nmr/Frequency-and-Phase-Correction-of-MRS-signals-Using-Cross-Correlation>.

## ORCID

Amirmohammad Shamaei  <https://orcid.org/0000-0001-8342-3284>

Jana Starcukova  <https://orcid.org/0000-0003-0337-7893>

Iveta Pavlova  <https://orcid.org/0000-0003-0961-8272>

Zenon Starcuk Jr.  <https://orcid.org/0000-0002-1218-0585>

## TWITTER

Amirmohammad Shamaei  @AmirLand6

## REFERENCES

- de Graaf RA. In vivo NMR spectroscopy. *Vivo NMR Spectrosc.* 2019;560:375-437.
- Hui SCN, Mikkelsen M, Zöllner HJ, et al. Frequency drift in MR spectroscopy at 3T. *Neuroimage.* 2021;241:118430.
- Near J, Edden R, Evans CJ, Harris A, Jezzard P. Frequency and phase drift correction of magnetic resonance spectroscopy data by spectral registration in the time domain. *Magn Reson Med.* 2015;50:44-50.
- Near J, Harris AD, Juchem C, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR Biomed.* 2021;34:e4257.
- Wieggers EC, Philips BWJ, Heerschap A, van der Graaf M. Automatic frequency and phase alignment of in vivo J-difference-edited MR spectra by frequency domain correlation. *MAGMA.* 2017;30:537-544.
- Keating B, Deng W, Roddey JC, et al. Prospective motion correction for single-voxel 1H MR spectroscopy. *Magn Reson Med.* 2010;64:672-679.
- Tapper S, Mikkelsen M, Dewey BE, et al. Frequency and phase correction of J-difference edited MR spectra using deep learning. *Magn Reson Med.* 2021;85:1755-1765.
- Helms G, Piringner A. Restoration of motion-related signal loss and line-shape deterioration of proton MR spectra using the residual water as intrinsic reference. *Magn Reson Med.* 2001;46:395-400.
- Wilson M. Robust retrospective frequency and phase correction for single-voxel MR spectroscopy. *Magn Reson Med.* 2019;81:2878-2886.
- Mikkelsen M, Tapper S, Near J, Mostofsky SH, Edden RAE. Correcting frequency and phase offsets in MRS data using robust spectral registration. *NMR Biomed.* 2020;33:e4368.
- Ernst T, Li J. A novel phase and frequency navigator for proton magnetic resonance spectroscopy using water-suppression cycling. *Magn Reson Med.* 2011;65:13-17.
- Waddell KW, Avison MJ, Joers JM, Gore JC. A practical guide to robust detection of GABA in human brain by J-difference spectroscopy at 3T using a standard volume coil. *Magn Reson Imaging.* 2007;25:1032-1038.
- Goodfellow I, Bengio Y. *Courville A. Deep Learning.* MIT Press; 2016.
- Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys.* 2019;29:102-27.
- Gurbani SS, Sheriff S, Maudsley AA, Shim H, Cooper LAD. Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting. *Magn Reson Med.* 2019;81:3346-3357.
- Lee HH. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. *Magn Reson Med.* 2019;82:33-48.
- Kreis R, Kyathanahally SP. Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy. *Magn Reson Med.* 2018;863:851-863.
- Lee H, Lee HH, Kim H. Reconstruction of spectra from truncated free induction decays by deep learning in proton

- magnetic resonance spectroscopy. *Magn Reson Med.* 2020;84:559-568.
19. Klukowski P, Augoff M, ZieRba M, Drwal M, Gonczarek A, Walczak MJ. NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. *Bioinformatics.* 2018;34:2590-2597.
  20. Iqbal Z, Nguyen D, Hangel G, Motyka S, Bogner W, Jiang S. Super-resolution 1H magnetic resonance spectroscopic imaging utilizing deep learning. *Front Oncol.* 2019;9:1-13.
  21. Gurbani SS, Schreibmann E, Maudsley AA, et al. A convolutional neural network to filter artifacts in spectroscopic MRI. *Magn Reson Med.* 2018;80:1765-1775.
  22. Ma DJ, Le HAM, Ye Y, et al. MR spectroscopy frequency and phase correction using convolutional neural networks. *Magn Reson Med.* 2022;87:1700-1710.
  23. Zhang Z, Zha H. Linear low-rank approximation and nonlinear dimensionality reduction. *Sci China, Ser A Math.* 2004;47:908-920.
  24. Melas-Kyriazi L. The Mathematical Foundations of Manifold Learning. arXiv:2011.01307. 2020.
  25. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science.* 2006;313:504-507.
  26. Mikkelsen M, Barker PB, Bhattacharyya PK, et al. Big GABA: edited MR spectroscopy at 24 research sites. *Neuroimage.* 2017;159:32-45.
  27. Mikkelsen M, Rimbault DL, Barker PB, et al. Big GABA II: water-referenced edited MR spectroscopy at 25 research sites. *Neuroimage.* 2019;191:537-548.
  28. Edden RAE, Puts NAJ, Harris AD, Barker PB, Evans CJ. Gannet: a batch-processing tool for the quantitative analysis of gamma-aminobutyric acid-edited MR spectroscopy spectra. *J Magn Reson Imaging.* 2014;40:1445-1452.
  29. Kreis R, Cudalbu C, De GRA, et al. Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: background and experts' consensus recommendations. *NMR Biomed.* 2021;34:e4347.
  30. Lam F, Li Y, Peng X. Constrained magnetic resonance spectroscopic imaging by learning nonlinear low-dimensional models. *IEEE Trans Med Imaging.* 2020;39:545-555.
  31. Tkáč I, Starčuk Z, Choi I-Y, Gruetter R. In vivo 1 H NMR spectroscopy of rat brain at 1 ms echo time. *Magn Reson Med.* 1999;41:649-656.
  32. Srivastava N, Hinton G, Krizhevsky A, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929-1958.
  33. Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern.* 1980;36:193-202.
  34. Schmidhuber, J. (2014). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>
  35. Sokolenko S, Jézéquel T, Hajjar G, Farjon J, Akoka S, Giraudeau P. Robust 1D NMR lineshape fitting using real and imaginary data in the frequency domain. *J Magn Reson.* 2019;298:91-100.
  36. Brunton SL, Kutz JN. *Data-driven science and engineering*; Cambridge University Press; 2019.
  37. Bisong E. Google Colaboratory. New York: Cambridge University Press; 2019.
  38. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst.* 2019;32:1-12.
  39. Falcon W. PyTorch lightning. GitHub. <https://github.com/PyTorchLightning/pytorch-lightning>.
  40. Falkner S, Klein A, Hutter F. BOHB: Robust and Efficient Hyperparameter Optimization at Scale. In: 35th International Conference on Machine Learning, PMLR; 2018:1437-1446. Accessed March 30, 2019.
  41. Liaw R, Liang E, Nishihara R, Moritz P, Gonzalez JE, Stoica I. Tune: a research platform for distributed model selection and training. 2018. <https://arxiv.org/abs/1807.05118v1>. Accessed January 25, 2022.
  42. Kingma DP, Ba JL. Adam: a method for stochastic optimization. In: Proceedings of the Third International Conference on Learning, San Diego, California, 2015.
  43. Simpson R, Devenyi GA, Jezzard P, Hennessy TJ, Near J. Advanced processing and simulation of MRS data using the FID appliance (FID-A)—an open source MATLAB-based toolkit. *Magn Reson Med.* 2017;77:23-33.
  44. Sun Q, Liu Y, Chua T-S, Schiele B. Meta-transfer learning for few-shot learning. <https://github.com/y2l/meta-transfer-learning-tensorflow>. Accessed August 15, 2022.
  45. Grøvik E, Yi D, Iv M, et al. Handling missing MRI sequences in deep learning segmentation of brain metastases: a multicenter study. *npj Digit Med.* 2021;41:1-7.
  46. Liu Z, Mao H, Wu C-Y, et al. A ConvNet for the 2020s. 2022. <https://arxiv.org/abs/2201.03545v1>. Accessed February 11, 2022.
  47. Evans CJ, Puts NAJ, Robson SE, et al. Subtraction artifacts and frequency (Mis-)alignment in J-difference GABA editing. *J Magn Reson Imaging.* 2013;38:970-975.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**Figure S1** Bin-based visualization of the result of the Monte Carlo analysis of the simulated data set

**Figure S2** Online monitoring of precision of a set of transients in the validation subset during training

**Figure S3** Validation loss (mean square error [MSE]) versus training steps

**Figure S4** Process flow of one-shot frequency and phase correction of J-difference-edited MR spectra using a single deep neural network

**Figure S5** Illustration of the proposed convolutional encoder-model decoder for deep learning-based spectral registration (dSR) method

**Figure S6** Reference scan selected by our proposed machine learning-based algorithm for the deep spectral registration method

**Table S1** Parameters for generating the simulated nuisance peak

**Table S2** Summary of the encoder's network

**Text S1** Bayesian hyper-parameterization

**Text S2** A novel machine learning-based algorithm for finding the reference scan for the dSR method

**Equation S1** Equation for deriving each element of the similarity matrix

**How to cite this article:** Shamaei A, Starcukova J, Pavlova I, Starcuk Z. Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals. *Magn Reson Med*. 2023;89:1221-1236. doi: 10.1002/mrm.29498

# DEEP LEARNING FOR MAGNETIC RESONANCE SPECTROSCOPY QUANTIFICATION: A TIME-FREQUENCY ANALYSIS APPROACH

**Amirmohammad Shamaei**

Doctoral Degree Programme (1st year at FEEC BUT), Czech Academy of Sciences, Institute of Scientific Instruments

E-mail: [amirshamaei@isibrno.cz](mailto:amirshamaei@isibrno.cz)

Supervised by: Radovan Jiřík, Jana Starčuková

E-mail: [jirik@feec.vutbr.cz](mailto:jirik@feec.vutbr.cz); [jana@isibrno.cz](mailto:jana@isibrno.cz)

**Abstract:** Magnetic resonance spectroscopy (MRS) is a technique capable of detecting chemical compounds from localized volumes in living tissues. Quantification of MRS signals is required for obtaining the metabolite concentrations of the tissue under investigation. However, reliable quantification of MRS is difficult. Recently deep learning (DL) has been used for metabolite quantification of MRS signals in the frequency domain. In another study, it was shown that DL in combination with time-frequency analysis could be used for artifact detection in MRS. In this study, we verify the hypothesis that DL in combination with time-frequency analysis can also be used for metabolite quantification and yields results more robust than DL trained with MR signals in the frequency domain. We used the complex matrix of absolute wavelet coefficients (WC) for the time-frequency representation of the signal, and convolutional neural network (CNN) implementation for DL. The comparison with DL used for quantification of data in the frequency domain is presented.

**Keywords:** magnetic resonance spectroscopy; quantification; deep learning; machine learning;

## 1 INTRODUCTION

Magnetic Resonance Spectroscopy (MRS) has attracted the MR community over the past 7 decades [1]. A significant part of the interest in biomedical MRS stems from the possibility of noninvasive measurements of metabolites. Information about tissue metabolites can help in clinical diagnostics. For example, detection of metabolic pathway changes may facilitate diagnosing disease in earlier stages before anatomy changes can be observed [1], [2], and thus enable more efficient treatment. E.g., in glioma, a decrease of N-acetylaspartate (NAA) and creatine concentrations of NAA and creatine and an increase of choline, lipids, and lactate predicts an increase of the glioma grade. To reach such a goal, at first, we need to quantify metabolic concentrations. Because there are many obstacles to reaching an accurate estimate of the metabolite concentrations, the use of MRS in daily clinical practice is still not common. The existing MRS quantitation methods are based on model fitting of a signal either in the time or the frequency domain [3]. Even though, in theory there is no difference in which domain is used for fitting, the reality in practice could be different.

Deep learning has achieved many accomplishments in a wide range of tasks, including the MRI field [4]. Due to the poor signal-to-noise ratio (SNR), chemical shift displacement, and overlapping of signal components of the MRS signal, deep learning can be a useful tool. Recently, Hatami et al. showed the first step in this area by using the deep learning approach for MRS signal quantification [5]. Kim et al. conducted a comprehensive study on brain metabolite quantification using deep learning [6]. The input of both studies is a signal in the frequency domain (metabolite spectra), and their network is a 1D convolutional neural network (CNN). As we mentioned earlier, there are differences between time and frequency domain quantification in practice. Be a case in point, elimination of the first few distorted data points of a signal in the time domain does not significantly dis-



turb the time-domain analysis, whereas the missing time-domain data points can result in complicated modulations throughout the entire spectrum [2]. To overcome the difficulties of the signal analysis in a single domain, time-frequency analysis has been carried out for decades in other areas [4]. Nevertheless, finding an accurate tool for the time-frequency analysis is fraught with difficulty. Here is where deep learning comes to play. Thomas et al. constructed time-frequency images of a speech signal and used them as an input to a CNN for classification [7]. Kyathanahally et al. learned a CNN with time-frequency data to detect and remove ghosting artifacts in clinical magnetic resonance spectra of human brain [8]. Given the mentioned accomplishments of deep learning and time-frequency analysis in a variety of different areas, in particular in MRS for signal artifacts detection, this paper describes to our knowledge the first attempt to use this state-of-the-art technique to quantify MRS signal by deep learning and time-frequency analysis. First, we generate simulated MRS signals. Second, we transform the signals to the time-frequency representation. Third, we train a CNN with the new time-frequency representation. Finally, the result is compared with the previous study.

## 2 METHODS

A framework is created to generate MRS signals with different amplitudes, damping factors, and frequency shifts. Second, these one-dimensional signals are transformed into their two-dimensional time-frequency representation using wavelet transformation (WT). Finally, the data are split into two datasets, the training and testing datasets. The input of the CNN is the time-frequency representation of signals, and the output is 21 values, which are the concentration-related amplitudes of 20 metabolites and the amplitude of the background signal. The CNN is trained with a training dataset of signals of known amplitudes. Then, the trained CNN is used to estimate the metabolite amplitudes of the test dataset. Finally, the techniques for accuracy evaluation are used.

### 2.1 SIGNAL GENERATION

Deep learning approaches need a considerable amount of data. For this purpose, we need a basis set (metabolite signals with known concentrations) either simulated or acquired. To be able to compare our results with the previous studies [5], [9], we used the same simulated basis set as used in those studies, i.e., the basis set provided for the ISMRM challenge 2016 [10]. The MRS signal is defined as a combination of amplitude-scaled phase-shifted metabolite basis set signals, the baseline and noise (in this study we use a noiseless signal). The mathematical model for the parametric part of the MRS signal is given by:

$$S[n] = \left[ \sum_{m=1}^M A_m \cdot X_m[n] \cdot e^{(\Delta\alpha_m + 2i\pi\Delta f_m n\Delta T)} \right] + A_{MM} \cdot MM[n] \cdot e^{(\Delta\alpha_{MM} + 2i\pi\Delta f_{MM} n\Delta T)} \quad (1)$$

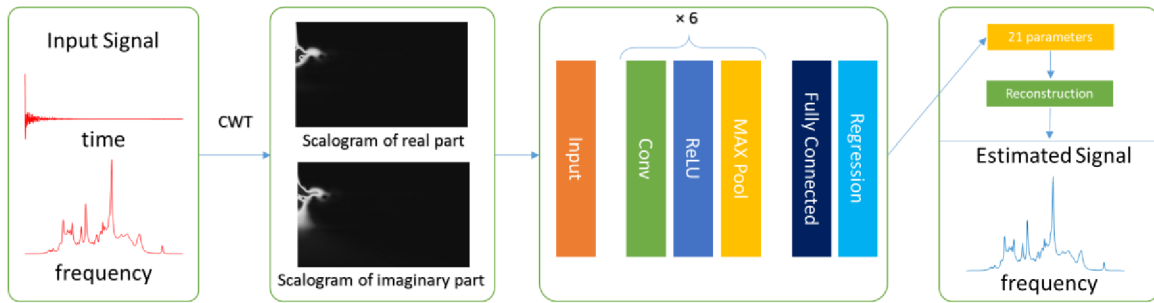
where  $X_m[n]$  is the  $n$ -th sample of the  $m$ -th simulated metabolite,  $\Delta T$  is a sampling period,  $A_m$  is the scaling factor of the metabolite ( $A_m \cdot X_m[0]$  is an indication of the metabolite concentration),  $\Delta\alpha_m$  is the damping factor,  $\Delta f_m$  is the frequency shift of the  $m$ -th metabolite, and  $M$  is the number of metabolites. For our signal simulation the values of the amplitude, damping, and frequency shift are chosen randomly from a defined range with a uniform distribution ( $A = [0, 1]$ ,  $\Delta\alpha = [-10, 10]$  and  $\Delta f = [-10, 10]$ ). The known background signal  $MM$  is considered as another metabolite, then is added to  $S[n]$  with a random scaling factor, damping, and frequency shift ( $A_{MM}$ ,  $\Delta\alpha_{MM}$  and  $\Delta f_{MM}$ ). Ten thousand signals are generated, in which the process of value selection is entirely random, thus preventing any bias to our train dataset. The basis set used was simulated for sequence PRESS, magnetic field 3T, echo time  $TE = 30$  ms, spectrum width  $SW = 4000$  Hz, and 2048 time-domain samples.

## 2.2 SIGNAL PROCESSING

The time-frequency representation of the 1D signal shows a signal in both the time and frequency domain simultaneously. One of the forms of the time-frequency representation of the signal is a scalogram (a matrix of absolute values of the continuous wavelet coefficients (CWC) of a signal ) that can be plot as a function of time and frequency. The scalogram is calculated using the Matlab Wavelet Toolbox (R2019a, Mathworks Inc.,Natick, MA, USA). We use Morse wavelet to compute the CWC. The last 512 points of the time signals are cut off to reduce the amount of computation for the CWC calculation. The selection of the number of points was decided by visual inspection of the signals to ensure that no significant information will be lost. The wavelet coefficients are computed. The minimum and maximum scales are determined automatically based on the energy spread of the wavelet in frequency and time by the toolbox. The coefficients matrix is a matrix where each row corresponds to one scale, and its column size is equal to the length of signal. Scalogram with 340 frequency bins and 1536 time points ( $340 \times 1536$  matrix) is created. Finally, the real and imaginary parts of 10000 matrices are stored in two channels.

## 2.3 CNN

A convolutional neural network is developed using the Matlab Deep Learning Toolbox (R2019a, Mathworks Inc.,Natick, MA, USA) on NVIDIA GTX 1050Ti graphics processing units. The architecture of the CNN is shown in Fig. 1 . This network includes one input layer with two channels, six convolutional layers, five max pool layers, and one regression layer. Rectified linear unit (ReLU) activation functions are used between CL and MP layers. The mean square error is implemented as the loss function. The output of regression layer is 21 parameters which correspond to twenty metabolites and one background MM. Using these parameters and Eq (1), the estimated signal is reconstructed.



**Figure 1:** A schematic of the proposed approach. The generated signal based on a linear combination of metabolites basis sets is converted to two gray Scalogram images (real and imaginary). These images feed to CNN as inputs. The CNN includes 6 blocks, which comprise a Convolutional, Rectifier and Max-pooling layer. The last layers are a fully connected and a regression layer (which has 21 outputs). The estimated signal is reconstructed with the estimated parameters.

## 2.4 ACCURACY EVALUATION

Two methods are used to measure the accuracy of the model. First, the mean absolute error, which is the most straightforward regression error metric. MAE is defined as below for each metabolite:

$$\text{MAE}[m] = \frac{1}{N} \sum_{n=1}^N |A_{mn} - A'_{mn}| \quad (2)$$

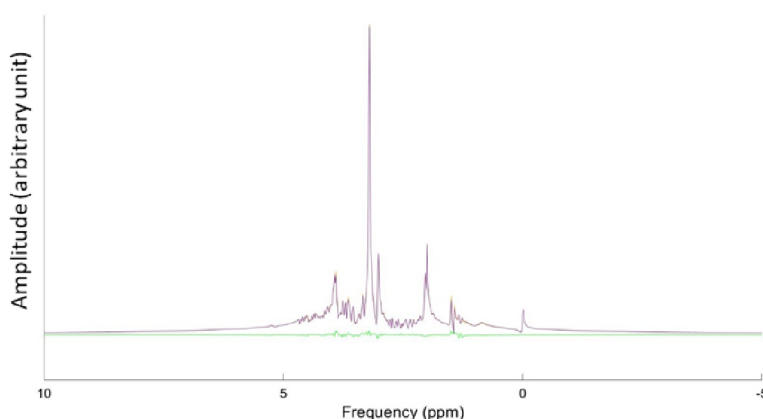
where  $m$ ,  $N$ ,  $A$ , and  $A'$  are the metabolite index, the number of test datasets, the ground truth, and the estimated amplitude, respectively. The second method is the Symmetric mean absolute percentage error (SMAPE) which is given by:

$$\text{SMAPE}[m] = \frac{\sum_{n=1}^N |A_{mn} - A'_{mn}|}{\sum_{n=1}^N |A_{mn} + A'_{mn}|} \quad (3)$$

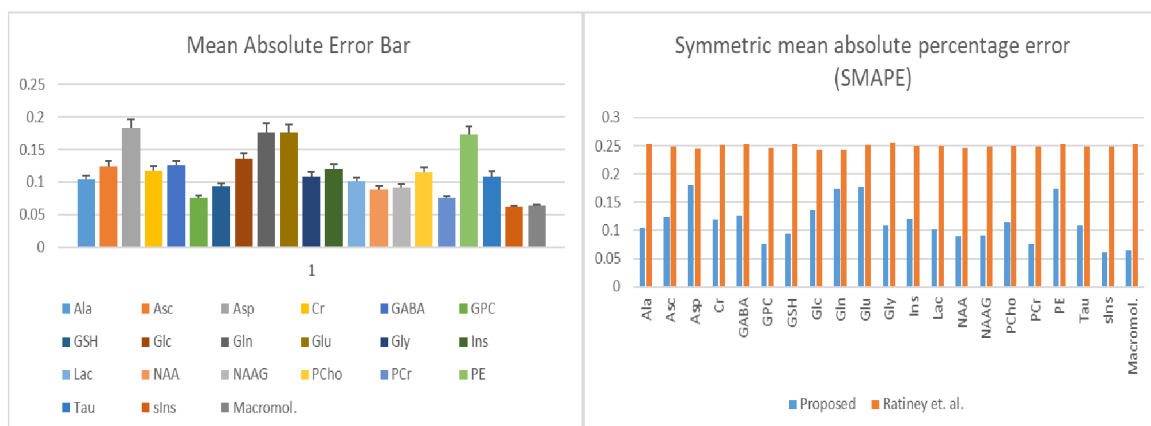
## 2.5 RESULTS

The dataset is separated into two datasets, namely a training dataset and a test dataset. The training dataset contains 80% of the data and the remaining 20% are the test dataset. CNNs with different hyperparameters such as minimum batch size, initial learning rate, and validation frequency are tested, and the CNN with the best result is chosen. The minimum batch size, initial learning rate, and validation frequency are 30, 1e-5, and 10, respectively. It has been shown that increasing the training sample would decrease the value of loss function [5]. Nonetheless, to be able to compare results obtained with our new approach (DL with time-frequency domain input) with the results of the Hatami et al approach (DL with frequency domain input) in a reasonable time, we decided to use only 10000 samples for CNN training and testing. Training and validation loss for the given dataset are 0.18 and 0.23, respectively.

Fig. 2 shows one of the tested (ground truth) signal, its estimate, and residual. The following conclusions may be drawn from this figure. First, the method used is able to estimate the tested signal. Second, residuals mainly occur when the signal shows rapid fluctuation.



**Figure 2:** Example of the signal estimation – ground truth signal (orange), estimated signal (violet) and residual signal (green).



**Figure 3:** (left) Mean absolute error bar of every metabolite and its variance. (right) The symmetric mean absolute percentage error of each metabolite for (blue) our study (orange) Hatami et al.[5]

The mean absolute errors (MAE) of metabolites are shown in Fig. 3 (left). Even though the amount of error is not too low compared to the amplitude range ([0, 1]), the variance of the error is small. Fig. 3 (right) shows the comparison of the Symmetric mean absolute percentage error (SMAPE) between our approach and Hatami et al. approach. To avoid any bias in comparison, we used

the CNN described in this study and train it with our training set but in one case in the form of scalogram and in the other case (Hatami et al.) with the data in the frequency domain. The proposed approach shows less amount of error compared to their method.

### 3 CONCLUSION

Quantification of MRS is an important topic where a robust and universal panacea approach to quantify signals is needed. It was shown in this study that time-frequency deep learning quantification could outperform single domain quantification used in the previous studies [2, 5] and hopefully as a method using information from both MRS domains be successfully used also for quantification of signals with artifact patterns [8]. The next steps may be to verify the tested approach on 1) the simulated noisy MRS with different signal-to-noise-ratios and for different pulse sequences 2) on real MRS acquired from a phantom, 3) on MRS acquired from a rat, and 4) to implement this approach as a plugin in the jMRUI software package [11].

### ACKNOWLEDGEMENT

This research was supported by European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 813120 (INSPIRE-MED).

### REFERENCES

- [1] M. Van Der Graaf, "In vivo magnetic resonance spectroscopy: Basic methodology and clinical applications," *Eur. Biophys. J.*, vol. 39, no. 4, pp. 527–540, 2010.
- [2] Robin A. de Graaf, "In Vivo NMR Spectroscopy," *In Vivo NMR Spectroscopy*. pp. 1–42, 13-Feb-2019.
- [3] J. B. Pouillet, D. M. Sima, and S. Van Huffel, "MRS signal quantitation: A review of time- and frequency-domain methods," *J. Magn. Reson.*, vol. 195, no. 2, pp. 134–144, 2008.
- [4] H. Alaskar, "Deep Learning-Based Model Architecture for Time-Frequency Images Analysis," no. January 2018, 2019.
- [5] N. Hatami, B. Lyon, and U. Etienne, "Magnetic Resonance Spectroscopy Quantification using Deep Learning," 2018.
- [6] H. H. Lee, "Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain," no. January, pp. 33–48, 2019.
- [7] S. Thomas, S. Ganapathy, G. Saon, and H. Soltau, "Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2014, pp. 2519–2523.
- [8] R. Kreis and S. P. Kyathanahally, "Deep Learning Approaches for Detection and Removal of Ghosting Artifacts in MR Spectroscopy," vol. 863, pp. 851–863, 2018.
- [9] D. Das, E. Coello, R. F. Schulte, and B. H. Menze, "Quantification of Metabolites in Magnetic Resonance Spectroscopic Imaging Using Machine Learning," vol. 3, pp. 462–469, 2017.
- [10] ISMRM, "MRS Fitting Challenge ([ismrm.org/workshops/Spectroscopy16/mrs\\_fitting\\_challenge/](http://ismrm.org/workshops/Spectroscopy16/mrs_fitting_challenge/))," 2016. [Online]. Available: [ismrm.org/workshops/Spectroscopy16/mrs\\_fitting\\_challenge/](http://ismrm.org/workshops/Spectroscopy16/mrs_fitting_challenge/).
- [11] D. Stefan *et al.*, "Quantitation of magnetic resonance spectroscopy signals: The jMRUI software package," *Meas. Sci. Technol.*, vol. 20, no. 10, 2009.



# A Wavelet Scattering Convolutional Network for Magnetic Resonance Spectroscopy Signal Quantitation

Amirmohammad Shamaei<sup>1,2</sup><sup>a</sup>, Jana Starčuková<sup>1</sup><sup>b</sup> and Zenon Starčuk Jr.<sup>1</sup><sup>c</sup>

<sup>1</sup>*Institute of Scientific Instruments of the CAS, Královopolská 147, 612 64 Brno, Czech Republic*

<sup>2</sup>*Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 3058/10, 616 00 Brno, Czech Republic*


**Keywords:** Magnetic Resonance Spectroscopy, Quantification, Deep Learning, Machine Learning.


**Abstract:** Magnetic resonance spectroscopy (MRS) can provide quantitative information about local metabolite concentrations in living tissues, but in practice the quantification can be difficult. Recently deep learning (DL) has been used for quantification of MRS signals in the frequency domain, and DL combined with time-frequency analysis for artefact detection in MRS. The networks most widely used in previous studies were Convolutional Neural Networks (CNN). Nonetheless, the optimal architecture and hyper-parameters of the CNN for MRS are not well understood; CNN has no knowledge about the nature of the MRS signal and its training is computationally expensive. On the other hand, Wavelet Scattering Convolutional Network (WSCN) is well-understood and computationally cheap. In this study, we found that a wavelet scattering network could hopefully be also used for metabolite quantification. We showed that a WSCN could yield results more robust than QUEST (one of quantitation methods based on model fitting) and the same as a CNN while being faster. We used wavelet scattering transform to extract features from the MRS signal, and a superficial neural network implementation to predict metabolite concentrations. Effects of phase, noise, and macromolecules variation on the WSCN estimation accuracy were also investigated.


## 1 INTRODUCTION

Magnetic Resonance Spectroscopy (MRS) has attracted the MR community over the past seven decades (Van Der Graaf, 2010). A significant part of the interest in biomedical MRS stems from the possibility of noninvasive measurements of metabolites. Information about tissue metabolites can help in clinical diagnostics. For instance, detection of metabolic pathway changes may facilitate diagnosing disease in earlier stages before anatomy changes can be observed, and thus enable more efficient treatment. E.g., in glioma, a decrease of N-acetylaspartate (NAA) and creatine concentrations and an increase of choline, lipids, and lactate predicts an increase of the glioma grade (Robin A. de Graaf, 2019; Van Der Graaf, 2010). To detect such changes, quantification of MRS signals is required for obtaining the metabolite concentrations in the tissue. However,

reliable quantification of MRS is difficult. The existing MRS quantitation methods are based on model fitting of the signal in either the time or the frequency domain (Pouillet et al., 2008). In recent years, new novel machine learning solutions have been proposed for quantification, one of which is deep learning (DL). Even though the first usage of machine learning dates back to the 1970s, it was unpractical until the past decade due to lack of high-performance hardware and novel algorithms (Chen et al., 2020). DL has achieved many accomplishments in a wide range of tasks, including the MRI field (Alaskar, 2019). Due to the poor SNR, chemical shift displacement, and overlapping signal components in MRS signals, only recently has DL been used for metabolite quantification of MRS signals in the frequency domain (Hatami et al., 2018; Lee & Kim, 2019)

<sup>a</sup>  <https://orcid.org/0000-0001-8342-3284>

<sup>b</sup>  <https://orcid.org/0000-0003-0337-7893>

<sup>c</sup>  <https://orcid.org/0000-0002-1218-0585>

Hatami et al. showed the first step in this area by using the Convolutional Neural Network (CNN) approach for simulated MRS signal quantification (Hatami et al., 2018). Kim et al. conducted a comprehensive study on brain metabolite quantification using DL (Lee and Kim, 2019). Nonetheless, the practical application of DL in MRS has not been limited to quantifications only. Kyathanahally et al. taught a CNN with time-frequency data to detect and remove ghosting artifacts in clinical magnetic resonance spectra of the human brain (Kreis & Kyathanahally, 2018).

However, the optimal architecture and hyper-parameters of CNN for MRS are not well understood. Besides, training a CNN is a computationally expensive and time-consuming task, and it usually needs a big dataset (Bruna & Mallat, 2013). Moreover, in the case of MRS signals, CNN has no understanding of the nature of the signal, and therefore, any shape difference of the signal under investigation from signals in the training data set can lead to CNN failure. If we look at a CNN as a transformation from the time domain to a features domain, due to the nature of MRS signals, the transformation should be invariant to time shift, deformation in the time domain, and frequency shift. To satisfy such requirements, CNN could be designed as

- an optimized and simple deep architecture which pools the features using a nonlinear averaging measure.
- a network with a fast computational algorithm which is stable to time-shifting, deformation in the time domain, and frequency shift.

Wavelet Scattering Convolutional Network (WSCN) can be a method of choice. WSCNs are well-understood, computationally cheap, and fast for a deep learning task (Andén & Mallat, 2014; Bruna & Mallat, 2013). Wavelet-based methods have previously been used for MRS quantification and water removal (Pouillet et al., 2008; Suvichakorn et al., 2008); but as far as we are aware, wavelet transform has not been implemented by a deep convolutional neural network to quantify MRS signals.

Given the mentioned accomplishments of machine learning in MRS for signal quantification, this paper describes to our knowledge the first attempt to use this state-of-the-art technique to quantify MRS signals by WSCNs. We used wavelet scattering transform to extract features from the free induction decay (FID, i.e. the MRS signal in the time domain) and a superficial neural network implementation to predict metabolite concentrations.

In this study, we used two different basis sets. The first basis set was the ISMRM challenge 2016 simulated basis set for comparing results of our method with the results published for a CNN and another conventional quantification method, QUEST (Graveron-Demilly, 2014). For the second basis set, we simulated our own metabolite signals and generated different synthetic datasets from them for evaluating our method against phase changing, noise, and presence of macromolecule signals.

## 2 METHODS

All steps were run on a laptop with a 4-core Intel i7 processor running at 2.6 GHz and an NVIDIA GTX 1050Ti graphics processing units using Matlab (R2019a, Mathworks Inc., Natick, MA, USA) software.

### 2.1 Simulation of Metabolites

To build a basis-set signals, fifteen metabolites – Alanine (Ala), Aspartate (Asp), Creatine (Cr), Choline (Cho), Gamma Aminobutyric Acid (GABA), Glutathione (GSH), Glutamine (Gln), Glutamate (Glu), Lactate (Lac), N-Acetylaspartate (NAA), N-acetyl-aspartyl-glutamate (NAAG), Phosphatidylcholine (PC), Phosphocreatine (PCr), Taurine (Tau) and myo-Inositol (mIns) – were simulated at 9.4 T magnetic field with the PRESS sequence (TE = 20 ms; TR = 2500 ms; acquisition points: 2048; acquisition bandwidth: 4401.41 Hz; three PRESS pulses with Hermite shapes and flip angles: P1 = 90°, P2 = 180°, P3 = 180°). The simulation was performed in NMRScopeB (Starčuk & Starčuková, 2017; Stefan et al., 2009). The parameters selected in the sequence were taken from an *in-vivo* experiment, which allows reusing the simulated basis set.

### 2.2 Baseline and Macromolecule Simulation

The baseline signals were simulated as a linear combination of several Gaussian lines identified by Osorio-Garcia (Opstad et al., 2008; Osorio-Garcia et al., 2011). The number and parameters of Gaussian lines were extracted from *in-vivo* signals using inversion recovery (Osorio-Garcia et al., 2011).

### 2.3 Signal Generation Framework

The MRS signal was defined as a linear combination

of amplitude-scaled frequency- and phase-shifted metabolite signals, the baseline, and noise.

The model describing a time-domain MRS signal  $s[n]$  as a combination of several metabolite profiles is (Pouillet et al., 2007):

$$s[n] = \sum_{m=1}^M A_m X_m[n] e^{(\Delta\alpha_m + i2\pi\Delta f_m)n\Delta t} e^{i\Delta\theta_m} + A_{MM} MM[n] e^{(\Delta\alpha_{MM} + i2\pi\Delta f_{MM})n\Delta t} e^{i\Delta\theta_{MM}} + \varepsilon[n] \quad (1)$$

where  $X_m[n]$  is the  $n$ -th sample of the  $m$ -th simulated metabolite,  $\Delta T$  is the sampling period,  $A_m$  is the scaling factor of the metabolite,  $\Delta\alpha_m$  is the damping factor,  $\Delta f_m$  is the frequency shift of the  $m$ -th metabolite affected by the static magnetic field inhomogeneity, pH, temperature and chemical composition of the tissue,  $\Delta\theta_m$  is the phase of the  $m$ -th metabolite,  $\Delta t$  is the time step, and  $M$  is the number of metabolites.

Table 1 specifies the range of parameter values used for generating different datasets according to equation (1). For a comparison of our results with the previous study (Hatami et al., 2018), the basis set provided for the ISMRM challenge 2016 (ISMRM, 2016) was used to generate dataset DSS1 (20 metabolite and one macromolecule components). All other datasets were generated using the basis set simulated with NMRScopeB (15 metabolites). The same parameter ranges that were used in the previous study (Hatami et al., 2018) were also used in this study for DSS1, but we decided to choose ranges of parameters for other datasets (DSS2-DSS7) in the same manner as we would do if we evaluated real acquired spectra.

Instead of generating 500 000 signal samples per dataset, in our study only 10 000 signal samples were generated for validating the hypothesis that our network is as robust as Hatami et al.'s approach (Hatami et al., 2018) even with a smaller number of samples but faster. Parameters were chosen randomly from defined ranges with a uniform distribution. In DSS1, random complex Gaussian noise was added to signal samples based on the previous study (Hatami et al., 2018). In the rest of the datasets, the SNR of the signal samples was adjusted by adding random noise such that the SNR was in the range of  $\sim 5$  to  $\sim 15$ . In this study, we used MATLAB built-in `snr` function which calculates the signal-to-noise ratio (SNR) of an MRS signal by computing the ratio of its summed squared magnitude to that of the noise. In Table 1, the presence of a parameter is marked by a tick and the absence of a parameter by a cross.

## 2.4 Deep Learning

### 2.4.1 Invariant Wavelet Scattering Network

Invariant wavelet scattering network is a transform from the time domain to the features domain, which has three stages, namely, Convolution (wavelet), Non-linearity, and Averaging (scaling factor).

In contrast to the classical wavelet transform, the Complex wavelet transform is translation invariant. In this study, we chose Morlet (Gabor) wavelets, a type of complex wavelet transform, because they have a simple mathematical representation.

Figure 1 illustrates the wavelet scattering transform processes (see (Andén & Mallat, 2014; Bruna & Mallat, 2013) for more details). In practice, a scattering decomposition framework was created with a signal input length of 1024 samples.

Table 1: Specification of datasets.

Name	Amplitude ( $A_m$ )	Frequency shift ( $\Delta f_m$ )	Damping range ( $\Delta\alpha_m$ )	Noise ( $\epsilon$ )	Phase ( $\Delta\theta_m$ )		MM ( $A_{MM}$ )	
					Common	Separated	Constant	Changing
DSS1 (Hatami et al., 2018)	[0, 1]	[-10, 10]	[-10, 10]	✓	×	×	×	✓
DSS2	[0.5 1]	[-10, 10]	[-5, 2.5]	×	×	×	×	×
DSS3	[0.5 1]	[-10, 10]	[-5, 2.5]	×	$[-\frac{\pi}{8}, \frac{\pi}{8}]$	×	×	×
DSS4	[0.5 1]	[-10, 10]	[-5, 2.5]	×	×	$[-\frac{\pi}{8}, \frac{\pi}{8}]$	×	×
DSS5	[0.5 1]	[-10, 10]	[-5, 2.5]	✓	$[-\frac{\pi}{8}, \frac{\pi}{8}]$	×	×	×
DSS6	[0.5 1]	[-10, 10]	[-5, 2.5]	✓	✓	×	✓	×
DSS7	[0.5 1]	[-10, 10]	[-5, 2.5]	✓	✓	×	×	Within $\pm 10$ percent of initial values

The framework had two filter banks; in other words, the depth of the framework was 3. The quality factor (the number of wavelet filters per octave) of the first and second filter banks were 8 and 1, respectively. For the given signal length and quality factors, the output of the framework was a matrix with a size of 154 by 8 by 2. There were 154 scattering paths and 8 scattering windows for each of the real and imaginary parts of the signal.

### 2.4.2 Regression

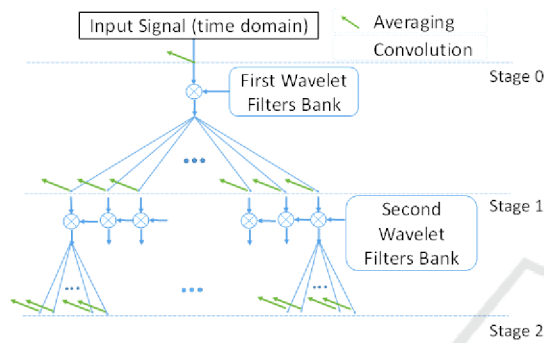


Figure 1: The process of wavelet scattering network; averaging and convolution of a signal with wavelet filters are showed by an arrow (green) and a circled star, respectively.

Flattening and fully-connected layers were what we had at the last stage of our network. The first step, so-called flattening, was converting a feature matrix into a 1-dimensional array. The matrices from the output of WSCN were flattened to create a single long feature vector. The flattening layer was connected to a fully-connected layer, which was a feedforward artificial neural network for the regression task. Neural networks with the different number of neurons in hidden layers were investigated. The best fully-connected layer structure was obtained by trial and error on the basis of the lowest error on the training and validation dataset. The results showed that one hidden-layer network with 20 neurons in the hidden layer yielded better results than other network types. The modeling performance and training were evaluated by the mean square error (MSE) and scaled conjugate gradient, respectively.

Figure 2 demonstrates the process of transformation, flattening, and regression. The input and output of a fully-connected layer were the features vector and the relative amplitudes of various metabolite basis spectra, respectively.

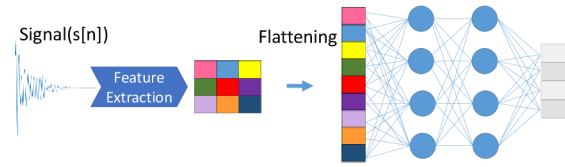


Figure 2: A schematic of feature extraction and flattening and the training of an artificial neural network.

### 2.4.3 Quantification

80% of each dataset was allocated to the training set, 10% for validation and the rest 10% for the test set. It applied to all datasets, DSS1 to DSS7, and then they were fed to the network. First, the network was trained with the training dataset; then, it was used to predict the test dataset. The output of the network was a vector in which each element represents the relative amplitude of each metabolite.

### 2.5 Accuracy Evaluation

The Symmetric mean absolute percentage error (SMAPE) is used to measure the accuracy of the model. SMAPE is defined as below for each metabolite:

$$\text{SMAPE}[m] = \frac{\sum_{n=1}^N |A_{mn} - A'_{mn}|}{\sum_{n=1}^N (A_{mn} + A'_{mn})} \quad (2)$$

Where  $m$ ,  $N$ ,  $A$ , and  $A'$  are the metabolite index, the number of test datasets, the ground truth, and the estimated amplitude, respectively.

## 3 RESULTS

### 3.1 Comparison between the Quantification Result of QUEST, CNN, and WSCN for ISMRM Challenge Dataset

Figure 3 shows the comparison between different methods, namely Quest, CNN, and WSCN, for dataset DSS1, where the SNR of signals was set to 10. The result for CNN and QUEST were extracted from (Hatami et al., 2018).



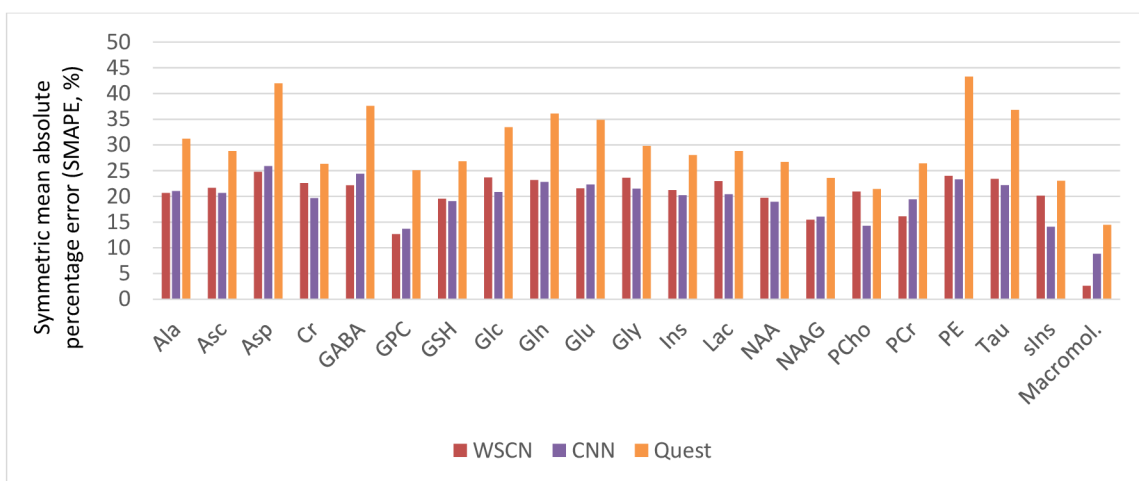


Figure 3: Comparison between SMAPEs of each metabolite for the WSCN (red), the CNN (yellow), and Quest (green).

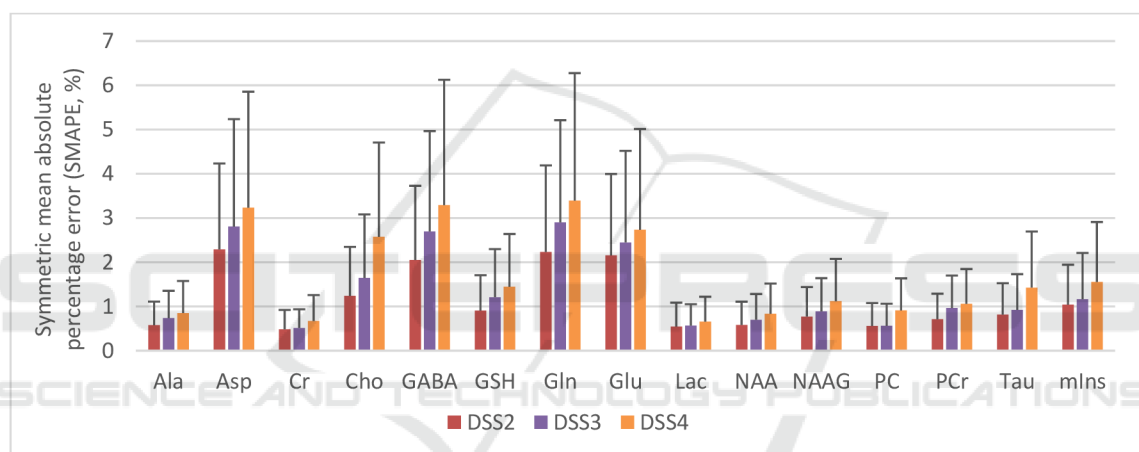


Figure 4: Comparison between SMAPEs of the concentration of all metabolites with fixed phases (DSS2), common phase varied (DSS3) and independently varied phases (DSS4) (different phase changes for different metabolites). (Test datasets, N=1000). The error bars represent the standard deviation.

### 3.2 Effect of Phase Variation and Noise on WSCN Estimation Accuracy

The performance of WSCN was evaluated on different datasets (DSS2 to DSS7) in table 1. Figure 4 shows the effect of metabolite phase variation in the signals under test. We compared the result of signals with a fixed phase, a common varied phase, and independently varied phases. The average of SMAPEs for DSS2, DSS3, and DSS4 were 1.13%, 1.38%, and 1.7%, respectively.

The results of the metabolite quantification for DSS5 (DSS3 with added noise) is shown in Figure 5. For all 15 metabolites, the average of SMAPE was  $3.46\% \pm 2.81\%$ . Asp with SMAPE of  $6.00 \pm 4.48$  and NAAG with SMAPE of  $13.20\% \pm 10.12\%$  were quantified as highest and lowest SMAPE,

respectively. The average SMAPE of DSS5 was increased by 151% compared to DSS3 (without noise).

### 3.3 Effect of Macromolecules Variation on WSCN Estimation Accuracy

Figure 6 shows a comparison between DSS6 and DSS7. In dataset DSS6, the parameters of baseline signals (11 Gaussian lines) are constant, while in DSS7, amplitudes of Gaussian lines were randomly varied in the range of  $\pm 10\%$  of their initial values. For all metabolites of DSS6 and DSS7, the average SMAPEs were  $5.92\% \pm 4.40\%$  and  $6.12\% \pm 4.55\%$ , respectively. The average SMAPE of DSS6 and DSS7 was increased by 73% compared to DSS5 (without Macromolecules inclusion).

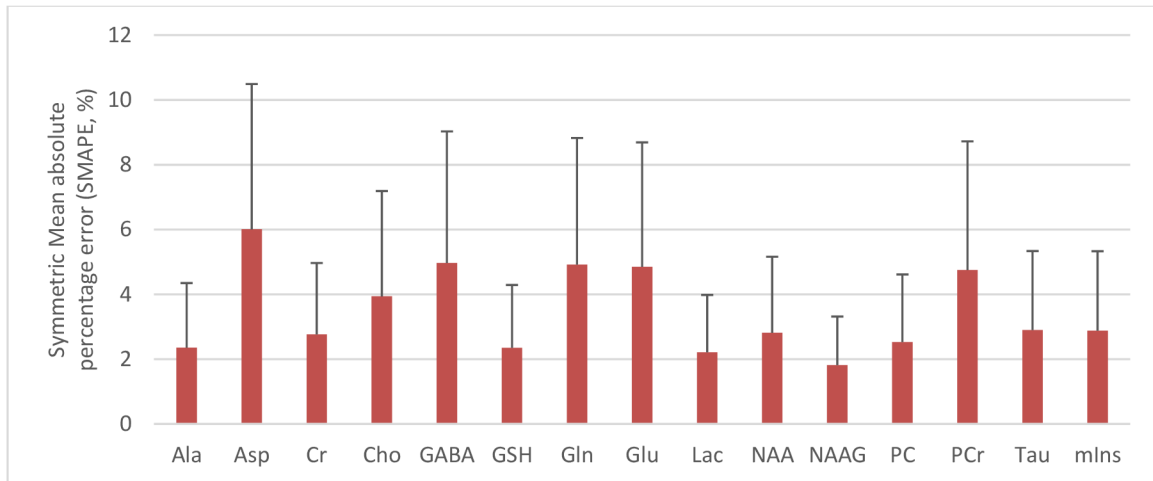


Figure 5: Symmetric mean absolute percent error (SMAPE) of the concentrations of all metabolites in dataset DSS5, which contains noisy signal (N = 5000). The error bars represent the standard deviation.

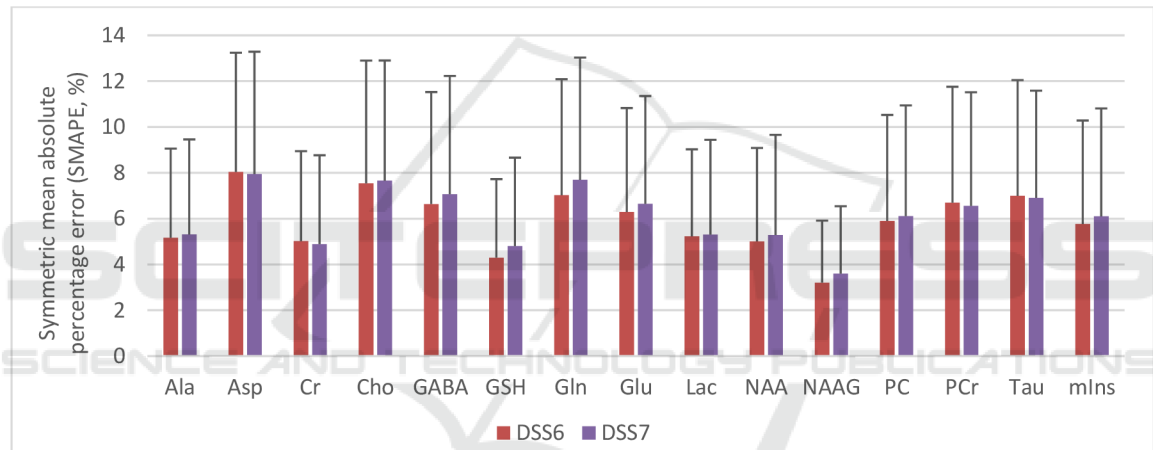


Figure 6: Comparison between SMAPEs of the concentration of all metabolites in dataset DSS6, and DSS7 (Test datasets, N=1000). In DSS6, the amplitudes of macromolecules lines were constant. In contrary, the amplitudes were varying within  $\pm 10\%$  of initial range in DSS7. Both datasets are noisy and with common phase changing.

## 4 CONCLUSIONS

The aim of MRS signal quantification is to estimate the amplitudes/areas (in time/frequency domain) of different metabolites in the signal. The estimated amplitudes/areas then can be converted to meaningful numbers as the concentration of metabolites. The conventional and widely used approach is to estimate amplitudes of single sinusoids (areas of single peaks) in MRS signal or to estimate the amplitudes (areas) of whole metabolite signals (spectra). In the former approach, the model is fitted to data using non-linear least-squares analysis; the latter approach uses a basis set of metabolite profiles in the model function and uses a semi-parametric fitting technique. The oldest

method, so-called peak integration, is calculating peaks area in a selected frequency interval. Nonetheless, using these approaches for quantification is challenging (Stagg & Rothman, 2014).

On the other hand, the quantification of the MRS signal using deep learning has attracted huge interest in recent years (Chen et al., 2020). DL can detect important features in the MRS signal and subsequently determine a non-linear mapping between these features and the outputs, which can be the concentrations of the metabolites. The most widely used DL approach for quantification is CNN. Nevertheless, this approach has drawbacks, for example, poor understanding of the CNN architecture and hyper-parameters for MRS, expensive and time-

consuming computation, and the need of a big dataset for CNN training (Bruna and Mallat, 2013).

These shortages motivated us to develop a deep network for MRS signal quantification, which can be fast, well-understood, and works with a small dataset of training samples. For this purpose, we used a WSCN.

In every DL task, determining the proper input and output of the network is an important step. In our study, the input is an FID, i.e., time-domain signal, and the network estimates amplitudes of the first points of metabolite signals (what corresponds to areas under metabolite signals in metabolite spectra). In this work, we demonstrated that the use of the wavelet scattering network could achieve better results than the semi-parametric fitting technique QUEST and similar results as the computationally more demanding CNN (Figure 3).

It is known that the accuracy of estimation in the peak integration approaches is influenced by phases of peaks (Stagg & Rothman, 2014), and that phase should be included in the model as one of the unknown parameters. Therefore, we also investigated whether WSCN is capable of estimating amplitudes of metabolites in case that metabolite phases change. It resulted in an increase in the complexity of the model, but WSCN proved to have the capability of handling this task. Figure 4 shows the WSCN can quantify signals with common varied phases (with SMAPE of 1.38%) as well as signals without fixed phases (with SMAPE of 1.13%). The average of MAPEs for DSS4 is increased by 36% and 17% compared to DSS2 and DSS3, respectively. It indicates that quantification can be moderately harder for a dataset with independently varied phases.

Another source of error in quantification are macromolecular signals, which stem in macromolecules present in the tissue under investigation. In conventional quantification approaches, macromolecule signals can either be removed in the preprocessing step or modeled in the quantitation step. However, the risk of errors will be increased and accumulated in fitting error in the former approach, and therefore the latter approach is recommended. However, macromolecule signals often overlap with metabolite components, for which DL can be a method of choice for disentangling. As we showed in Figure 3, the WSCN could estimate macromolecules better than other approaches. Later in this study, we modeled the macromolecules signal as a set of Gaussian lines using parameters (like linewidth, frequency) measured using the inversion-recovery recovery (Osorio-Garcia et al., 2011). Figure 6 demonstrates that the WSCN showed nearly the same

error for signals with randomly varied macromolecule lines and signals with fixed macromolecule lines. This could indicate that despite the changing of background signals parameters, the WSCN is stable against nuisance components in MRS, such as macromolecules. Additional research should be done however with simulated signals that will imitate *in-vivo* data.

To compare the learning times of both networks, i.e., Hatami et al.'s CNN and our WSCN, we rebuilt their CNN and fed both networks with the DSS5 dataset, and ran both networks in the earlier mentioned system. Our proposed approach is estimated to be 45 times faster than Hatami et al.'s approach (the WSCN's learning time was 5 min 40 sec precisely and the CNN's learning time was 268 min). The WSCN showed that it could be faster than the CNN due to using fixed-size filters and less parameter optimization.

It should be noted that even though deep learning showed promising results in areas like speech recognition and image processing (Chen et al., 2020), this study is one of the very initial steps in the application of DL in MRS and more studies are needed for proving DL suitability for *in-vivo* spectroscopy. Below some of the limitations and open issues are addressed:

1. In this study, we only quantified simulated data. The amplitudes of metabolites in our simulated data did not imitate the metabolite concentrations in *in-vivo* data. Quantification of simulated data with concentrations close to *in-vivo* data should be investigated as the next step together with data acquired from a phantom.
2. Real MRS data is influenced by numerous factors such as voxel size, voxel placement, radiofrequency (RF) coil sensitivity, receiver gain, and other experimental factors. Further research must take all factors into account.
3. A potential application of our proposed approach is the quantification of MRSI data, where a fast method is needed for quantification of a set of MRS signal. Learning a network and using it for only a single voxel may not be efficient as using it for a set of signals.

## ACKNOWLEDGEMENTS

This research was supported by European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 813120 (INSPIRE-MED) and by European Regional Development Funds under project "National

infrastructure for biological and medical imaging" of the Ministry of Education, Youth and Sports of the CR (No. CZ.02.1.01/0.0/0.0/16\_013/0001775).

## REFERENCES

- Alaskar, H. (2019). *Deep Learning-based Model Architecture for Time-Frequency Images Analysis*. January 2018.
- Andén, J., & Mallat, S. (2014). Deep scattering spectrum. *IEEE Transactions on Signal Processing*, 62(16), 4114–4128.
- Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1872–1886.
- Chen, D., Wang, Z., Guo, D., Orekhov, V., & Qu, X. (2020). Review and Prospect: Deep Learning in Nuclear Magnetic Resonance Spectroscopy. *Chemistry - A European Journal*, 26(46), 10391–10401.
- Graveron-Demilly, D. (2014). Quantification in magnetic resonance spectroscopy based on semi-parametric approaches. In *Magnetic Resonance Materials in Physics, Biology and Medicine* (Vol. 27, Issue 2, pp. 113–130). Springer Verlag.
- Hatami, N., Lyon, B., & Etienne, U. (2018). *Magnetic Resonance Spectroscopy Quantification using Deep Learning*.
- ISMRM. (2016). *MRS Fitting Challenge* ([ismrm.org/Workshops/Spectroscopy16/mrs\\_fitting\\_challenge/](http://ismrm.org/Workshops/Spectroscopy16/mrs_fitting_challenge/)).
- Kreis, R., & Kyathanahally, S. P. (2018). *Deep Learning Approaches for Detection and Removal of Ghosting Artifacts in MR Spectroscopy*. 863, 851–863.
- Lee, H. H., & Kim, H. (2019). Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. *Magnetic Resonance in Medicine*, 82(1), 33–48.
- Opstad, K. S., Bell, B. A., Griffiths, J. R., & Howe, F. A. (2008). Toward accurate quantification of metabolites, lipids, and macromolecules in HRMAS spectra of human brain tumor biopsies using LCModel. *Magnetic Resonance in Medicine*, 60(5).
- Osorio-Garcia, M. I., Sima, D. M., Nielsen, F. U., Dresselaers, T., Van Leuven, F., Himmelreich, U., & Van Huffel, S. (2011). Quantification of in vivo 1H magnetic resonance spectroscopy signals with baseline and lineshape estimation. *Measurement Science and Technology*, 22(11), 1–17.
- Pouillet, J. B., Sima, D. M., Simonetti, A. W., De Neuter, B., Vanhamme, L., Lemmerling, P., & Van Huffel, S. (2007). An automated quantification of short echo time MRS spectra in an open source software environment: AQSES. *NMR in Biomedicine*, 20(5), 493–504.
- Pouillet, J. B., Sima, D. M., & Van Huffel, S. (2008). MRS signal quantitation: A review of time- and frequency-domain methods. *Journal of Magnetic Resonance*, 195(2), 134–144.
- Robin A. de Graaf. (2019). In Vivo NMR Spectroscopy. In *In Vivo NMR Spectroscopy* (pp. 1–42).
- Stagg, C., & Rothman, D. (2014). *Magnetic Resonance Spectroscopy: Tools for Neuroscience Research*.
- Starčuk, Z., & Starčuková, J. (2017). Quantum-mechanical simulations for in vivo MR spectroscopy: Principles and possibilities demonstrated with the program NMRScopeB. *Analytical Biochemistry*, 529.
- Stefan, D., Cesare, F. Di, Andrasescu, A., Popa, E., Lazariu, 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. *Measurement Science and Technology*, 20(10).
- Suvichakorn, A., Ratiney, H., Bucur, A., Cavassila, S., & Antoine, J. P. (2008). Quantification method using the Morlet wavelet for Magnetic Resonance Spectroscopic signals with macromolecular contamination. *Proceedings of the 30th Annual International Conf. of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology"*, 2681–2684.
- Van Der Graaf, M. (2010). In vivo magnetic resonance spectroscopy: Basic methodology and clinical applications. *European Biophysics Journal*, 39(4), 527–540.



## RESEARCH ARTICLE

# Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias

Rudy Rizzo<sup>1,2,3,4</sup> | Martyna Dziadosz<sup>1,2,3,4</sup> | Sreenath P. Kyathanahally<sup>5</sup> | Amirmohammad Shamaei<sup>6,7</sup> | Roland Kreis<sup>1,2,4</sup>

<sup>1</sup> MR Methodology, Department for Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland

<sup>2</sup> Department for Biomedical Research, University of Bern, Bern, Switzerland

<sup>3</sup> Translational Imaging Center (TIC), Swiss Institute for Translational and Entrepreneurial Medicine, Bern, Switzerland

<sup>4</sup> Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland

<sup>5</sup> Department of System Analysis, Integrated Assessment and Modelling, Data Science for Environmental Research Group, EAWAG, Dübendorf, Switzerland

<sup>6</sup> Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic, Brno, Czech Republic

<sup>7</sup> Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic

## Correspondence

Roland Kreis, MR Methodology, Department for Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland.  
Email: [roland.kreis@insel.ch](mailto:roland.kreis@insel.ch)

## Funding information

H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 813120; Nvidia; Swiss National Science Foundation, Grant/Award Number: 320030-175984

**Purpose:** The aims of this work are (1) to explore deep learning (DL) architectures, spectroscopic input types, and learning designs toward optimal quantification in MR spectroscopy of simulated pathological spectra; and (2) to demonstrate accuracy and precision of DL predictions in view of inherent bias toward the training distribution.

**Methods:** Simulated 1D spectra and 2D spectrograms that mimic an extensive range of pathological in vivo conditions are used to train and test 24 different DL architectures. Active learning through altered training and testing data distributions is probed to optimize quantification performance. Ensembles of networks are explored to improve DL robustness and reduce the variance of estimates. A set of scores compares performances of DL predictions and traditional model fitting (MF).

**Results:** Ensembles of heterogeneous networks that combine 1D frequency-domain and 2D time-frequency domain spectrograms as input perform best. Dataset augmentation with active learning can improve performance, but gains are limited. MF is more accurate, although DL appears to be more precise at low SNR. However, this overall improved precision originates from a strong bias for cases with high uncertainty toward the dataset the network has been trained with, tending toward its average value.

**Conclusion:** MF mostly performs better compared to the faster DL approach. Potential intrinsic biases on training sets are dangerous in a clinical context that requires the algorithm to be unbiased to outliers (i.e., pathological data). Active learning and ensemble of networks are good strategies to improve prediction performances. However, data quality (sufficient SNR) has proven as a bottleneck for adequate unbiased performance—like in the case of MF.

## KEYWORDS

active learning, bias, deep learning, ensemble of networks, model fitting, magnetic resonance spectroscopy, quantification

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Magnetic Resonance in Medicine* published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

## 1 | INTRODUCTION

MR Spectroscopy (MRS) provides a noninvasive means for extracting biochemical profiles from in vivo tissues. Metabolites are encoded with different resonance frequency patterns, and their concentrations are directly proportional to the signal amplitude.<sup>1,2</sup> Metabolite quantification is traditionally based on model fitting (MF), where a parameterized model function is optimized to explain the data via a minimization algorithm. Metabolite parameters are usually estimated by a nonlinear least-squares fit (either in time or frequency domain) using a known basis set of the metabolite signals.<sup>3</sup> However, despite various proposed fitting methods,<sup>3-7</sup> robust, reliable, and accurate quantification of metabolite concentrations remains challenging.<sup>8</sup> The major problems influencing the quantitative outcome are: (1) overlapping spectral patterns of metabolites, (2) low SNR, and (3) unknown background signals and line shape (no exact prior knowledge). Therefore, the problem is ill-posed, and current methods address it with different regularizations and constraint strategies (e.g., parameter bounds, penalizations, choice of the algorithm), with discrepancies in the results from one method to another.<sup>9</sup>

Supervised deep learning (DL) utilizes neural networks to discover essential features embedded in large data sets and to determine complex nonlinear mappings between inputs and outputs.<sup>10</sup> Thus, DL does not require any prior knowledge or traditional assumptions. Given the success of the method in different areas,<sup>10-14</sup> DL has been introduced into MRS as an alternative to conventional methods.<sup>15-22</sup> Quantification of MRS datasets has been explored as follows: (1) DL algorithms identify datasets' features and either help reduce the parameter space dimension or set reliable starting conditions for the fit (i.e., combining knowledge on the physics with DL). It showed rapid spectral fitting of a whole-brain MRSI datasets.<sup>23</sup> (2) Convolutional neural networks (CNNs) have been deployed to investigate combinations of spectral input of edited human brain MRS, which showed improved accuracy of straight metabolite quantitation when compared to traditional MF techniques.<sup>24</sup> (3) Regression CNNs have been used to mine the real part of rat brain spectra to predict highly resolved metabolite basis set spectra with intensities proportional to the concentrations of the contributions,<sup>17</sup> with results comparable to traditional MF approaches and showing readiness for (pre)clinical applications.<sup>22</sup> (4) Targeting localized correlated spectroscopy (L-COSY) datasets, DL algorithms have reported faster data reconstruction and quantification compared to alternative acceleration techniques.<sup>16</sup>

Nevertheless, despite the reported equivalence in quantitation performance compared to traditional MF,<sup>14,17,22,23</sup> questions arise concerning the robustness of

DL algorithms. A robust use within a clinical MRS context requires the algorithm to be unbiased also for pathological spectra. In imaging, DL has shown excellent performance for classification or segmentation tasks but may suffer from inherent weaknesses in subsets of representative outlier samples.<sup>11,25</sup> DL architectures for MRS quantitation have mostly been investigated for sample distributions of near-healthy spectral metabolite content. Hence, it can be suspected that high accuracy and precision are mainly found when DL is deployed for new entries of similar near-normal types. However, inaccurate estimates may result for tests with atypical datasets.<sup>26</sup> Here, strongly variable metabolite concentrations that vary uniformly and independently over the entire plausible parameter space are used in the training set. This mimics the full range from healthy to strongly pathological spectra, that is, the full complexity of a clinical setup.

MRS signals are acquired in time domain but viewed in frequency domain. Traditional MF works in either of the two equivalent domains, and fit packages may allow the user to switch from one to the other for fitting and viewing. However, DL architectures for MRS quantification have mainly explored the frequency domain, mostly motivated by the reduced overlap between the constituting metabolite signals. Spectrograms<sup>18</sup> present an extension into a simultaneous time/frequency domain representation and offer a 2D signal support that matches the input format for the original usage of CNN algorithms in computer vision. This work introduces a dedicated high-resolution spectrogram calculation focusing on signal-rich areas in both domains to be used as input for different CNN architectures. They are compared to other inputs and networks, inspired by previous MRS publications. Specifically, 24 network designs are investigated with differing input-output dataset types with a combined focus on depth (i.e., number of layers) and width (i.e., number of nodes/kernels) of the networks. This focus was motivated by the fact that the exploitation of spectrograms in deep learning has shown top-notch performance for speech and audio processing when deploying architectures with few layers and large convolutional kernels.<sup>27-29</sup> Moreover, wide and shallow networks are more suitable to detect simple and small but fine-grained features. In addition, they are easier and faster to train.<sup>30</sup> Network linearity (i.e., activation function) and locality (i.e., kernel size) are also investigated.

Besides investigating multiple architectures and input formats, two established main strategies for improving the outcome of predictions are also explored: *active learning*<sup>31</sup> (data augmentation for critical types of spectra) and *ensemble learning*<sup>32,33</sup> (combination of outputs from multiple architectures).

*Active learning* can improve labeling efficiency,<sup>31,34,35</sup> where the learning algorithm can interactively select a

subset of examples that needs to be labeled. This is an iterative process where (1) the algorithm selects a subset of examples; (2) the subset is provided with labels; and (3) the learning method is updated with the new data.<sup>36</sup> *Uncertainty sampling*<sup>37</sup> is a specific strategy used in active learning that prioritizes selecting examples whose predictions are more uncertain (i.e., targeted data augmentation). Because these cases are usually close to the class separation boundaries, they contain most of the information needed to separate different classes.<sup>38,39</sup> In different applications, uncertainty sampling has been shown to improve the effectiveness of the labeling procedure significantly.<sup>34,35,37,40</sup>

DL algorithms are sensitive to the specifics of the training.<sup>41</sup> Hence, they usually find a different set of weights each time they are trained, producing different predictions.<sup>10</sup> A successful approach for reducing the variance is to train multiple networks instead of one and combine their predictions.<sup>41</sup> This is called *ensemble learning*, where the model generalization is maintained, but predictions improve compared to any of the single models.<sup>33</sup> From a range of different techniques,<sup>42–44</sup> here, *stacking of models* is implemented.<sup>32</sup>

To evaluate pros and cons of all these approaches, in silico ground truth (GT) knowledge is used (and hence no in vivo data was included in this evaluation) to assess performances via a dedicated set of metrics based on bias and SD. The CNN-predicted distributions of concentration are then compared to those from traditional MF. Furthermore, to emphasize the analysis at the core of the quantification task, the focus is placed on an idealized simulated setting with typical single-voxel spectra that have been pre-processed to eliminate phase as well as frequency drifts.<sup>3</sup> This assumption aims at (1) freeing the MF algorithm from problems with local  $\chi^2$  minima and (2) designing DL models optimized for the quantification task only.

## 2 | METHODS

### 2.1 | Simulations

This work is based on in silico simulations. A dataset of 22,500 entries was randomly split into 18,000 for training, 2000 for validation, and 2500 for testing. Larger dataset sizes are also explored, see section 2.4.

#### 2.1.1 | MR spectra

Brain spectra were simulated using actual RF pulse shapes for 16 metabolites at 3 T using *Vespa*<sup>45</sup> for a semi-LASER<sup>46</sup>

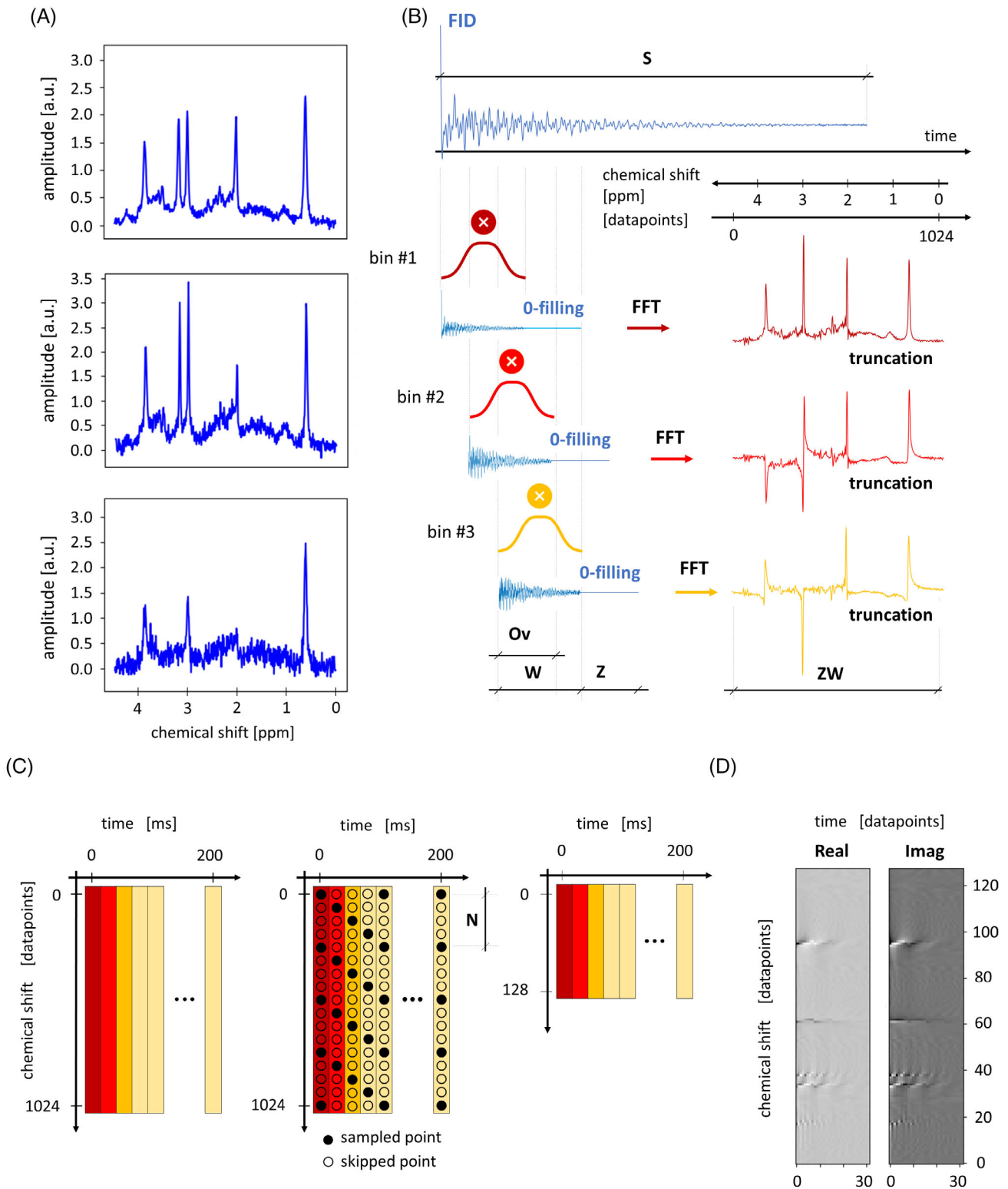
protocol with TE = 35 ms, a sampling frequency of 4 kHz, and 4096 datapoints.

Further specifics of the simulations include: (1) Voigt line shapes, (2) metabolite concentration range, (3) addition of macromolecular background signal (MMBG), (4) noise generation, and (5) spectrum or spectrogram calculation.<sup>47</sup> Metabolite concentrations vary independently and uniformly between 0 and twice a normal reference concentration for healthy human brain.<sup>1,48–50</sup> Maximal concentrations in mM units—NAA 25.8, tCr (1:1 sum of creatine + phosphocreatine spectra): 18.5, mI (myo-inositol): 14.7, Glu (glutamate): 20, Glc (glucose): 2, NAAG (N-acetylaspartylglutamate): 2.8, Gln (glutamine): 5.8, GSH (glutathione): 2, sI (sylo-inositol): 0.6, Gly (glycine): 2, Asp (aspartate): 3.5, PE (phosphoethanolamine): 3.3, Tau (taurine): 2, Lac (lactate): 1, and GABA ( $\gamma$ -aminobutyric acid): 1.8. The concentration for tCho (1:1 sum of glycerophosphorylcholine + phosphorylcholine spectra) ranges from 0 to 5 mM to mimic tumor conditions.<sup>51</sup> A constant down-scaled water reference (64.5 mM) is added at 0.5 ppm to ease quantitation. Metabolite  $T_2s$  in ms (and hence Lorentzian broadening) are fixed to reference values from literature—tCr ( $CH_2$ ): 111, tCr ( $CH_3$ ): 169, NAA ( $CH_3$ ): 289, and all other protons: 185.<sup>49,52,53,54</sup> MMBG content, shim, and SNR mimicked in vivo acquisitions and varied independently and uniformly (time-domain water referenced SNR 5–40, Gaussian shim 2–5 Hz, MMBG amplitude  $\pm 33\%$ ). The MMBG pattern was simulated as a sum of overlapping Voigt lines as reported in Refs. 49 and 55 (Figure 1A).

#### 2.1.2 | Spectrograms

A spectrogram is a complex 2D representation of a spectrum, where frequencies vary with time: Every image column represents the frequency content of a particular time portion of the FID. Time information is binned along every row of the image. It is calculated via application of a short-time Fourier transform,<sup>18</sup> where, depending on the size of the Fourier analysis window, different levels of frequency and time resolution can be achieved. A long window size modulated via zero-filling combined with a small overlap interval is chosen to increase frequency resolution and minimize the expense of time resolution (Figure 1B). Diagonal downsampling is designed to reduce the spectrogram size, keeping the original resolution grid at least as part of the time-frequency information on consecutive bins and reducing the spectrogram size (Figure 1C) to allow reasonable computation time for a CNN architecture (i.e., 128 frequency bins  $\times$  32 time bins) (Figure 1D).





**FIGURE 1** Illustration of input formats. (A) Samples of spectra, real part, view of the central 1024 points. (B) Spectrogram computation via short-time Fourier transform. Specifically, in datapoints units (corresponding to time and frequency resolution of 0.25 ms and 1 Hz, respectively):  $S = 4096$ ,  $Z = 6000$ ,  $W = 1024$ ,  $Ov = 1000$ ,  $ZW = 1024$ . Zero-filling is tuned to select the relevant part of the spectrum with  $W = 1024$  datapoints. (C) (Left) Arrangement on a 2D frame of short-time Fourier transforms over time bins. Color code reference to windows in part (B). A truncation at 32 bins (200 ms) in time domain is used to limit the matrix space, given an almost complete  $T_2$  relaxation of the FID at that point. (C) (Middle) Diagonal undersampling reduces the vertical (frequency domain) matrix size. Size reduction is about a factor  $N = 8$ . (C) (Right) Undersampled spectrogram:  $128 \times 32$  datapoints. (D) Example of constructed spectrogram matrix. FFT, fast Fourier transform; S, support of the signal; Ov, window overlap; W, Hamming window size; Z, zero filling; ZW, truncated support of zero-filled FFT.

## 2.2 | Design and training of CNN architectures

A total of 24 different CNN architectures combined with different spectroscopic input representations are compared for MRS metabolite quantification. Current state-of-the-art networks have been taken as reference models and adapted to the purpose and datasets used.

Scripts were written in Python<sup>56</sup> using Keras library<sup>57</sup> on a Tensorflow<sup>58</sup> backend. Code ran on either of three graphic-processing units (GPUs; NVIDIA [Santa Clara, USA] Titan Xp, Titan RTX, or GeForce RTX 2080 Ti) or Google [Mountain View, USA] Colaboratory.<sup>59</sup> Samples of the design are reported in Figure 2. Overall network designs are given in Table S1; Figures S1, S2, S3, S4, S5; and Text S1.

### 2.2.1 | Architectures for straight numeric quantification of concentrations

A total of 22 architectures were fed with 1D (spectra) or 2D (spectrograms) input and mapped as output a vector of 17 normalized concentrations (i.e., in [0–1] interval) of 16 metabolites and the water reference, as listed in Table S1. Networks fed with 1D input exploit one channel with truncated spectra of 1024 datapoints from –0.5 to +6 ppm with concatenated real and imaginary parts (i.e.,  $2048 \times 1 \times 1$  datapoints, Figure 2A). Networks fed with 2D input can either be configured in two channels (real and imaginary components of the spectrogram,  $32 \times 128 \times 2$  datapoints) or one channel (real and imaginary components concatenated,  $64 \times 128 \times 1$  datapoints, Figure 2B).

Five networks receive 1D input: two deep convolutional neural networks (*DeepNet*),<sup>60</sup> two residual networks (*ResNet*)<sup>61</sup> and one inception network (*InceptionNet*).<sup>62–64</sup>

This work investigates deep and shallow architectures either exploiting large or small convolutional kernel sizes. A total of 10 networks receive two-channel spectrograms as input. Given the limited size of the input FOV, the architecture is limited to be shallow (i.e., pooling operations to downsampling features directly following a convolutional layer are limited). However, a deeper architecture with multiple convolutional operations with sparse pooling is also compared. A further comparison is performed regarding the optimal activation function, comparing batch normalization + rectified linear unit (ReLU) versus exponential linear unit (ELU).<sup>65,66</sup> Seven networks receive one-channel spectrograms as input. With this configuration, deeper architectures are explored: two *DeepNets*, four *ResNets*, and one *InceptionNet*.

Architectures are analyzed either in a preconfigured parameter state or in a parameter space that had been optimized via Bayesian hyperparameterization.<sup>67</sup> The optimization procedure is given in Text S1. In addition, to limit biases around zero for small concentrations,<sup>68</sup> all network designs are characterized by a final layer with linear activation, allowing the prediction of negative concentrations.

### 2.2.2 | Architectures for estimation of metabolite base spectra

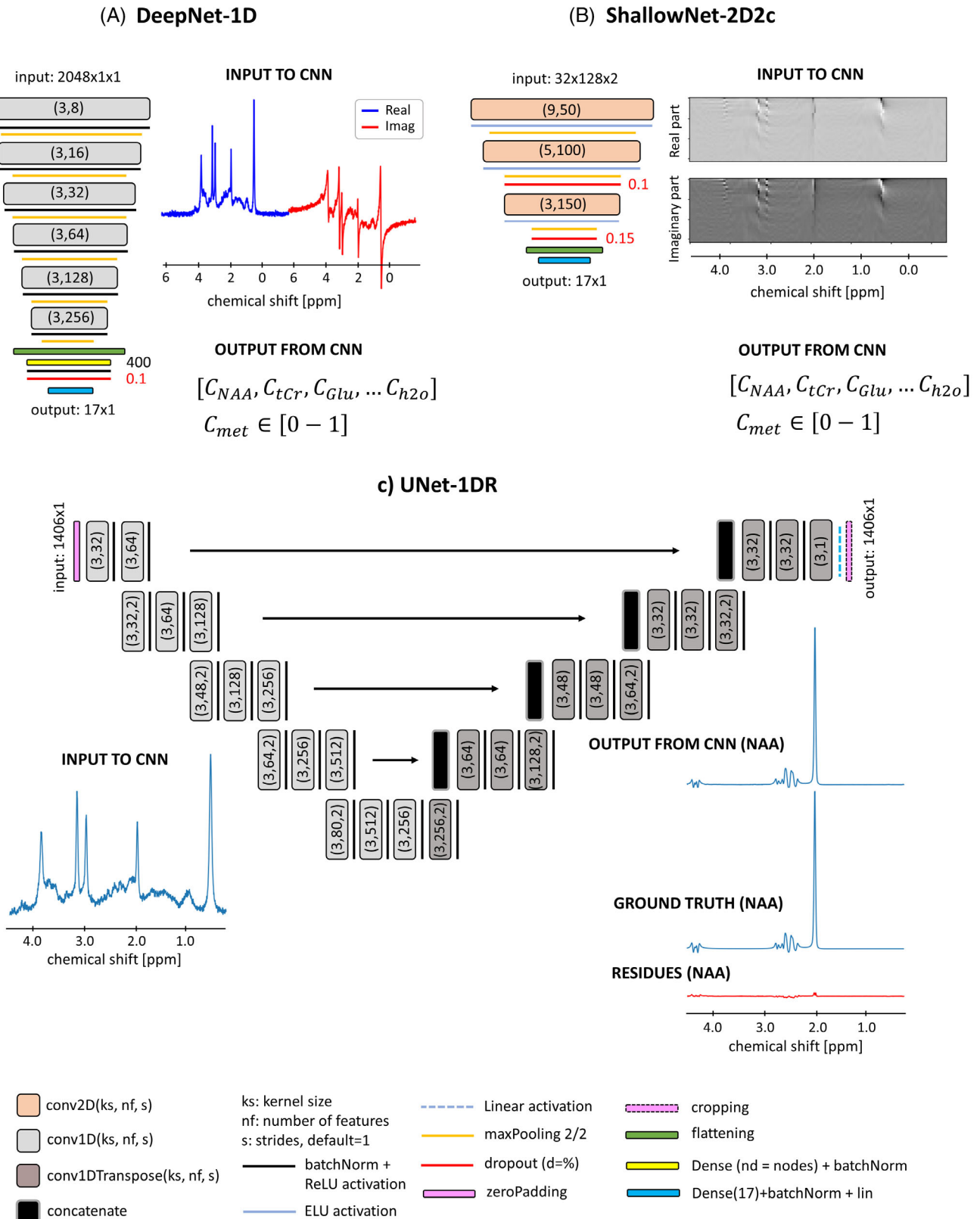
1D input (real part only, 0–4.7 ppm,  $1406 \times 1 \times 1$  datapoints after zero-filling of original FID) was used to input and output to/from the CNNs. U-Net architectures<sup>69</sup> analogous to those of Ref. 22 are implemented here to map the ideal high-resolved noiseless base spectrum of a target metabolite as output. CNNs are trained one by one for each metabolite such that each CNN filters out signals only from the designated target metabolite. A base U-Net design (Figure 2C) is optimized for individual metabolites as follows:

1. *UNet-1DR-hp*: A total of 17 different networks with the same base architecture but adapted weights for each metabolite;
2. *Unet-1DR-hp-met*: A total of 17 different networks with adapted Bayesian-optimized architecture and weights for each metabolite.

Configurations are reported in Figure S5. First, metabolite concentrations are evaluated by feeding an input spectrum to the 17 metabolite-specific CNNs. Integration of the predicted metabolite base spectrum is then referenced to the integrated water reference to produce concentrations for a fully automated quantification pipeline.<sup>22</sup>

### 2.2.3 | Training

Training and validation sets were randomly assigned for training the CNN on a maximum of 200 epochs with batch normalization of 50. The adaptive moment estimation algorithm (ADAM)<sup>70</sup> was used with dedicated starting learning rates for each network.<sup>71,72</sup> The loss function was the mean-squared error (MSE). Visualization of training and validation loss over epochs combined with implementing an early-stopping criterion monitoring minimization of validation loss with patience = 10 has been used for tuning the network parameter space.<sup>57</sup> Training time and test loss function are listed in Table S1.



**FIGURE 2** Examples of three CNN structures and schematic input–output relationships. (A) and (B) depict architectures for straight quantification, with metabolites relative concentrations as output. (C) depicts a U-Net architecture similar to what was proposed in Ref. 22 for NAA basis set prediction. Input details: (A) *Deep neural network* with 1D-spectral input from concatenated real and imaginary parts (-1D). (B) *Shallow neural network* with 2D-spectral input from two-channel spectrograms (-2D2c). (C) U-Net architecture fed with only the real part of a spectroscopic input (-1DR). CNN, convolutional neural network.



## 2.2.4 | Evaluation

Regression plots mapping GT concentrations versus CNN predicted concentrations from the whole test set are taken as indicators of the network's prediction performance. Four scores are defined:

- $a$  (slope of the regression line): must be close to 1 for ideal mapping of concentrations over the whole range of simulated metabolite content;
- $q$  (intercept of the regression line, mM): must be close to 0 to minimize prediction offsets/biases;
- $R^2$  (coefficient of determination): must be close to 1 to assess full model explanation of the variability of the data;
- $\sigma$  (RMS error [RMSE] of prediction vs. GT, mM): as low as possible. However, expected to be comparable to Cramer Rao Lower Bounds (CRLBs) from MF.<sup>73</sup>

To easily compare different networks and input setups quantitatively in the Results section, these scores or combinations thereof have been used. The combinations are referred to as *concise scores*:  $a \cdot R^2$  as measure of linearity,  $\sigma$  to compare with CRLBs.  $q$  was excluded because it is mostly negligible.

## 2.3 | Influence of inclusion of water reference peak

For the evaluation of the potential benefit of including a water reference peak, two slightly different *ShallowNet-2D2c-hp* networks are compared. *Network A* outputs 17 neurons (16 metabolites and water), whereas *network B* outputs 16 neurons only (no water output). Two adapted datasets are used for the investigation, one with (*dataset A*), and one without (*dataset B*) downscaled water reference at 0.5 ppm. Metabolite concentrations are calculated for both cases (assuming known water content in case A). Networks have been independently trained

five times to monitor network variability over multiple trainings.

## 2.4 | Active learning and dataset size

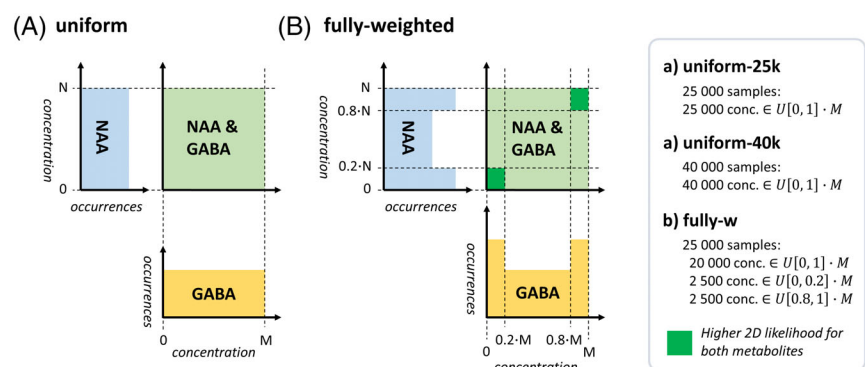
In this part, data augmentation techniques to smartly generate training sets are investigated. Subsets with 5000 new entries of the dataset where predictions scored worst are defined: specific subsets of spectrally weakly represented metabolites in either very low or very high concentrations and spectra with low SNR. New weighted datasets of 25,000 entries (20,000 training – 5000 validation set) or 40,000 entries (35,000 training – 5000 validation set) are generated (example in Figure 3, full description in Figure S6). Datasets with matching size and the testing set are kept unchanged from the previous simulation, thus with uniformly distributed concentrations and SNR. *ShallowNet-2D2c-hp* is selected as architecture and trained 10 times with a given augmented training set to minimize training variance.

Complementarily, given the network trained on a uniform span of concentrations, active learning is investigated in the testing phase on three different test sets where concentrations are clipped to a progressively smaller range of 20%–80%, 20%–80% with SNR >20, and 40%–60% concentration range relative to the training set.

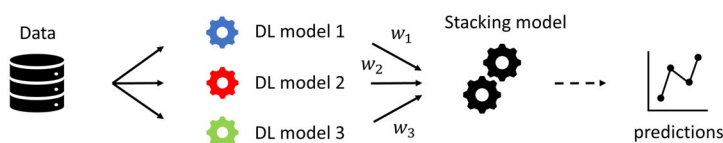
## 2.5 | Ensemble of networks

In this section, ensembles of networks are implemented via *stacking of models*.<sup>32</sup> This consists of designing a DL architecture called *stacking model* (a multilayer perceptron (MLP) with two hidden layers) is selected for this case) that will take as input the combination of a given number of independently pretrained models. The stacking model aims at weighting predictions from single models. It is trained using the same training and validation sets

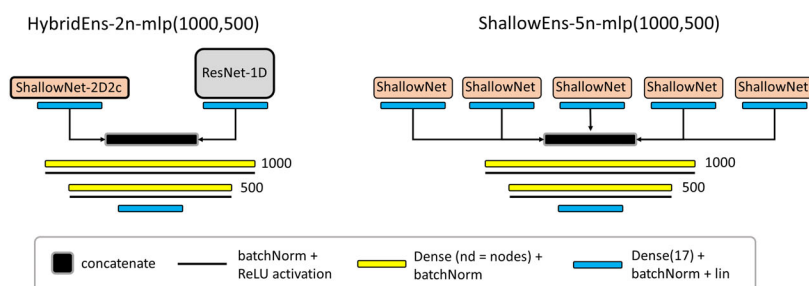
**FIGURE 3** Examples of dataset augmentation techniques representing sample distributions for two metabolites (NAA and GABA). (A) Dataset size increment with uniformly distributed concentrations. (B) Active learning weighted on higher occurrences of low and high concentrations for all metabolites. GABA,  $\gamma$ -aminobutyric acid.



## (A) Ensemble of networks via Stacking Model



## (B) Example of Stacking Models



**FIGURE 4** Illustration of ensemble learning. (A) Stacking model concept. (B) Examples of considered models: the stacking model consists of the two-layer MLPs (i.e., first layer with 1000 neurons, second layer with 500 neurons). *HybridEns*: an ensemble of two different networks ( $-2n$ ). In this study, *ShallowNet-2D2c* and *ResNet-1D* are combined with two or 10 networks. *ShallowEns*: an ensemble of five different networks ( $-5n$ ) of the same type, specifically *ShallowNet-2D2c*. *HybridEns*, hybrid ensemble; MLP, multi-layer perceptron; ResNet, residual network; *ShallowNet*, shallow network.

used to train single models while keeping the weights of the pretrained input models fixed. Three different ensembles are investigated: *ShallowEns-5n* groups five identical *ShallowNet-2D2c-hp* architectures, whereas *HybridEns* tests heterogeneous inputs grouping either two or 10 different networks (*ShallowNet-2D2c-hp* and *ResNet-1D-hp*) (Figure 4).

## 2.6 | Model fitting

Spectra are fitted using FiTAID<sup>7</sup> given its top performance in the ISMRM fitting challenge<sup>9</sup> and to be expected for the spectra as used in the current setup (in particular, without undefined spurious baseline). The model consists of a linear combination of the metabolite base spectra with Voigt lineshape, where the Lorentzian component was kept fixed at the known GT value. The areas of the metabolites are restricted in a range corresponding to  $[-0.5 + 2.5 \mu]$ , where  $\mu$  is the average concentration in the testing set distribution (i.e., the normal tissue content). These bounds mimic the effective boundaries of the DL algorithms. CRLBs are used as a precision measure<sup>74</sup> and are considered for three subgroups of the testing set (high [SNR > 28.4], medium [ $16.7 < \text{SNR} < 28.4$ ], and low [SNR < 16.7] relative SNR, respectively).

## 3 | RESULTS

### 3.1 | S1Metabolite quantification referenced to the downscaled water peak

As illustrated for three different networks, Figure 5 shows that CNN predictions perform better if the spectra are

referenced to a downscaled water peak: Regression slope  $a$  and  $R^2$  are closer to 1;  $\sigma$  is appreciably lower. Moreover, the spread of the scores is on average reduced, displaying improved stability over multiple trainings. Extended results are presented in Figures S7 and S8.

### 3.2 | Network design

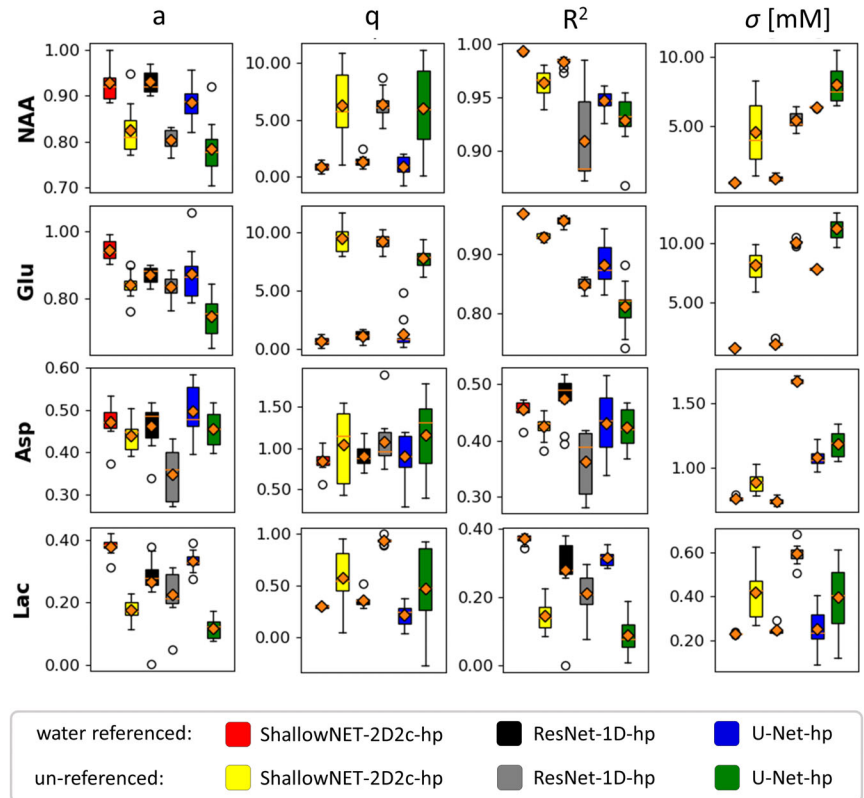
Figure 6 reports CNN predictions versus GT values of a *ResNet-1D-hp* architecture for nine metabolites (see Figures S9 and S10 for extended results on 16 metabolites or different CNN architecture). Distributions of GT and predicted values are displayed for the test set (as for all results). Predictions relate very well to the GT for well-represented metabolites (top row). However, for metabolites with lower relative SNR, predicted distributions of concentrations tend to be less uniform and are biased toward average values of the GT distributions. Thus, concentrations at distribution boundaries are systematically mispredicted, particularly for low SNR. This is reflected in lower  $a$  and  $R^2$  values and higher  $\sigma$ . Figures S11 and S12 include a comparison of multiple networks via bar graphs (which are ill-suited to express the systematic bias) and a plot of distributions of predictions.

The performance of all networks and fitting models for nine metabolites is reported in Figure 7 via a 2D plot of the concise scores  $a \cdot R^2$  and  $\sigma$  (see Figure S13 for extended results on 16 metabolites). Top performance corresponds to the top-left corner where  $a \cdot R^2$  approaches 1 and  $\sigma$  is low. Metabolites can roughly be divided into three groups:

1. *Well-represented* metabolites: NAA, tCho, tCr, mI, Glu with averaged DL scores  $a \cdot R^2 > 0.80$  and  $\sigma < 15\%$ , as well as MF scores  $a \cdot R^2 > 0.95$  and  $\sigma < 10\%$ ;



**FIGURE 5** Boxplot statistics of the prediction scores for four metabolites showing the effect of water referencing. Results reported for *ShallowNET-2D2c-hp*, *ResNet-1D-hp*, and *U-Net-hp* trained and tested on datasets with (red, black, or blue) and without (yellow, gray, or green) water reference (mean values plotted in orange). On average, water referencing yields better performance with higher coefficients  $a$  and  $R^2$  as well as lower offset  $q$  and lower RMSE  $\sigma$ . RMSE, RMS error.



2. *Medium-represented* metabolites: Glc, NAAG, Gln, GSH with averaged DL scores  $0.50 < a \cdot R^2 < 0.75$  and  $20\% < \sigma < 35\%$ , as well as MF scores  $0.75 < a \cdot R^2 < 0.90$  and  $15\% < \sigma < 35\%$ ;
3. *Weakly represented* metabolites: sI, Gly, Asp, PE, Tau, Lac, GABA with averaged DL scores  $a \cdot R^2 < 0.40$  and average  $\sigma > 35\%$ , as well as MF scores  $a \cdot R^2 < 0.65$  and  $\sigma > 35\%$ .

Overall, multiple DL networks perform similarly, but some general differences are noteworthy. Optimized spectrogram representation via two channels combined with a shallow architecture (i.e., dark blue squares) is found to be well suited for MRS quantification, showing mostly better performances than alternative deeper designs (i.e., light blue, pink, and gray squares), with one-channel designs (diamonds) or 1D spectra as signal representation (circles). Benefits are evident for medium and weakly represented metabolites. Performances of direct quantification and two-step quantification via base spectrum prediction followed by integration (stars) are similar. MF is found superior to DL for all medium- and weakly represented metabolites with significant average improvements for  $a \cdot R^2$ . However,  $\sigma$  tends to be higher for many cases. A more detailed presentation of performance is given in Figures S14 and Text S2.

Figure 8 displays plots of prediction errors (i.e.,  $\Delta = \text{prediction} - \text{GT}$ ) and their spread  $\sigma$  as a function

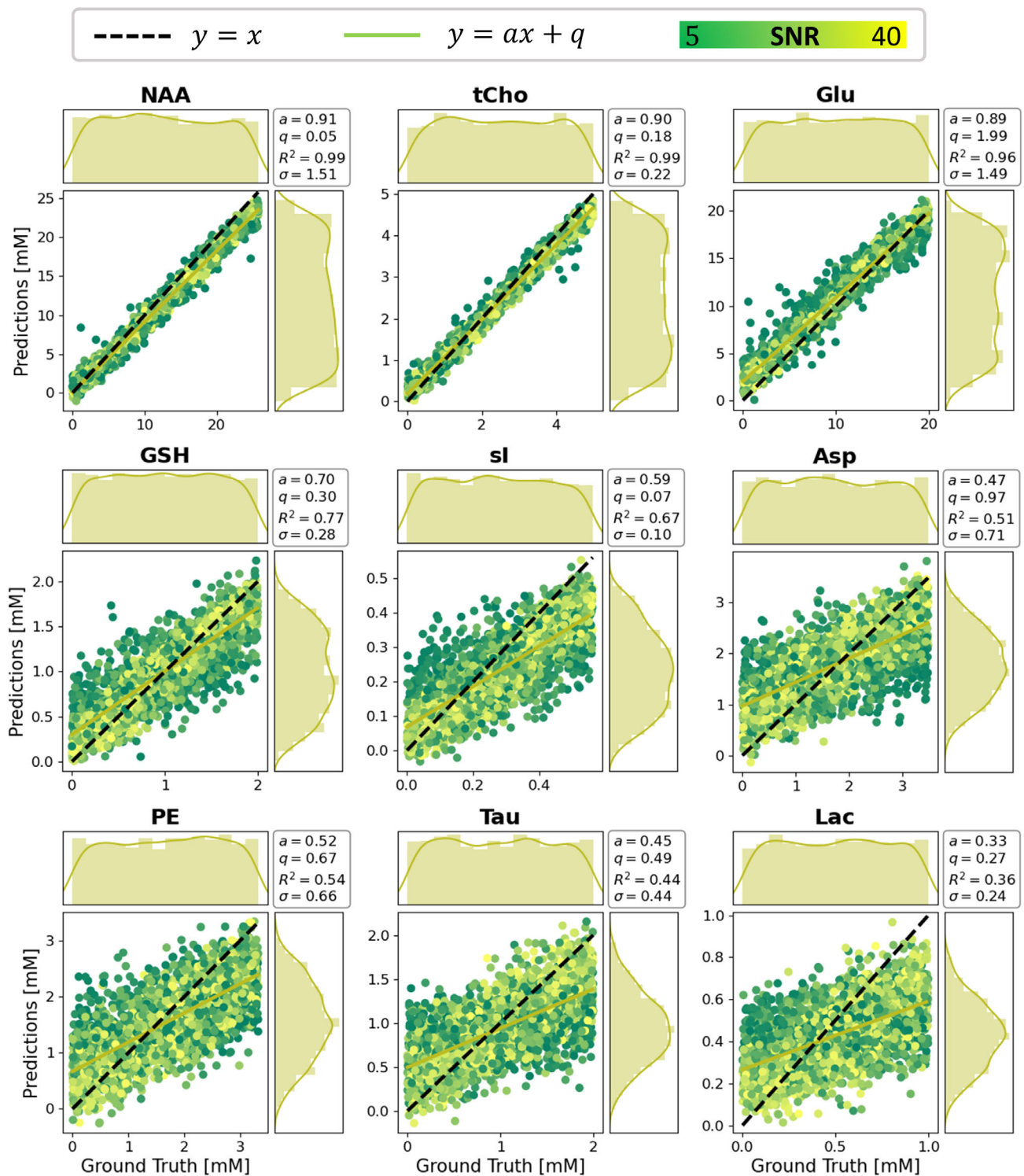
of SNR and shim for tCho, NAAG, and sI. Prediction uncertainties increase with noise level approximately linearly with  $1/\text{SNR}$  and reach a plateau for weakly represented metabolites when the spread represents essentially the whole training range. No dependence on shim is apparent for the investigated range.

### 3.3 | Dataset size, active learning, and ensembles of networks

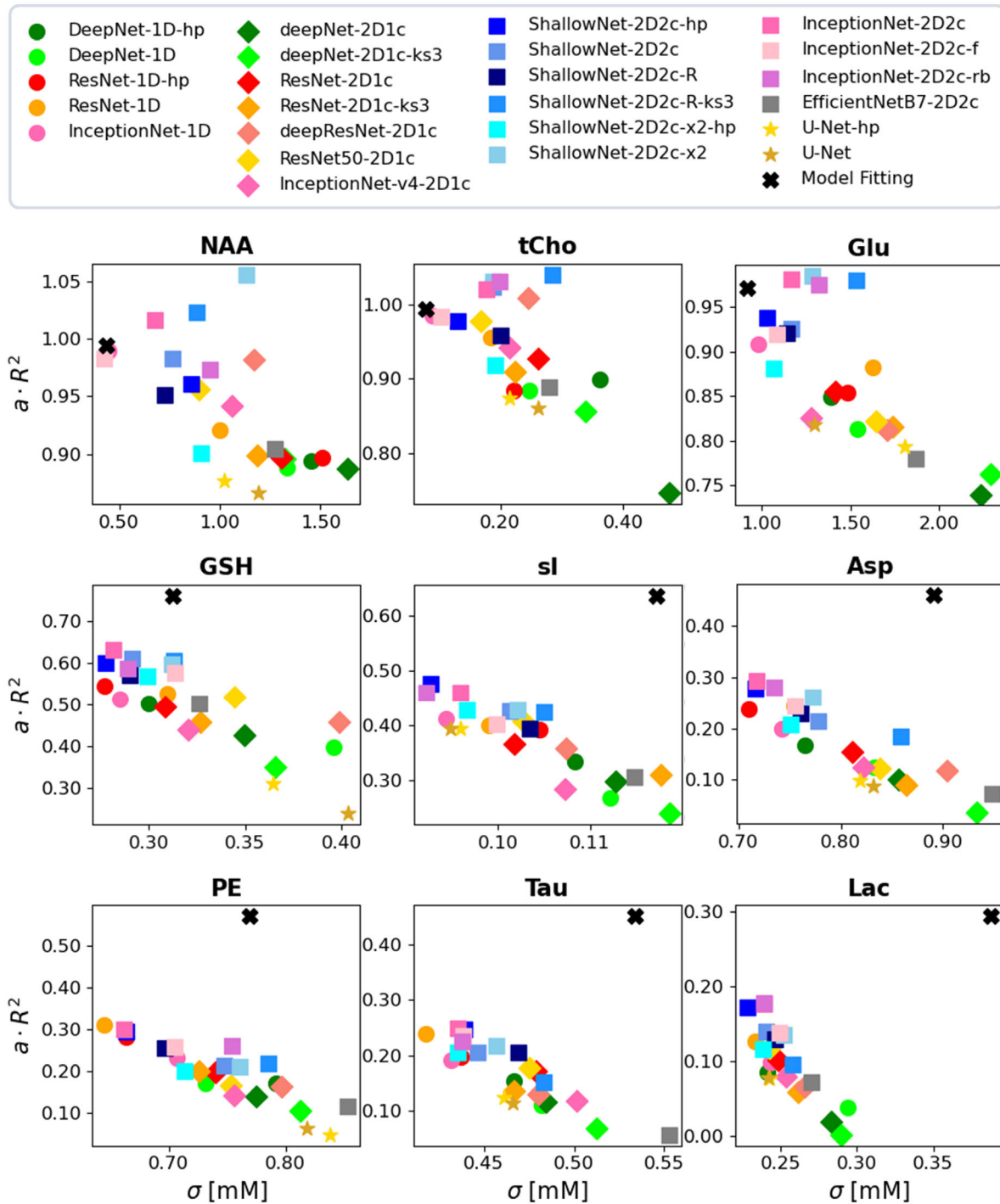
Figure 9 reports on performance improvements by active learning in training phase and dataset sizing (part 9A) as well as by using an ensemble of networks (part 9B) for four metabolites as reflected by *concise scores*. Outcomes of emulated active learning approaches in limiting the testing sets are illustrated through regression plots for Gln in Figure 9C. Detailed comparisons for 16 metabolites are given in Figure S15, Table S2, Table S3, Figure S16, and Table S4.

#### 3.3.1 | Dataset size

The performance showed moderate improvements for most metabolites when dataset size was increased from 25,000 to 40,000 samples (Figure S9).



**FIGURE 6** Maps and marginal distributions of predictions versus GT for a *ResNet\_1D\_hp* network. Results for nine metabolites are arranged in approximate decreasing order of relative SNR from top left to bottom right. RMSE ( $\sigma$ ) is reported as an overall measure of variability. A regression model ( $y = ax + q$ ) is also provided to judge prediction quality.  $R^2$  measures how well a linear model explains the overall data. Mispredictions can be monitored either by a decrease in  $a$  and  $R^2$  or by visual biases in distributions of predictions (bell shape). The prediction bias toward the mean value of the training distribution is evident for medium- to weakly represented metabolites (e.g., si, Asp, PE, Tau, Lac). On average, metabolites with lower SNR yield higher errors ( $q$  and  $\sigma$  in mM units). Further metabolite results are shown in Figure S11 and results for *ShallowNet-2D2c\_hp* in Figure S15. GT, ground truth; Asp, aspartate; Lac, lactate; PE, phosphoethanolamine; si, sylo-inositol; Tau, taurine.



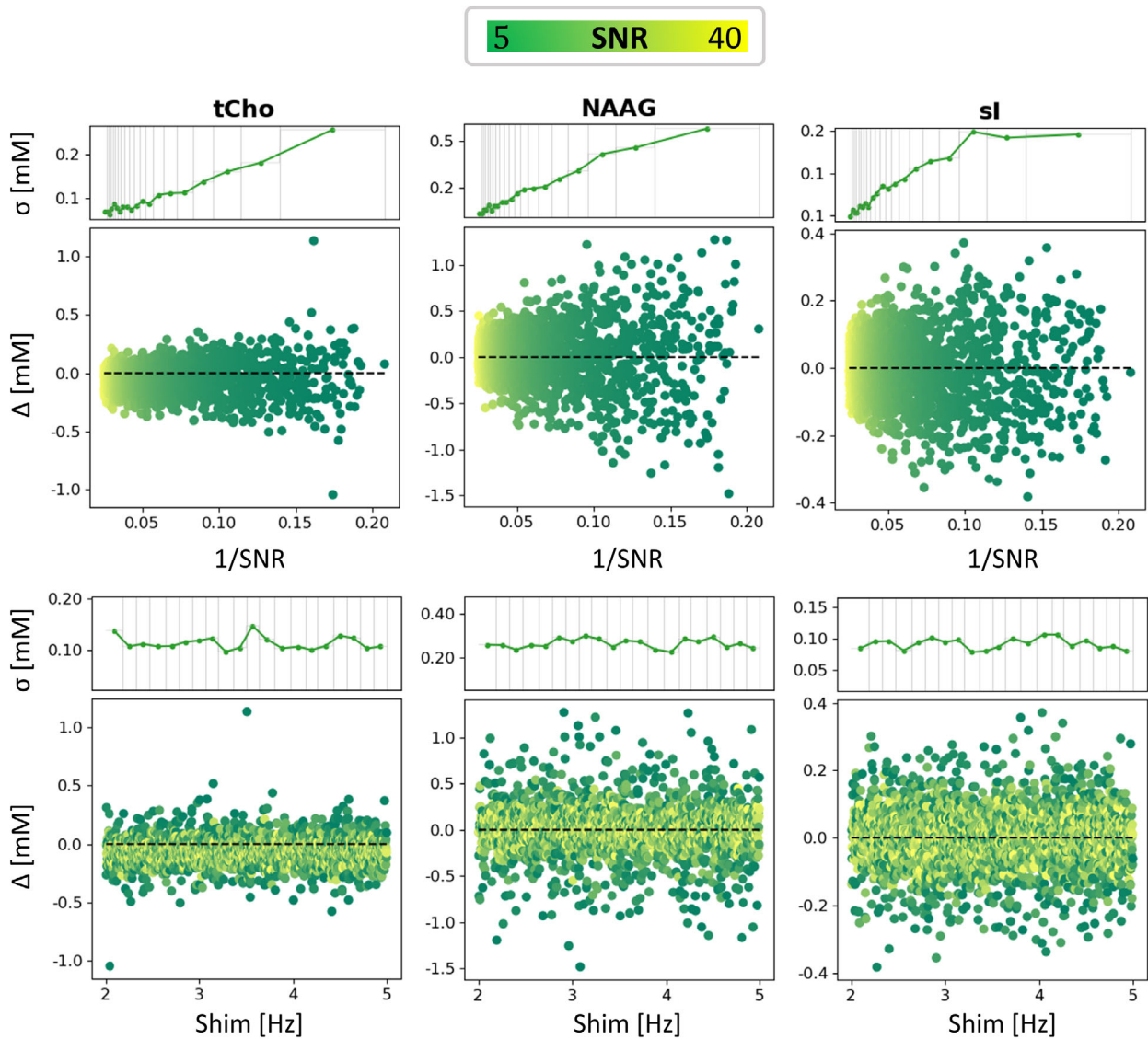
**FIGURE 7** Concise scores presented to compare quantification quality for different networks and input setups (all with water reference). Network identification is chosen as follows: *NetworkType-InputType-properties*. 1c, 1 channel; 1D, spectra; 2D, spectrograms; f, factorized convolution; hp, Bayesian hyperparameterized architecture; ks3, convolutional kernel size = 3; R, exploiting ReLU activations; rb, downsampling via reduction blocks; x2, double convolution before MaxPooling.

### 3.3.2 | Active learning

Dataset augmentation to favor training with combinations of low or high concentrations of weakly represented metabolites (see Figure S6B–S6D) does not substantially improve performance (Figure 9A, Figure S15,

Table S2). Mild improvements (<6% for  $a$ ,  $q$ ,  $R^2$  and  $\sigma$ ) are seen for GABA and sl, respectively, when exploiting metabolite-specifically augmented datasets (*GABA-w*, *sl-w*). Increased dataset size combined with data augmentation to favor high and low concentrations of different metabolites (*GSPT-w*) moderately improves performances





**FIGURE 8** Illustration of the SNR and shim dependence of prediction quality. The CNN's prediction error  $\Delta$  ( $prediction - GT$ ) and the RMSE ( $\sigma$ ) are plotted as a function of SNR (top row) and shim (bottom row) for four metabolites. Results reported for network type *ShallowNet-2D2c-hp* with water reference. RMSE is averaged over bins with an equal number of samples. Bins' width increases for low SNR values. Errors scale approximately linearly with  $1/SNR$  and are insensitive to different shim setups.

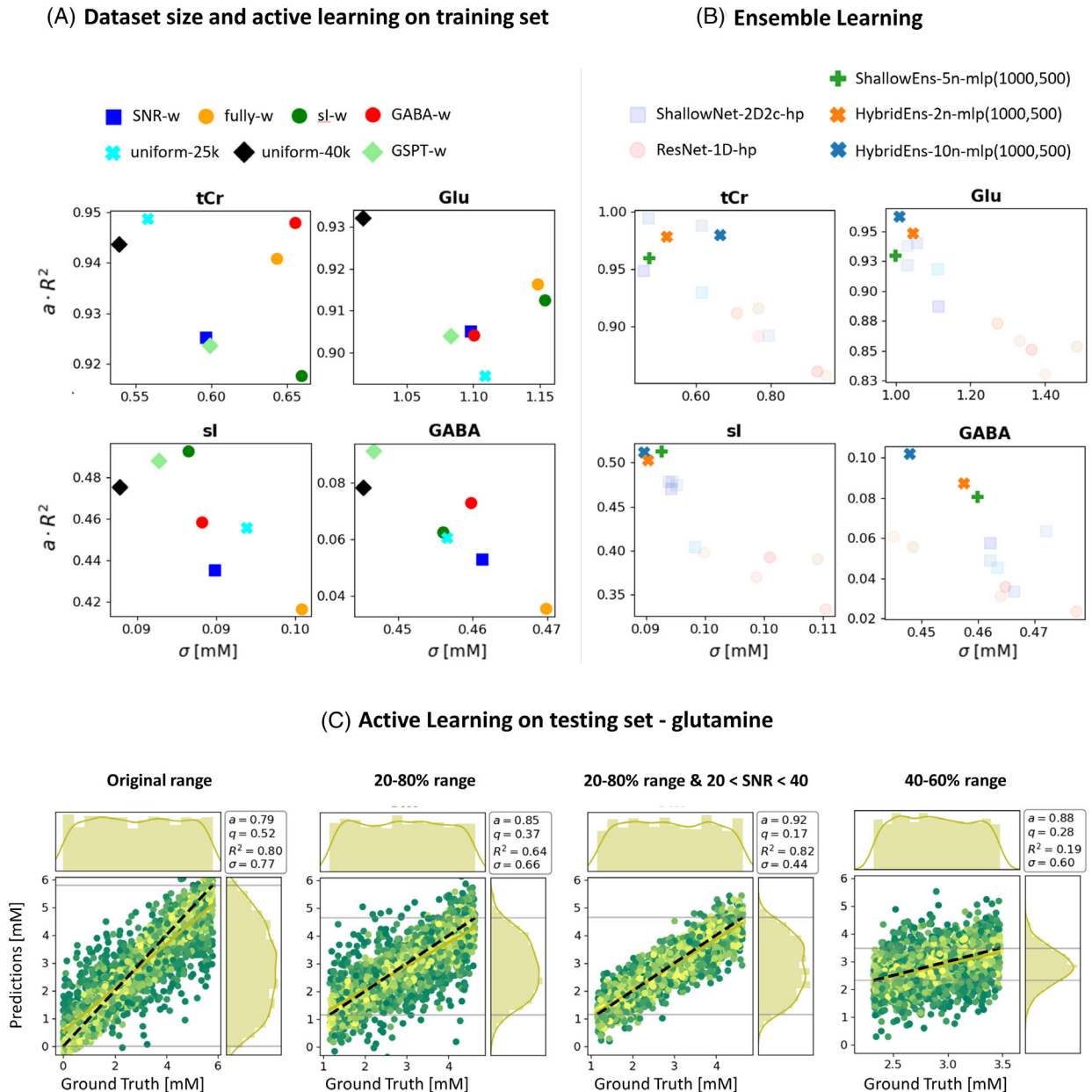
for the augmented metabolites (GABA, sI, PE, Tau). It also extends mild improvements to medium- to weakly represented metabolites that have not undergone data augmentation (e.g., Lac, Gly, Gln). A dataset that is strongly weighted toward extreme combinations of low or high concentration for all metabolites (*fully-w*) or a dataset weighted toward low SNR (*SNR-w*) deteriorated performances.

Clipping the test set to 20%–80% or 40%–60% of the concentration range in training renders improved performances (on average  $a + 4.5\%$ ,  $q - 10.2\%$ ,  $\sigma - 23.9\%$  and  $a + 4\%$ ,  $q - 37.5\%$ ,  $\sigma - 36.2\%$ , respectively), which is even enhanced further when the testing set includes samples with higher SNR (on average  $a + 15.4\%$ ,  $q - 45.4\%$ ,

$\sigma - 36.2\%$ ). Given the limited range on the y-axis,  $R^2$  is less representative (Figure 9C, Table S3).

### 3.3.3 | Ensemble of networks

Ensembles of Bayesian-optimized networks show consistent and relevant  $a \cdot R^2$  improvements for medium- to weakly represented metabolites without deteriorating performance for well-defined metabolites. A hybrid ensemble outperforms the ensemble of networks of the same type. The performance of the ensemble increases with the number of combined networks (Figures 9B, S16) (Table S4).



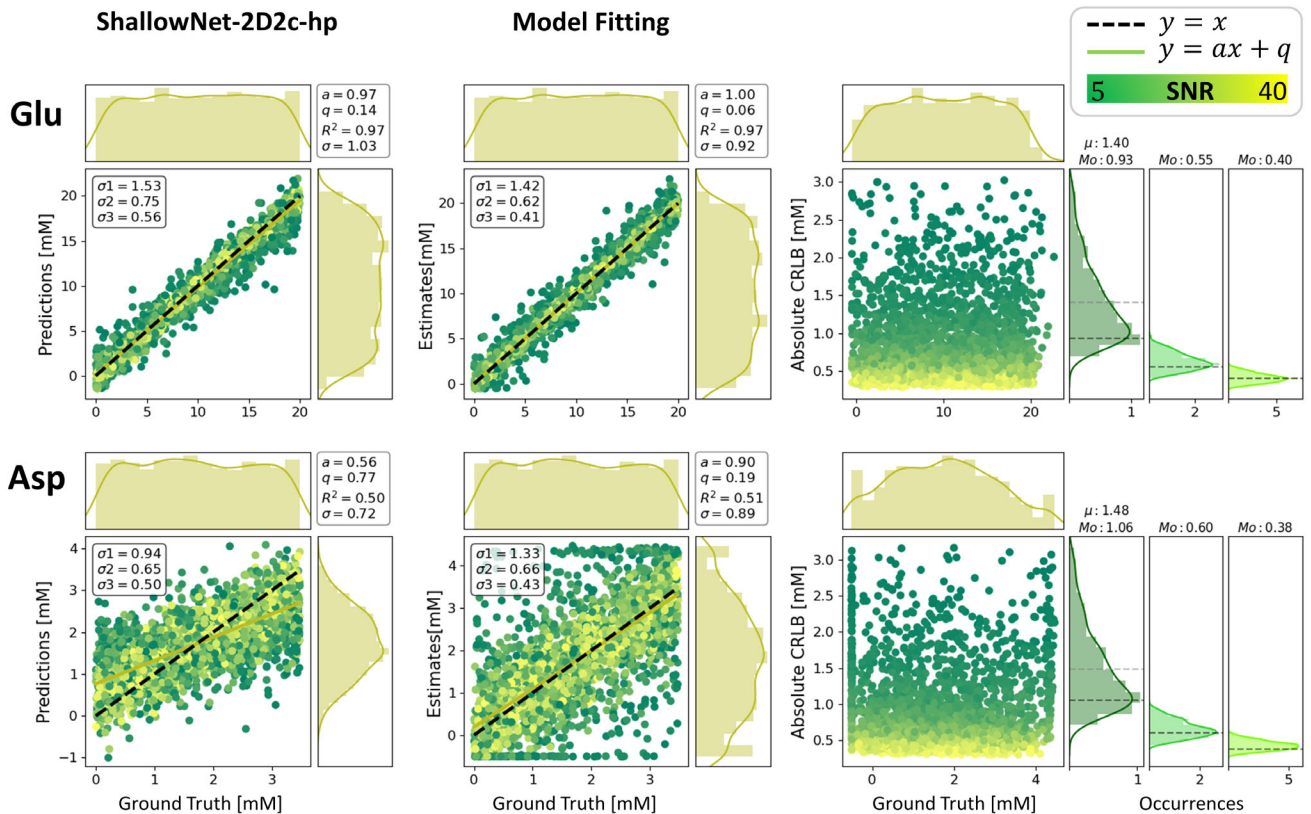
**FIGURE 9** Outcome comparison for the influence of dataset size, active learning approaches, and ensemble of networks (all with water reference). Concise scores evaluated on the same testing set for tCr, Glu, sl, and GABA in different setups: (A) Dataset size and active learning on the training set (for abbreviations alluding to types of active learning extensions, see Results 3.3.2). (B) Ensemble of networks (for naming, see Figures 3 and 4). Ensemble models improve predictions for weakly to medium-represented metabolites without worsening the already good single-network performances for well-represented metabolites (higher  $a \cdot R^2$  and lower  $\sigma$ ). (C) Active learning on the testing set monitored via maps and marginal distribution of predictions versus GT for glutamine. Improvements for clipped concentration ranges can be monitored via scores. However, the 40%–60% interval shows a significant number of outliers. Prediction distributions are still far from being uniform. GABA,  $\gamma$ -aminobutyric acid; Glu, glutamate; sl, syllo-inositol; tCr, total creatine.

### 3.4 | CNN predictions versus model fitting estimates

A general juxtaposition of CNN and MF performance is contained in Figure 7. In Figure 10, detailed results are presented for two metabolites in the form of regression plots for *ShallowNet-2D2c-hp* and MF with FiTAID. In addition,

the estimated CRLBs from MF are displayed and then compared in subgroups of SNR with the variance found in MF estimates and CNN predictions.

Area-constrained MF shows biases at the parameter boundaries for weakly represented metabolites (e.g., Asp). However, traditional MF outperforms quantification via DL: regression lines show less bias ( $a$  and  $q$ ), and the



**FIGURE 10** Comparison of performance for deep learning and model fitting reported for two illustrative metabolites. (Left) DL prediction versus GT mapped via *ShallowNet-2D2c-hp* with water reference. (Middle) Estimates versus GT for the MF approach. (Right) CRLBs evaluated on the fitted estimates. Histograms on the right group three subsets of an equal number of samples with different levels of SNR—group 1: SNR < 16.7, group 2: 16.7 < SNR < 28.4, and group 3: SNR > 28.4 displaying the distribution of estimated CRLBs. For group 1, given the skewness of distribution, mode (*Mo*) and mean ( $\mu$ ) values are reported. For comparison, RMSEs ( $\sigma$ ) are reported as estimated for each SNR group for both DL and MF. DL's RMSEs ( $\sigma$ ) underestimate CRLBs for low relative SNR metabolites.

distribution shape of estimates is closer to a uniform pattern within the GT range. RMSEs ( $\sigma$ ) are higher in the case of MF for medium- to weakly represented metabolites (e.g., Asp) but lower for well-defined metabolites (e.g., Glu) (as formerly noted in Figure 7). Consequently, although  $\sigma$ s of MF are bigger than the CRLBs estimated for their SNR reference group,  $\sigma$ s of DL overestimate CRLBs for well-defined metabolites and underestimate CRLBs for weakly represented metabolites.

## 4 | DISCUSSION

Quantitation of brain metabolites using deep learning methods with spectroscopy data in 1D, 2D, and a combined input format was implemented in multiple network architectures. The main aim of the investigation was to compare the core performance of quantification in an idealized setting of simulated spectra. In fact, the analysis of the optimal performance of both, MF and DL, may otherwise be blurred by additional experimental inaccuracies or

artifacts from actual in vitro or in vivo spectra. Moreover, these nuisance contributors may be tackled in separate traditional or DL preprocessing steps that are beyond the current analysis. Many of the methods proved successful in providing absolute concentration values even when using a very large concentration range for the tested metabolites that goes way beyond the near-normal range that has often been used in the past. In addition, different forms of network input were tested, including a specifically tailored time-frequency domain representation and a downscaled water peak for easing of quantification. Whereas data augmentation by active learning schemes showed only modest improvements, ensembles of heterogeneous networks that combine both input representation domains improve the quantitation tasks substantially.

Results from DL predictions were compared to estimates from traditional MF, where it was found that MF is more accurate than DL at high and modest relative noise levels. MF yields higher variance at low SNR, with estimated concentrations artificially aggregated at the boundaries of the fitting parameter range. Predictions



obtained with DL algorithms delusively appear more precise (lower RMSE) in the low SNR regime, which may misguide nonexperts to believe that the DL predictions are reliable even at low SNR. However, these predicted concentrations are strongly biased by the dataset the network has been trained with. Hence, in case of high uncertainty (e.g., metabolites with low relative SNR or present in concentrations at the edge of the parameter/training space), the predicted concentration tends toward the most likely value: the average value from the training set.

#### 4.1 | Forms of input to networks

Previously, 1D spectra have mostly been used as input for DL algorithms. Here, they have been compared and combined with 2D time-frequency domain spectrograms that had explicitly been designed to be of manageable size while retaining those areas of the high-resolved standard spectrogram that contain the most relevant information, that is, rich in detail in frequency domain to distinguish overlapping spectral features but also maintaining enough temporal structure to characterize  $T_2^*$  signal decay. This comes at the cost that the spectrogram creation cannot be reversed mathematically. However, this is irrelevant when serving as input to a DL network. It was found that this tailored time-frequency representation as input in combination with a shallow CNN architecture performs best and outperforms the use of traditional 1D frequency-domain input for straight quantification or for metabolite basis spectrum isolation with subsequent integration. Furthermore, DL quantitation performance improved upon the inclusion of a downscaled water peak for reference, likely solving scaling issues if no reference is provided.

#### 4.2 | Active learning

Active learning has been explored by extending the training dataset with cases that appeared challenging to predict in the original setup. In particular, new training data with nonequal distribution of metabolite concentrations have been used with a predominance of single or multiple metabolites at low or high concentrations. None of these trials led to substantial improvements, although it might be helpful if specific metabolites are targeted primarily. Such data augmentation for all metabolites simultaneously even deteriorated the overall network performance. This can be understood given that augmentation at the border of the concentration range inherently leads to an underrepresentation of intermediate cases, which are

equally relevant for the overall performance. Extending the size of the training set even further in an unspecific manner appears to still yield modest improvements.<sup>75</sup> In addition, an unconventional way of active learning was probed by using unequal dataset ranges in training and testing by limiting testing on the central portion of the training range. This setup clearly ameliorated some of the issues at the edges of the testing range found in the typical setup. This approach was only implemented by reducing the test range rather than expanding the training range, which would yield better comparable outcome scores (e.g.,  $R^2$ ). However, expanding the training range to negative concentrations may be questionable.

While data augmentation with a bigger proportion of low SNR spectra leads to worse performance, the theoretical prediction limits for good SNR data are probed in the noiseless scenario in which training and testing are run with GT data. Example results for a *ShallowNet* architecture are reported in Figure S17 for NAA, GSH, and Lac. This, combined with the results discussed, suggests that the bottleneck that limits higher prediction performances is SNR, just like in traditional MF, regardless of the implementation of state-of-the-art networks, network optimization, or dataset augmentation. It thus reflects limitations in clinical applications where high enough SNR is just not available. According to this study, DL cannot do miracles unless one accepts the bias toward training conditions.<sup>73</sup>

#### 4.3 | Ensemble of networks

An ensemble of networks has been implemented, and it shows improvements for quantifying metabolites. A combination of networks is less sensitive to the specifics of the training and helps reduce the variance in the predictions. Furthermore, ensembles of networks where multiple noise-sensitive predictions are weighted are more robust to noise. However, even the thus optimized networks underperform in comparison to MF. For MF, CRLBs clearly indicate limits for the confidence in the fit results. For DL, including the optimized ensemble of networks, such limits can only vaguely be deduced from the distributions of predicted values with the major danger of bias toward training data norms.<sup>76,77</sup> The CRLB would provide good guidance for the valid range of DL predictions as well—although of course they are not readily available without the model. New tools to estimate precision and replace CRLB in the case of DL<sup>76,77</sup> still have to prove their value in practice. The situation will be different again if the DL quantification is trained to include cleaning of spectra from artifacts (ghosts, baseline interference) where CRLBs are not available.



#### 4.4 | Low SNR regime

Both MF and DL show lower reliability in quantifying metabolites in the low SNR regime. Clear-cut SNR limits for validity of concentration estimates are not available, neither for MF nor for DL, although SNR values are often indicated as measure of spectral quality. While CRLBs provide a widely used and easy-to-interpret reliability measure that includes the influence of SNR, a similar widely accepted concept does currently not extend to DL approaches.<sup>77</sup> Obviously, a SNR threshold for DL reliability would have to be metabolite-SNR specific, but already the definition of a meaningful metabolite-specific SNR would be cumbersome given that peak-splitting patterns and number of contributing protons as well as lineshape introduce ambiguity. On top, such a metabolite SNR would depend on the estimated metabolite content, whose reliability is at stake. Therefore, just like for MF, global or metabolite-specific SNR will not be informative enough. An uncertainty measure is needed that is based on the predictions and noise distribution but also integrating the uncertainty propagation of the DL model prediction<sup>78,79</sup> (like the inverse of the Fisher information matrix used in the CRLB definition<sup>74</sup>). Despite flourishing literature,<sup>80,81</sup> addressing uncertainty estimation as a complementary tool for DL interpretability, a full-scale analysis of the robustness and reliability of such models is still challenging.<sup>82-84</sup> First attempts to extend these concepts in DL for MRS quantification are just subject of recent investigations<sup>76,77</sup> but far from general acceptance.

#### 4.5 | Limitations

The current investigation focused on probing multiple DL techniques and input forms for a full range of metabolite concentrations but a limited range of spectral quality. In particular, the shim remained in a broadly acceptable range, no phase or frequency jitter was considered, and no artifactual data was included. Such features could have been integrated in the current setup to arrive at a more realistic framework. However, the core of the findings (performance of the actual quantification step) is expected to remain in place. In addition, it is recommended to add separate preprocessing steps to prepare the data for the presented algorithms rather than to combine processing and quantification in a single process.<sup>3</sup> They could be realized in the form of dedicated DL networks, such as those proposed for phase and frequency drift corrections,<sup>20,85,86</sup> and stacked before the quantification model. This would also ensure the essential gain in speed expected from DL quantification models.

Direct comparison with previously proposed successful DL quantification implementations like Ref. 22 was not possible or meaningful for lack of open access network details and differences in the considered spectra.

Our particular implementation used to create spectrograms was optimized to maintain relevant resolution but downweights the initial part of the FID (initialization of Hamming window). CNN inputs may thus not be fully susceptible to changes in broad signals. Alternative recipes with, for example, prefilled filters or circular datasets, were not explored.

Furthermore, active learning has been explored for a single network type and could in principle be more beneficial for other networks or types of input than what has been found here.

## 5 | CONCLUSIONS

Quantification of MR spectra via diverse and optimized DL algorithms and using 1D and 2D input formats have been explored and have shown adequate performance as long as the metabolite-specific SNR is sufficient. However, as soon as SNR becomes critical, CNN predictions are strongly biased to the training dataset structure.

Traditional MF requires parameter tuning and algorithm convergence, making it more time consuming than DL-based estimates. On the other hand, we have shown that ideally (i.e., with simulated cases) and statistically (i.e., within a variable cohort of cases), it can achieve higher performances when compared to a faster DL approach. DL does not require feature selection by the user, but the potential intrinsic biases at training set boundaries act like soft constraints in traditional modeling,<sup>9</sup> leading estimated values to the average expected concentration range, which is dangerous in a clinical context that requires the algorithm to be unbiased to outliers (i.e., pathological data).

Active learning and ensemble of networks are attractive strategies to improve prediction performances. However, data quality (i.e., high SNR) has proven as bottleneck for adequate unbiased performance.

## ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement # 813120 (inspire-med) and the Swiss National Science Foundation (#320030-175984). We acknowledge the support of NVIDIA Corporation for the donation of a Titan Xp GPU used for some of this research. The authors thank Prof Maurico Reyes (ARTORG Center for Biomedical

Engineering Research, University of Bern, Switzerland) for very helpful discussions.

#### DATA AVAILABILITY STATEMENT

The main part of the code will be available on GitHub (<https://github.com/bellarude>). In addition, simulated datasets will be available on MRSHub (<https://mrshub.org/>). For questions, please contact the authors.

#### ORCID

Rudy Rizzo  <https://orcid.org/0000-0003-4572-5120>

Sreenath P. Kyathanahally  <https://orcid.org/0000-0002-7399-8487>

Amirmohammad Shamaei  <https://orcid.org/0000-0001-8342-3284>

Roland Kreis  <https://orcid.org/0000-0002-8618-6875>

#### REFERENCES

- De Graaf RA. *In Vivo NMR Spectroscopy: Principles and Techniques*. 3rd ed. John Wiley & Sons; 2019.
- Kreis R, Boer V, Choi IY, et al. Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: background and experts' consensus recommendations. *NMR Biomed*. 2020;34:e4347. doi:10.1002/nbm.4347
- Near J, Harris AD, Juchem C, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR Biomed*. 2021;34:e4257. doi:10.1002/nbm.4257
- Ratney H, Sdika M, Coenradie Y, Cavassila S, van Ormondt D, Graveron-Demilly D. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR Biomed*. 2005;18:1-13. doi:10.1002/nbm.895
- Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn Reson Med*. 1993;30:672-679. doi:10.1002/mrm.1910300604
- Wilson M, Reynolds G, Kauppinen RA, Arvanitis TN, Peet AC. A constrained least-squares approach to the automated quantitation of in vivo <sup>1</sup>H magnetic resonance spectroscopy data. *Magn Reson Med*. 2011;65:1-12. doi:10.1002/mrm.22579
- Chong DGQ, Kreis R, Bolliger CS, Boesch C, Slotboom J. 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. *MAGMA*. 2011;24:147-164. doi:10.1007/s10334-011-0246-y
- Bhogal AA, Schür RR, Houtepen LC, et al. <sup>1</sup>H-MRS processing parameters affect metabolite quantification: the urgent need for uniform and transparent standardization. *NMR Biomed*. 2017;30:e3804. doi:10.1002/nbm.3804
- Marjańska M, Deelchand DK, Kreis R, et al. Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM. *Magn Reson Med*. 2022;87:11-32. doi:10.1002/mrm.28942
- Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444. doi:10.1038/nature14539
- Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2019;29:102-127. doi:10.1016/j.zemedi.2018.11.002
- Lam F, Li Y, Peng X. Constrained magnetic resonance spectroscopic imaging by learning nonlinear low-dimensional models. *IEEE Trans Med Imaging*. 2020;39:545-555. doi:10.1109/TMI.2019.2930586
- Klukowski P, Augoff M, ZieRba M, Drwal M, Gonczarek A, Walczak MJ. NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. *Bioinformatics*. 2018;34:2590-1597. doi:10.1093/bioinformatics/bty134
- Hatami N, Sdika M, Ratney H. Magnetic resonance spectroscopy quantification using deep learning. *Lect Notes Comput Sci*. 2018;11070:467-475. doi:10.1007/978-3-030-00928-1\_53
- Lee H, Lee HH, Kim H. Reconstruction of spectra from truncated free induction decays by deep learning in proton magnetic resonance spectroscopy. *Magn Reson Med*. 2020;84:559-568. doi:10.1002/mrm.28164
- Iqbal Z, Nguyen D, Thomas MA, Jiang S. Deep learning can accelerate and quantify simulated localized correlated spectroscopy. *Sci Rep*. 2021;11:8727. doi:10.1038/s41598-021-88158-y
- Lee HH, Kim H. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. *Magn Reson Med*. 2019;82:33-48. doi:10.1002/mrm.27727
- Kyathanahally SP, Döring A, Kreis R. Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy. *Magn Reson Med*. 2018;80:851-863. doi:10.1002/mrm.27096
- Gurbani SS, Schreiber E, Maudsley AA, et al. A convolutional neural network to filter artifacts in spectroscopic MRI. *Magn Reson Med*. 2018;80:1765-1775. doi:10.1002/mrm.27166
- Tapper S, Mikkelsen M, Dewey BE, et al. Frequency and phase correction of J-difference edited MR spectra using deep learning. *Magn Reson Med*. 2021;85:1755-1765. doi:10.1002/mrm.28525
- Jang J, Lee HH, Park JA, Kim H. Unsupervised anomaly detection using generative adversarial networks in <sup>1</sup>H-MRS of the brain. *J Magn Reson*. 2021;325:106936. doi:10.1016/j.jmr.2021.106936
- Lee HH, Kim H. Deep learning-based target metabolite isolation and big data-driven measurement uncertainty estimation in proton magnetic resonance spectroscopy of the brain. *Magn Reson Med*. 2020;84:1689-1706. doi:10.1002/MRM.28234
- Gurbani SS, Sheriff S, Maudsley AA, Shim H, Cooper LAD. Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting. *Magn Reson Med*. 2019;81:3346-3357. doi:10.1002/mrm.27641
- Chandler M, Jenkins C, Shermer SM, Langbein FC. MRSNet: metabolite quantification from edited magnetic resonance spectra with convolutional neural networks. 2019 arXiv:1909.03836v1 [eess.IV]. 10.48550/arXiv.1909.03836
- Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. doi:10.1016/J.MEDIA.2017.07.005
- Gyori NG, Palombo M, Clark CA, Zhang H, Alexander DC. Training data distribution significantly impacts the estimation of tissue microstructure with machine learning. *Magn Reson Med*. 2022;87:932-947. doi:10.1002/MRM.29014
- Espi M, Fujimoto M, Kinoshita K, Nakatani T. Exploiting spectro-temporal locality in deep learning based acoustic event detection. *J Audio Speech Music Proc*. 2015;2015:26. doi:10.1186/s13636-015-0069-2



28. Thomas S, Ganapathy S, Saon G, Soltau H. Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2014:2519–2523. doi:10.1109/ICASSP.2014.6854054
29. Alaskar H. Deep learning-based model architecture for time-frequency images analysis. *Int J Adv Comput Sci Appl*. 2018;9:486–494. doi:10.14569/IJACSA.2018.091268
30. Zagoruyko S, Komodakis N. Wide residual networks. In *ArXiv*; 2017:arXiv:1605.07146. doi:10.5244/C.30.87
31. Cohn DA, Ghahramani Z, Jordan MI. Active learning with statistical models. *J Artif Intell Res*. 1996;4:129–145. doi:10.1613/JAIR.295
32. Hansen LK, Salamon P. Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell*. 1990;12:993–1001. doi:10.1109/34.58871
33. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60:84–90. doi:10.1145/3065386
34. Patra S, Bruzzone L. A cluster-assumption based batch mode active learning technique. *Pattern Recognit Lett*. 2012;33:1042–1048. doi:10.1016/J.PATREC.2012.01.015
35. Maiora J, Ayerdi B, Graña M. Random forest active learning for AAA thrombus segmentation in computed tomography angiography images. *Neurocomputing*. 2014;126:71–77. doi:10.1016/J.NEUCOM.2013.01.051
36. Kutsuna N, Higaki T, Matsunaga S, et al. Active learning framework with iterative clustering for bioimage classification. *Nat Commun*. 2012;3:1032. doi:10.1038/ncomms2030
37. Lewis DD, Gale WA. A sequential algorithm for training text classifiers. In the *17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1994: 3–12. doi:10.1007/978-1-4471-2099-5\_1
38. Tuia D, Ratle F, Pacifici F, Kanevski MF, Emery WJ. Active learning methods for remote sensing image classification. *IEEE Trans Geosci Remote Sens*. 2009;48:2218–2232. doi:10.1109/TGRS.2008.2010404
39. Silva C, Ribeiro B. Margin-based active learning and background knowledge in text mining. In the *4th International Conference on Hybrid Intelligent Systems*, 2005: 8–13. doi:10.1109/ICHIS.2004.70
40. Pedrosa de Barros N, McKinley R, Wiest R, Slotboom J. Improving labeling efficiency in automatic quality control of MRSI data. *Magn Reson Med*. 2017;78:2399–2405. doi:10.1002/mrm.26618
41. Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press; 2005.
42. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. 1997;55:119–139. doi:10.1006/jcss.1997.1504
43. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In the *22nd International Conference on Knowledge Discovery and Data Mining*; 2016: 785–794. doi:10.1145/2939672
44. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. In the *31st International Conference on Neural Information Processing Systems*; 2017: 3149–3157. doi:10.5555/3294996
45. Soher BJ, Semanchuk P, Todd D, Steinberg J, Young K. VeSPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis. In *Proceedings of the 19th Annual Meeting of ISMRM*, Montréal, Québec, Canada. 2011, 1410.
46. Oz G, Tkac I. Short-echo, single-shot, full-intensity proton magnetic resonance spectroscopy for neurochemical profiling at 4 T: validation in the cerebellum and brainstem. *Magn Reson Med*. 2011;65:901–910. doi:10.1002/mrm.22708
47. The Mathworks Inc. MATLAB (R2019a). MathWorks Inc 2019.
48. Marjańska M, McCarten JR, Hodges J, et al. Region-specific aging of the human brain as evidenced by neurochemical profiles measured noninvasively in the posterior cingulate cortex and the occipital lobe using 1H magnetic resonance spectroscopy at 7 T. *Neuroscience*. 2017;354:168–177. doi:10.1016/j.neuroscience.2017.04.035
49. Hoefemann M, Bolliger CS, Chong DGQ, van der Veen JW, Kreis R. Parameterization of metabolite and macromolecule contributions in interrelated MR spectra of human brain using multidimensional modeling. *NMR Biomed*. 2020;33:e4328. doi:10.1002/nbm.4328
50. Bottomley PA, Griffiths JR. *Handbook of Magnetic Resonance Spectroscopy in Vivo: MRS Theory, Practice and Applications*. 1st ed. Hoboken, NJ, Wiley & Sons; 2016.
51. Oz G, Alger JR, Barker PB, et al. Clinical proton MR spectroscopy in central nervous system disorders. *Radiology*. 2014;270:658–679.
52. Träber F, Block W, Lamerichs R, Gieseke J, Schild HH. 1H metabolite relaxation times at 3.0 Tesla: measurements of T1 and T2 values in normal brain and determination of regional differences in transverse relaxation. *J Magn Reson Imaging*. 2004;19:537–545. doi:10.1002/jmri.20053
53. An L, Li S, Shen J. Simultaneous determination of metabolite concentrations, T1 and T2 relaxation times. *Magn Reson Med*. 2017;78:2072–2081.
54. Zhang Y, Shen J. Simultaneous quantification of glutamate and glutamine by J-modulated spectroscopy at 3 Tesla. *Magn Reson Med*. 2016;76:725–732.
55. Cudalbu C, Behar KL, Bhattacharyya PK, et al. Contribution of macromolecules to brain 1H MR spectra: experts' consensus recommendations. *NMR Biomed*. 2021;34:e4393. doi:10.1002/nbm.4393
56. Van RG, Drake FL. *Python 3 Reference Manual*. CreateSpace; 2009.
57. Gulli A, Pal S. *Deep Learning with Keras*. Packt Publishing; 2017.
58. Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. In the *12th USENIX Symposium on Operating Systems Design and Implementation*; 2016: 265–283. doi:10.5555/3026877.3026899
59. Bisong E. Google colab. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress, New York City; 2019. doi:10.1007/978-1-4842-4470-8\_7
60. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In *arXiv*; 2015:1409.1556.
61. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In the *2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016: 770–778. doi:10.1109/CVPR.2016.90
62. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In the *2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016: 2818–2826. doi:10.1109/CVPR.2016.308
63. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In *2015 IEEE Computer Society Conference on Computer Vision*

- and Pattern Recognition (CVPR); 2015: 1–9. 10.1109/CVPR.2015.7298594
64. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. In the *31st AAAI Conference on Artificial Intelligence*; 2017: 4278–4284. 10.48550/arXiv.1602.07261
  65. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *ArXiv*; 2015:arXiv:1502.03167.
  66. Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). In: *ArXiv*; 2016:arXiv:1511.07289.
  67. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. In: *ArXiv*; 2012:arXiv:1206.2944.
  68. Rizzo R, Kreis R. Accounting for bias in estimated metabolite concentrations from cohort studies as caused by limiting the fitting parameter space. In *Proceedings of the 2021 ISMRM & SMRT Annual Meeting and Exhibition, Virtual meeting*, May 15–20, 2021. p. 2011.
  69. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci*. 2015;9351:234–241. doi:10.1007/978-3-319-24574-4\_28
  70. Kingma DP, Ba JL. Adam: a method for stochastic optimization. In *ArXiv*; 2017:arXiv:1412.6980.
  71. Bengio Y. Practical recommendations for gradient-based training of deep architectures. *Lect Notes Comput Sci*. 2012;7700:437–478.
  72. Bengio Y, Goodfellow IJ, Courville A. Optimization for training deep models. *Deep Learning*. MIT Press; 2016.
  73. Landheer K, Juchem C. Are Cramér-Rao lower bounds an accurate estimate for standard deviations in in vivo magnetic resonance spectroscopy? *NMR Biomed*. 2021;34:e4521. doi:10.1002/nbm.4521
  74. Bolliger CS, Boesch C, Kreis R. On the use of Cramér-Rao minimum variance bounds for the design of magnetic resonance spectroscopy experiments. *Neuroimage*. 2013;83:1031–1040. doi:10.1016/j.neuroimage.2013.07.062
  75. Hong S, Shen J. Impact of training size on deep learning performance in in vivo 1H MRS. In *Proceedings of the 2021 ISMRM & SMRT Annual Meeting and Exhibition, Virtual meeting*, May 15–20, 2021, p. 2015.
  76. Lee HH, Kim H. Bayesian deep learning-based 1 H-MRS of the brain: metabolite quantification with uncertainty estimation using Monte Carlo dropout. *Magn Reson Med*. 2022;88:38–52. doi:10.1002/MRM.29214
  77. Rizzo R, Dziadosz M, Kyathanahally SP, Reyes M, Kreis R. Reliability of quantification estimates in MR spectroscopy: CNNs vs traditional model fitting. *Med Image Comput Comput Assist Interv–MICCAI 2022 Lect Notes Comput Sci*. 2022;13438:715–724. doi:10.1007/978-3-031-16452-1\_68
  78. Gal Y. 2016 Uncertainty in deep learning. <https://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf>
  79. Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? In the *31st Conference on Neural Information Processing Systems (NIPS)*; 2017.
  80. Sanchez T, Caramiaux B, Thiel P, Mackay WE. Deep learning uncertainty in machine teaching. In *27th Annual Conference on Intelligent User Interfaces (IUI)*, Vol. 1, 2022. 10.1145/3490099.3511117
  81. Abdar M, Pourpanah F, Hussain S, et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf Fusion*. 2021;76:243–297. doi:10.1016/j.inffus.2021.05.008
  82. Jungo A, Reyes M. Assessing reliability and challenges of uncertainty estimations for medical image segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Proceedings, Part II*. Berlin, Heidelberg, Springer-Verlag. 10.1007/978-3-030-32245-8\_6
  83. Ennab M, McHeick H. Designing an interpretability-based model to explain the artificial intelligence algorithms in healthcare. *Diagnostics*. 2022;12:1557. doi:10.3390/DIAGNOSTICS12071557
  84. Meng C, Trinh L, Xu N, Enouen J, Liu Y. Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset. *Sci Rep*. 2022;12:1–28. doi:10.1038/s41598-022-11012-2
  85. Ma DJ, Le HAM, Ye Y, et al. MR spectroscopy frequency and phase correction using convolutional neural networks. *Magn Reson Med*. 2022;87:1700–1710. doi:10.1002/MRM.29103
  86. Shamaei AM, Starcukova J, Pavlova I, Starcuk Z. Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals. *bioRxiv*. 2022. doi:10.1101/2022.06.28.497332
  87. Lin A, Andronesi O, Bogner W, et al. Minimum reporting standards for in vivo magnetic resonance spectroscopy (MRSin-MRS): Experts' consensus recommendations. *NMR Biomed*. 2021;34:e4484. doi:10.1002/nbm.4484

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**Table S1.** List of probed networks for straight quantification of metabolites and some of their characteristics. The listed characteristics includes the complexity (defined as number of trainable parameters), test loss performance, and training time in sec/epoch. The network identifications were chosen as follows: *NetworkType-InputType-properties*. 1D: spectra, 2D: spectrograms, 1c: 1 channel, ks3: convolutional kernel size = 3, hp: Bayesian hyper-parameterized architecture, R: exploiting ReLU activations, x2: double convolution before Max-Pooling, f: factorized convolution, rb: down-sampling via Reduction-Blocks

**Figure S1.** Schemes of Residual Network configurations with 1D (a) and 2D (b) inputs, as well as a deep residual network (c). The basic network structure is sketched on the left, the architectures of Residual, Identity, and Convolutional Blocks are reported on the right, while specifications are detailed in the tables in the middle, and symbols are explained near the bottom. The deeper Residual Network configuration has two convolutional layers at the beginning without pooling.

**Figure S2.** Schemes of Deep CNN configurations with 2D (a) and 1D (b) inputs, as well as an InceptionNet with



1D inputs (c). Network specifications are detailed in the tables, while the architectures of Reduction Blocks are reported on the bottom right. Symbols are explained near the bottom.

**Figure S3.** Schemes of InceptionNet configurations with 2D inputs on 2 channels. Networks (a) and (b) share the same configuration but (b) exploits convolutional factorization to speed-up training time. (c) Simple concatenation in architectures (a) and (b) are replaced by Reduction Blocks. The architectures of the Reduction Blocks are reported in Figure-S6. Symbols are explained on the right.

**Figure S4.** Schemes of (a) InceptionNet with 2D inputs and 1 channel, (b) EfficientNetB7, (c) ResNet50 and (d–f) Shallow Network configurations. Networks (a), (b) and (c) are modified from [1–3], respectively. (d) Implements ELU activations, (e) implements RELU activations, whereas (f) implements a deeper configuration with consecutive convolutional layers with sparse pooling. Network specifications are detailed in the tables. Symbols are explained near the bottom.

**Figure S5.** Scheme and detail of U-Net-1DR-hp configurations for metabolite basis-set prediction. Metabolite-specific network specifications are detailed in the tables. Symbols are explained at the bottom left.

**Text S1.** Details of Bayesian hyper-parameterization

**Figure S6.** Examples of dataset augmentation techniques representing sample distributions for two metabolites (NAA and GABA). (a) Dataset size increment with uniform distributed concentrations. (b) and (c) Active Learning weighted on higher occurrences of small and high concentrations for all metabolites in (b) and for selected metabolites in (c). (d) Active Learning weighted on more occurrences of low SNR entries whereas concentration distributions are kept uniform.

**Figure S7.** Comparison of prediction scores for *well-represented* and *medium-represented* metabolites for three CNN architectures with datasets with (red, black, or blue) and without (yellow, gray, or green) water reference. Mean values in orange. On average, water referencing yields higher coefficients  $a$  and  $R^2$  and lower offset  $q$  and RMSE  $\sigma$ .

**Figure S8.** Comparison of prediction scores for *medium-represented* and *weakly-represented* metabolites for three CNN architectures with datasets with (red, black, or blue) and without (yellow, gray, or green) water reference. Mean values in orange. On average, water referencing yields higher coefficients  $a$  and  $R^2$  and lower offset  $q$  and RMSE  $\sigma$ .

**Figure S9.** Maps and marginal distributions of predictions vs. GT for a *ResNet\_1D\_hp* network. Results for 16 metabolites are arranged in approximate decreasing order of relative SNR from top left to bottom right. RMSE ( $\sigma$ ) is reported as an overall measure of variability. A regression

model ( $y = ax + q$ ) is also provided to judge prediction quality.  $R^2$  measures how well a linear model explains the overall data. Mis-predictions can be monitored either by a decrease in  $a$  and  $R^2$  or by visual biases in distributions of predictions (bell-shape). The prediction bias towards the mean value of the training distribution is evident for medium- to weakly-represented metabolites (e.g., sI, Gly, Asp, PE, Tau, Lac, GABA). On average, metabolites with lower SNR yield higher errors. ( $q$  and  $\sigma$  in mM units.)

**Figure S10.** Maps and marginal distributions of predictions vs. GT for a *ShallowNet-2D2c-hp* network. Results for 16 metabolites are arranged in approximate decreasing order of relative SNR from top left to bottom right. RMSE ( $\sigma$ ) is reported as an overall measure of variability. A regression model ( $y = ax + q$ ) is also provided to judge prediction quality.  $R^2$  measures how well a linear model explains the overall data. Mis-predictions can be monitored either by a decrease in  $a$  and  $R^2$  or by visual biases in distributions of predictions (bell-shape). The prediction bias towards the mean value of the training distribution is evident for medium- to weakly-represented metabolites (e.g., sI, Gly, Asp, PE, Tau, Lac, GABA). On average, metabolites with lower SNR yield higher errors. ( $q$  and  $\sigma$  in mM units.)

**Figure S11.** Boxplots comparing the distributions of predictions for 8 metabolites via 7 different CNN architectures vs. Model Fitting estimate distributions (MF) and uniform Ground Truth (GT) distributions. Mis-prediction is evident for *medium-* to *weakly-*represented metabolites (e.g., sI, Asp, Tau, Lac) and can be monitored by different degrees of skewness of the boxplot. However, the bias to training distribution is not evident given the visual limitation of boxplots. For better visibility of this outcome, see Figure S14.

**Figure S12.** Comparison of distributions of predictions for 8 metabolites via 7 different CNN architectures vs. Model Fitting's estimate distributions (MF) and Ground Truth (GT) uniform distributions. Mis-prediction is evident for *medium-* to *weakly-*represented metabolites (e.g., sI, Asp, Tau, Lac) and can be monitored by visual biases (bell-shape) towardstoward the mean value of the training distribution (i.e., regression to the mean). Note: y-axes scale inhomogeneously between different networks. However, all distributions integrate to 1.

**Figure S13.** *Concise scores* presented to compare quantification quality for different networks and input setups for 16 metabolites. Results reported using the proposed artificial water signal reference. Network identification is chosen as follows: *NetworkType-InputType-properties*. Keywords: 1D: spectra, 2D: spectrograms, 1c: 1 channel, ks3: convolutional kernel size = 3, hp: Bayesian hyper-parameterized architecture, R: exploiting ReLU activations, x2: double convolution before MaxPooling, f:

factorized convolution, rb: down-sampling via Reduction-Blocks.

**Figure S14.** Comparison of performance scores from different networks for 16 metabolites. Model fitting is included in the comparison.

**Text S2.** Comparison of predictions from different CNNs.

**Figure S15.** Comparison of outcomes of Active Learning approaches using concise scores.

**Figure S16.** Quantification outcome as reflected by *concise scores* for differently trained single networks and three ensembles of networks (identical training set for 16 metabolites).

**Figure S17.** Maps and marginal distributions of predictions vs. GT obtained for three metabolites using *ShallowNet-2D2c-hp* as contrasted for a realistic and noiseless dataset.

**Table S2.** Results of Active Learning on training set: scores of 16 metabolites for every augmented training set.

**Table S3.** Results of emulated Active Learning on test set: scores of 16 metabolites for every concentration range considered.

**Table S4.** Outcome for ensemble learning: scores for 16 metabolites for average network or ensemble of network considered.

**Table S5.** MRSinMRS checklist.<sup>87</sup>

**How to cite this article:** Rizzo R, Dziadosz M, Kyathanahally SP, Shamaei A, Kreis R. Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias. *Magn Reson Med.* 2022;1-21. doi: 10.1002/mrm.29561



Contents lists available at ScienceDirect

## Computers in Biology and Medicine

journal homepage: [www.elsevier.com/locate/combiomed](http://www.elsevier.com/locate/combiomed)

# Physics-informed deep learning approach to quantification of human brain metabolites from magnetic resonance spectroscopy data

Amirmohammad Shamaei<sup>a,b,\*</sup>, Jana Starcukova<sup>a</sup>, Zenon Starcuk Jr.<sup>a</sup><sup>a</sup> Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic<sup>b</sup> Department of Biomedical Engineering, Brno University of Technology, Czech Republic

## ARTICLE INFO

## Keywords:

MR spectroscopy  
Inverse problem  
Deep learning  
Machine learning  
Convolutional neural network  
Metabolite quantification

## ABSTRACT

**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. Our proposed method has the potential to facilitate clinical practice by enabling faster processing of large datasets such as high-resolution MRSI datasets, which may have thousands of spectra.

## 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 [1,2].

Linear combination model (LCM) fitting, peak fitting, and peak integration are the three most often used methods for spectral analysis of an MRS signal [3]. 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 [2–5]. 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 <sup>1</sup>H-MRS of the brain due to the excessive number of metabolites and peaks per metabolite [2–5].

LCM fitting is the expert-recommended approach [6] 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 [2–5,7,8].

A long-standing interest in LCM fitting has resulted in the development of a variety of fitting techniques, including time-domain and

\* Corresponding author. Magnetic Resonance Group, Institute of Scientific Instruments of the Czech Academy of Sciences, Kralovopolska 147, CZ 61264, Brno, Czech Republic.

E-mail address: [amirshamaei@isibrno.cz](mailto:amirshamaei@isibrno.cz) (A. Shamaei).

<https://doi.org/10.1016/j.combiomed.2023.106837>

Received 28 October 2022; Received in revised form 6 March 2023; Accepted 26 March 2023

Available online 5 April 2023

0010-4825/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



frequency-domain algorithms [4,9–15].

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) [16,17], suggests that it might also handle the spectral analysis of an MRS signal.

Moreover, supervised DL-based approaches have been used for ghosting artifacts detection and removal [18], spectral reconstruction [19], automatic peak picking [20], magnetic resonance spectroscopic imaging (MRSI) spatial resolution enhancement [21], localized correlated spectroscopy acceleration [22], frequency and phase correction of MRS signals [23,24], and poor-quality spectra identification [25].

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 [26]; 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.

A deep autoencoder (DAE) is typically designed to encode the input into a low-dimensional embedding space and then reconstruct the input from the encoding in a self-supervised manner [16]. 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 [27,28].

A common belief among traditional statisticians and DL practitioners is that "more data is always better" [29]. The effects of model complexity, dataset size, and data augmentation were empirically investigated in computer vision and natural language processing [29,30]; however, these effects have not been investigated in MRS data quantification using DL.

## 2. Authors' contributions

In this paper, we present 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 quantum-mechanically simulated basis spectra to the acquired MR spectrum. We utilize the LCM fitting and integrate it into a DAE to estimate relative concentrations of metabolites. We trained, validated, and evaluated our proposed network using simulated MRS data, and in-vivo MRS data obtained from the publicly accessible Big GABA repository [31,32]. We compared the performance of our proposed model with other LCM packages, namely, QUEST [13] and FiTAID [9], in terms of accuracy and time efficiency.

In addition, we designed an experiment to evaluate the effect of the dataset size and neural network architecture in our proposed model.

## 3. Related work

Hiltunen et al. [33] have demonstrated the feasibility of constructing a quantifying analyzer for long echo time (TE) in vivo proton magnetic resonance spectroscopy (1H NMR) spectra using artificial neural networks with magnitude spectra.

Bhat et al. [34] investigated the application of a radial basis function neural network (RBFNN) for the automatic quantification of short echo time, multi-voxel, and phased spectral data.

Hatami et al. [35] and Lee et al. [36] applied supervised DL approaches to metabolite quantification and presented results comparable to conventional LCM approaches. Chandler et al. [37] also applied a supervised DL approach to study metabolite quantification in edited human brain MRS spectra.

Gurbani et al. [38] presented a self-supervised DL architecture that integrates a convolutional neural network (CNN) with peak fitting. In

their approach, a DAE, introduced by Hinton et al. [39] as a specific type of neural network, is used as a framework for unsupervised learning.

Shamaei et al. [40] extracted characteristics from the MRS data using a wavelet scattering transform and predicted metabolite concentrations using a shallow neural network implementation.

Rizzo et al. [41] explored the effect of DL architectures, spectroscopic input types, and learning designs on the quantification of simulated MR spectroscopy. A list of related work is provided in Table 1.

## 4. Material and methods

### 4.1. Data sets

#### 4.1.1. 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:

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

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.  $\Delta\alpha$ ,  $\Delta f$ , 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 [43] (19 metabolites signals (see Table 2) and one MM signal, 3T, PRESS, TE = 30 ms, spectral width = 4000 Hz, 2048 points). The range of parameters ( $A_m$ ,  $\Delta\alpha$ ,  $\Delta f$ , and  $\Delta\theta$ ) were determined according to the literature [2,44] (listed in Table 2). 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  $67 \pm 15$  by introducing random complex Gaussian white noise ( $\xi$ ). 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.

#### 4.1.2. The in vivo dataset

Data from the public repository Big GABA [31,32] 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 T field strength, spectral width = 4000 Hz, 4096 points, TE = 35 ms, TR = 2000 ms; 64 transients;  $30 \times 30 \times 30 \text{ mm}^3$ ; 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 [16]. However, rich and adequate datasets are uncommon [27] 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

**Table 1**  
A summary of related work.

	Model architectures	Input types	Learning process	Data type (training)	Prior metabolite resonances model
Hiltunen et al. [33]	Shallow Neural Networks	Magnitude spectra (1D)	Supervised	Simulated	-
Bhat et al. [34]	Radial basis function neural network	Real components of spectra (1D)	Self-supervised	In-vivo	Lorentzian-Gaussian lineshape
Hatami et al. [35]	Convolutional neural network	Complex spectra (1D)	Supervised	Simulated	-
Lee et al. [36]	Convolutional neural network	Complex spectra (1D)	Supervised	Simulated	-
Chandler et al. [37]	Convolutional neural network	Real, imaginary, and magnitude components of spectra (2D)	Supervised	Simulated	-
Gurbani et al. [38]	Convolutional neural network	Real components of spectra (1D)	Self-supervised	In-vivo	Lorentzian-Gaussian lineshape
Shamaei et al. [42]	Convolutional neural network	Time-frequency domain scalogram (wavelet coefficients) (2D)	Supervised	Simulated	-
Shamaei et al. [40]	Wavelet scattering network	Complex time-domain signal (FID) (1D)	Supervised	Simulated	-
Rizzo et al. [41]	Convolutional neural network	Real and imaginary components of spectra, time-frequency domain spectrograms (2D)	Supervised	Simulated	-
Our work	Physics-informed convolutional neural network	Complex time-domain signal (FID) (1D)	Self-supervised	Simulated and in-vivo	Quantum-mechanics simulated metabolite responses

**Table 2**

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

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

the spectral registration method [45,46] 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 [47]. The first step was repeated 64 times, producing 64 signals for each subject and thus producing a new extended subject dataset [47].

Totally, 3072 ( $48 \times 64$ ) 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 ( $5 \times 64$ ) 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  $-9^\circ$ - $9^\circ$ , respectively, and added to apodized signals (ranges were extracted from in vivo brain spectra subjected to spectral fitting). The final training dataset consisted of 28864 ( $10 \times 2624 + 2624$ ) signals. The mean SNR of the training dataset was set to  $256 \pm 72$  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.

Example spectra from the simulated dataset and the in-vivo dataset are provided in Fig. 1.

## 4.2. The deep model

### 4.2.1. The proposed DAE

DAE is a sort of deep artificial neural network designed to learn an efficient data coding in a self-supervised way [16,39]. The basic idea behind DAEs is to utilize the input data as the target that should be reconstructed in the output layer [16]. A DAE is typically composed of two parts: an encoder and a decoder. The encoder function  $h = f(x)$  maps the  $n$ -dimensional input vector ( $x \in R^n$ ) to the  $n'$ -dimensional latent vector ( $h \in R^{n'}$ ), while the decoder function  $\hat{x} = g(h)$  aims to reconstruct the  $n$ -dimensional output vector ( $\hat{x} \in R^n$ ) 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

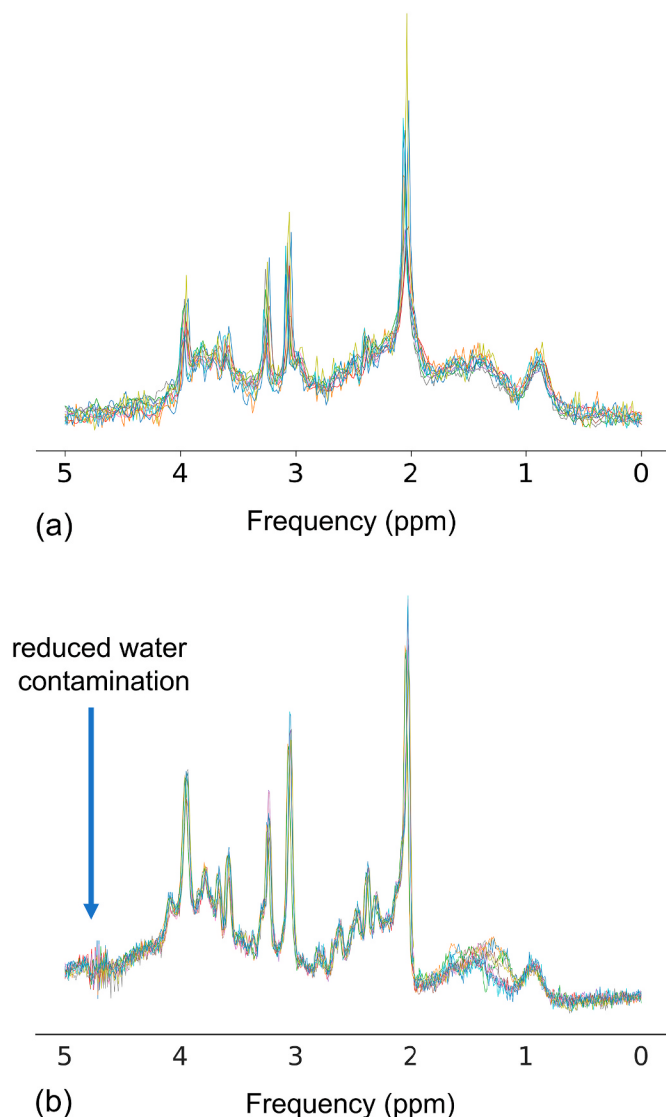


Fig. 1. Example spectra from (a) the simulated dataset and (b) the in-vivo dataset.

prior knowledge are the decoders. For the DAE to be useable 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 [39] is not useful for quantifying MRS signals.

To this end, a convolutional encoder–model decoder [38] 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 [48,49], Gaussian error linear units (GELU) [50], and a fully connected (FC) layer [16,49]. Furthermore, a Softplus function [16] 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 Fig. 2a. 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.

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:

$$\hat{w} = \underset{w}{\operatorname{argmin}} L(x, \hat{x}(w)), \quad [3]$$

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

where  $x$  and  $\hat{x}$  are the  $n$ -dimensional input vector ( $x \in C^n$ ) and the  $n$ -dimensional output vector ( $\hat{x} \in C^n$ ), 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  $P = \{A_1, A_2, \dots, A_M, A_b, \Delta\alpha, \Delta f, \Delta\theta\}$ . Gradients were computed with the help of Pytorch using its automatic gradient computation [51], which is a reverse automatic differentiation system.

#### 4.2.2. Modeling of macromolecules

Short-TE MRS signals contain a contribution of MMs that overlap with metabolites [52]. 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 [6]. MMs can be included as a numerical pattern (found, for instance, from metabolite-nulling inversion recovery measurements) [53]. Another approach is to include a set of parametric patterns in the spectral fitting algorithm [2,3].

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 [53] 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 [4], 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 [4].

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]).

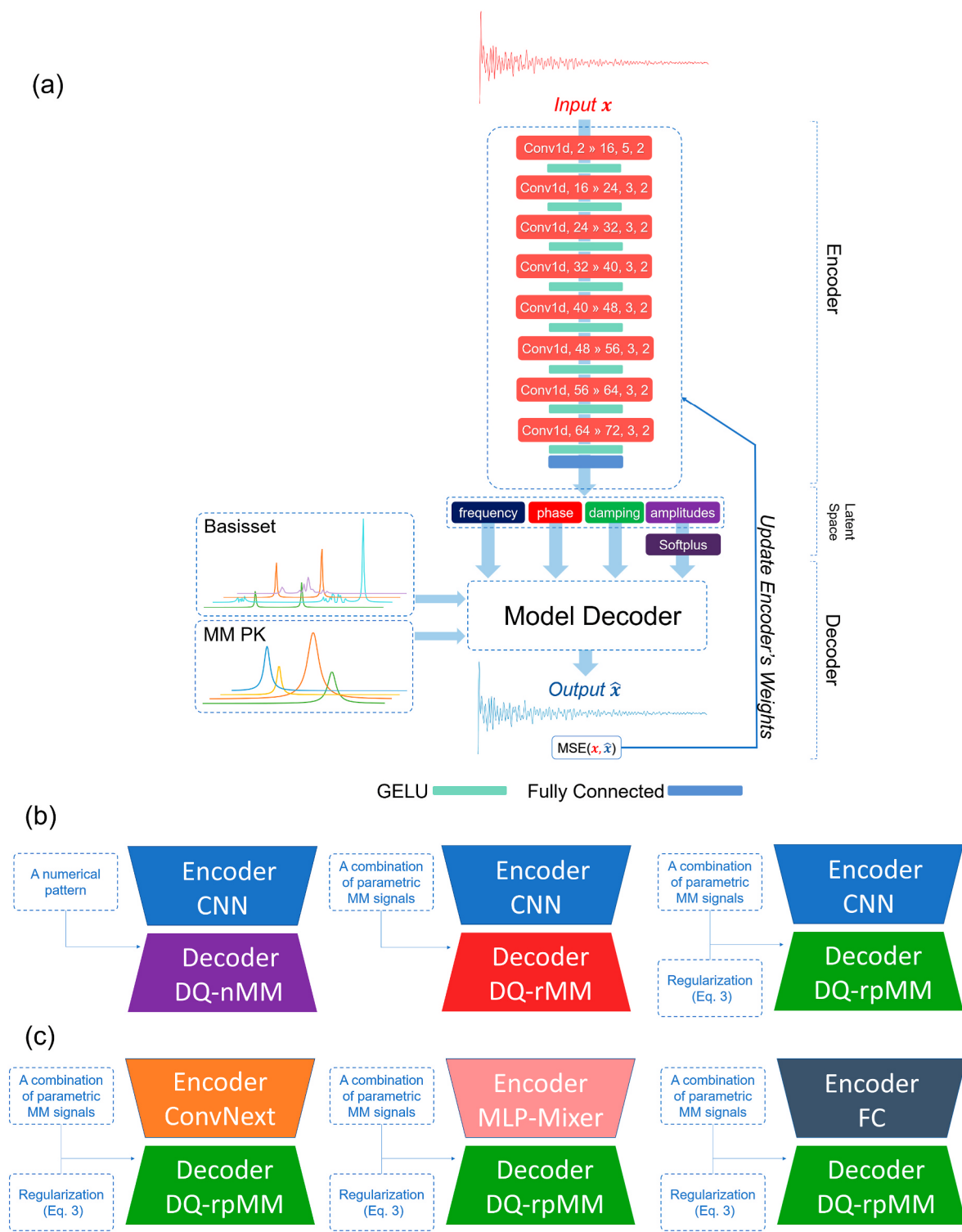
DQ-rpMM includes a combination of parametric MM components in Eq. (1) and adds a regularization term (Eq. (5)) to the loss function.

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

where  $\operatorname{sgn}$  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  $\lambda$  tunes the significance of the regularization term. Brackets  $\|\cdot\|_2^2$  and  $\|\cdot\|_1$  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  $x_{input}$ )





**Fig. 2.** (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.

should be prevented by the contradictory force of the second term. The architectures of the proposed models are depicted in Fig. 2b.

#### 4.3. Implementation details and training

All steps were run on a computer with a dual EPYC 7742 ( $2 \times 64$  cores) processor and one graphics processing unit (NVIDIA A100 40 GB). The DAE was implemented in Python programming language [51] with the help of the Pytorch lightning interface [54].

The initial architecture of the network and the training parameters were optimized using the Bayesian Optimization HyperBand algorithm [55] with the help of the Tune framework [56]. 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 [57] with a batch size of 128, a learning rate of 0.001, and 200 epochs. An early-stopping strategy [51] 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 [58], reducing the learning rate is beneficial when learning becomes stagnant. The learning was reduced twice by a factor of 10 during training. Our source codes and data (in NIFTI-MRS format [59]) are available at <https://github.com/isi-nmr/Deep-MRS-Quantification>.

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 to show the applicability of our method to in-vivo data in which a numeric pattern of MMs is not available. For the simulated dataset, the basis set and the numeric pattern of MMs from the ISMRM MRS study group's fitting challenge [43] were utilized in the model decoder. The provided numeric pattern of MMs was parametrized using AMARES [60] 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 [61] consisted of 19 metabolites was used in the model decoder: alanine (Ala), aspartate (Asp), creatine (Cr), negative creatine methylene (-CrCH<sub>2</sub>),  $\gamma$ -aminobutyric acid (GABA), glycerophosphocholine (GPC), glutathione (GSH), glutamine (Gln), glutamate (Glu), water (H<sub>2</sub>O), 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 Ref. [62]) were used to generate the lipid and MM basis signals.

#### 4.4. 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 [63]. In fact, MAPE does not reveal much information about the effectiveness of regression with regard to the distribution of the ground truth values.

#### 4.5. 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 ( $45^\circ$ ), 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  $\sim 50$ .

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.

#### 4.6. Data size effects

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.

#### 4.7. The regularization term effects

The significance of the regularization term (i.e.,  $\lambda$ ) 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  $\lambda$  on the performance of the proposed model. DQ-rpMM methods were trained seven times with varying  $\lambda$  ( $\{0, 1, 2, 3, 4, 6, 8\}$ ). 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.

#### 4.8. The architecture effects

To test the applicability of successful machine-vision architectures, we used two cutting-edge architectures, ConvNext [64] and MLP-Mixer [65], 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 [64] and MLP-Mixer [65] architectures to process 1D signals. The architectures of the proposed variants are depicted in Fig. 2c. The details of ConvNext, MLP-Mixer, and our proposed network with FC layers architectures are provided in Supporting Information Tables 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  $1 \times 10^{-4}$ .

Table 3 gives a list of the experiments that were carried out in this study.

## 5. Results and discussion

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 [43], in this work, two LCM fitting methods were used, namely, QUEST [13,66] and FiTAID [9], in order to compare the results obtained for the simulated data and

**Table 3**  
A list of experiments.

Experiment	Dataset	Models
Monte Carlo analysis	Simulated	DQ-nMM
Data size effects	Simulated	DQ-nMM
Regularization term effects	Simulated	DQ-rpMM
Architecture effects	Simulated	DQ-rpMM(CNN)
		DQ-rpMM(FC)
		DQ-rpMM(MLP-Mixer)
		DQ-rpMM(ConvNext)

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 [43]. FiTAID performed close to the top in the challenge [43] 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 [66]) 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 [13] (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) was utilized as a model in the QUEST and the FiTAID software for quantification. Table 4 contains a list of used methods.

Table 5 and Fig. 3 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<sup>2</sup> and MAPE than QUEST and FiTAID (time and freq.). QUEST-Subtract showed the worst performance in terms of R<sup>2</sup> 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 Fig. S1.

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

**Table 4**  
A list of all utilized models in this study.

Model	Dataset		MM model	Fitting domain
	simulated	in vivo		
DQ-nMM (ours) <sup>a</sup>	✓	-	Numerical pattern	Time
DQ-pMM (ours) <sup>a</sup>	✓	-	Parametric MM components	Time
DQ-rpMM (ours) <sup>a</sup>	✓	✓	Parametric MM components and regularization term	Time
FiTAID [time] <sup>b</sup>	✓	-	Pumerical pattern	Time
FiTAID [freq.] <sup>b</sup>	✓	-	Pumerical pattern	Frequency
QUEST <sup>b</sup>	✓	-	Pumerical pattern	Time
QUEST-Subtract <sup>b</sup>	✓	-	Subtract technique [13]	Time

<sup>a</sup> The source code is available at <https://github.com/isi-nmr/Deep-MRS-Quantification>.

<sup>b</sup> QUEST and FiTAID software are available in the jMRUI software package (version 7) at <http://www.jmrui.eu/>.

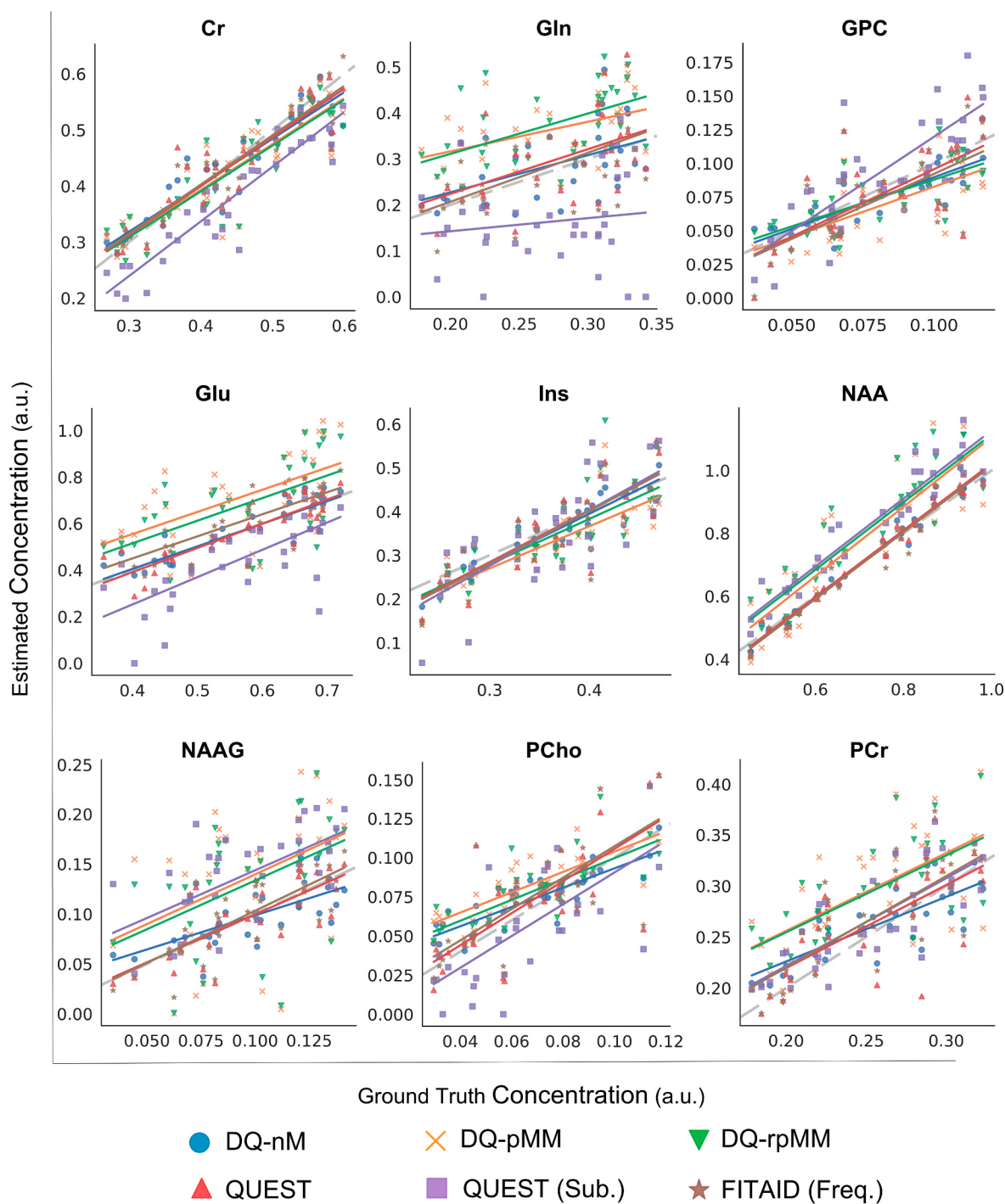
**Table 5**

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

MAPE, mean absolute percentage error; R<sup>2</sup>, 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; QUEST (Sub.), Subtract-QUEST; FiTAID (Time), FiTAID with the time-domain fit; FiTAID (freq.), FiTAID with the frequency-domain fit.

	R <sup>2</sup>	MAPE (%)	Ala	Asp	Cr	GABA	Glc	Gln	Glu	Gly	GPC	GSH	Ins	Lac	NAA	NAAG	PCho	PCr	PE	sHns	Tau
DQ-nMM	0.88	-11.80	-3.89	-4.75	-0.53	19.11	8.88	0.68	-2.42	0.67	-0.29	0.71	-0.47	0.98	0.46	0.62	0.54	-9.26	-2.37	0.66	
DQ-pMM	0.62	47.52	33.04	32.97	-2.32	-16.15	70.00	31.10	29.88	131.96	24.01	49.55	14.89	70.78	12.17	78.63	25.56	18.07	124.28	27.10	30.51
DQ-rpMM	0.73	-105.51	101.61	149.18	9.07	-57.36	14.55	-3.18	-1.56	-14.91	0.40	-5.96	0.45	-4.34	-4.00	0.50	-0.39	-35.00	-2.76	0.03	
FiTAID (Freq.)	0.80	-24.72	101.28	65.50	33.49	27.31	93.06	21.97	0.38	-19.87	0.03	-1.08	0.34	63.80	11.55	16.63	21.05	118.87	27.54	26.62	
FiTAID (Time)	0.85	-24.40	99.65	51.01	25.95	-8.72	-8.72	-1.94	0.44	94.65	27.32	22.68	12.80	75.75	0.24	25.84	27.02	12.29	78.81	36.48	18.78
QUEST	0.69	80.83	100.00	50.07	25.51	12.13	100.00	27.39	-15.06	0.03	-0.92	0.60	10.57	-6.09	0.95	0.11	-0.01	-0.47	-13.71	-4.42	0.45
QUEST (Sub.)	0.83	-22.82	-10.07	-8.86	-1.84	0.55	-18.97	0.10	-0.88	0.35	-5.97	0.97	0.30	75.05	3.92	27.52	27.15	13.55	69.89	36.96	18.84
	0.83	78.58	7.12	52.73	24.54	11.05	92.70	26.06	21.48	12.88	78.09	3.36	24.84	25.75	11.79	76.50	35.05	19.21	-8.21	-2.34	
	0.35	-106.81	-72.01	-22.98	-6.71	-1.27	-48.42	-0.22	-2.93	-7.44	0.51	-2.19	-0.30	0.00	0.00	-95.47	13.40	157.56	51.00	47.70	
	15.45	149.09	146.23	78.23	47.05	24.54	138.09	29.80	31.21	86.95	14.38	56.88	31.95	13.40	157.56	51.00	47.70	157.56	51.00	47.70	





**Fig. 3.** 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.

Fig. 3 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 Fig. S3. DQ-nMM revealed a very good performance in the estimation of Cr ( $R^2 = 0.88$ ), Glu ( $R^2 = 0.68$ ), GPC ( $R^2 = 0.67$ ), Ins ( $R^2 = 0.71$ ), NAA

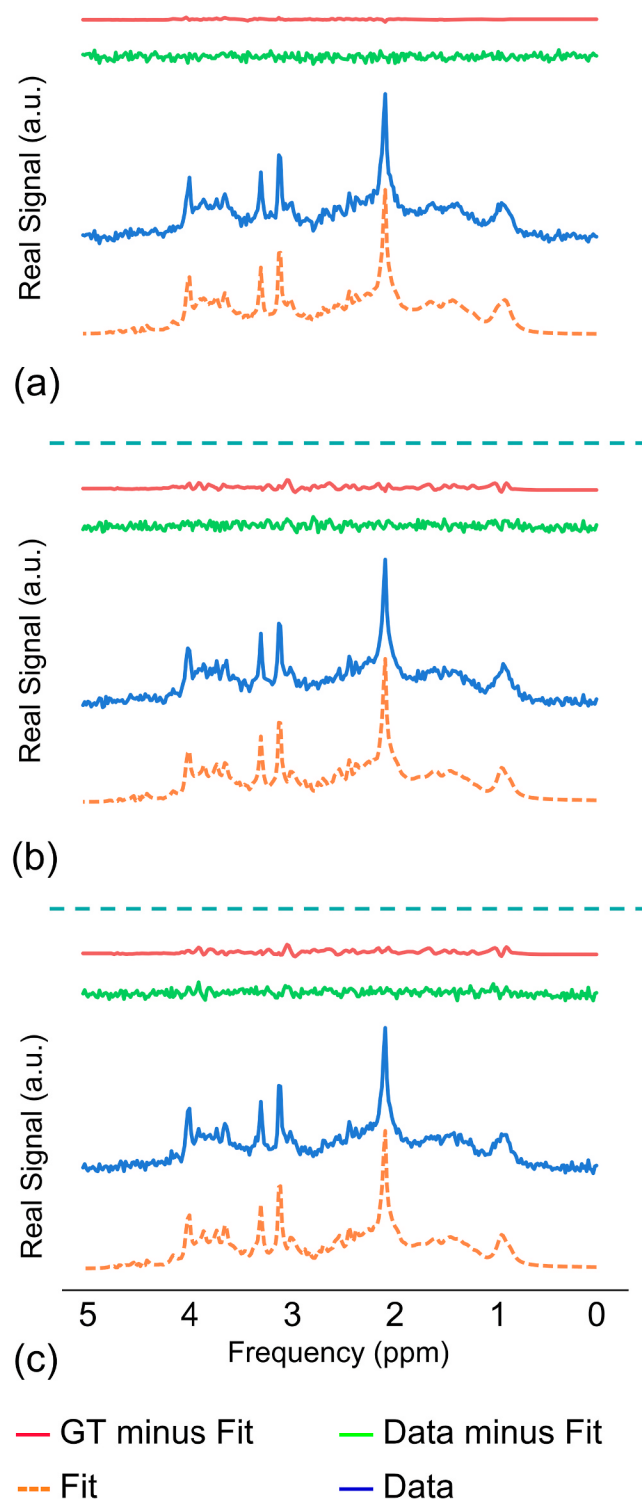
( $R^2 = 0.98$ ), PCho ( $R^2 = 0.62$ ), PCr ( $R^2 = 0.54$ ), and Tau ( $R^2 = 0.66$ ) concentration and a good performance in estimation of NAAG ( $R^2 = 0.46$ ). Interestingly, DQ-nMM outperformed QUEST in the estimation of

GPC ( $R^2 = 0.10$ ), Ins ( $R^2 = 0.35$ ), PCho ( $R^2 = 0.10$ ), PCr ( $R^2 = 0.19$ ), and Tau ( $R^2 = 0.42$ ).

Fig. 4 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.

Fig. 5 shows the results of the MC analysis for the DQ-nMM method. Fig. 5a shows the distribution of estimated concentrations as histogram plots. In most metabolites, the histogram peak is close to the true value,





**Fig. 4.** 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. The contribution of each basis spectrum to the total spectrum was illustrated in Supporting Information Figure S6.

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.

except that the concentration of NAAG was slightly underestimated. Fig. 5b 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. Fig. 5c 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.

Fig. 6a 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). Fig. 6b highlights the effect of  $\lambda$  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  $\lambda$  during training is shown.

The influence of the encoder architecture on the performance of the proposed model (DQ-rpMM) is illustrated in Fig. 7. The performance of the proposed networks with convolutional layers was superior to those with fully connected layers.

Fig. 8 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 6 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.

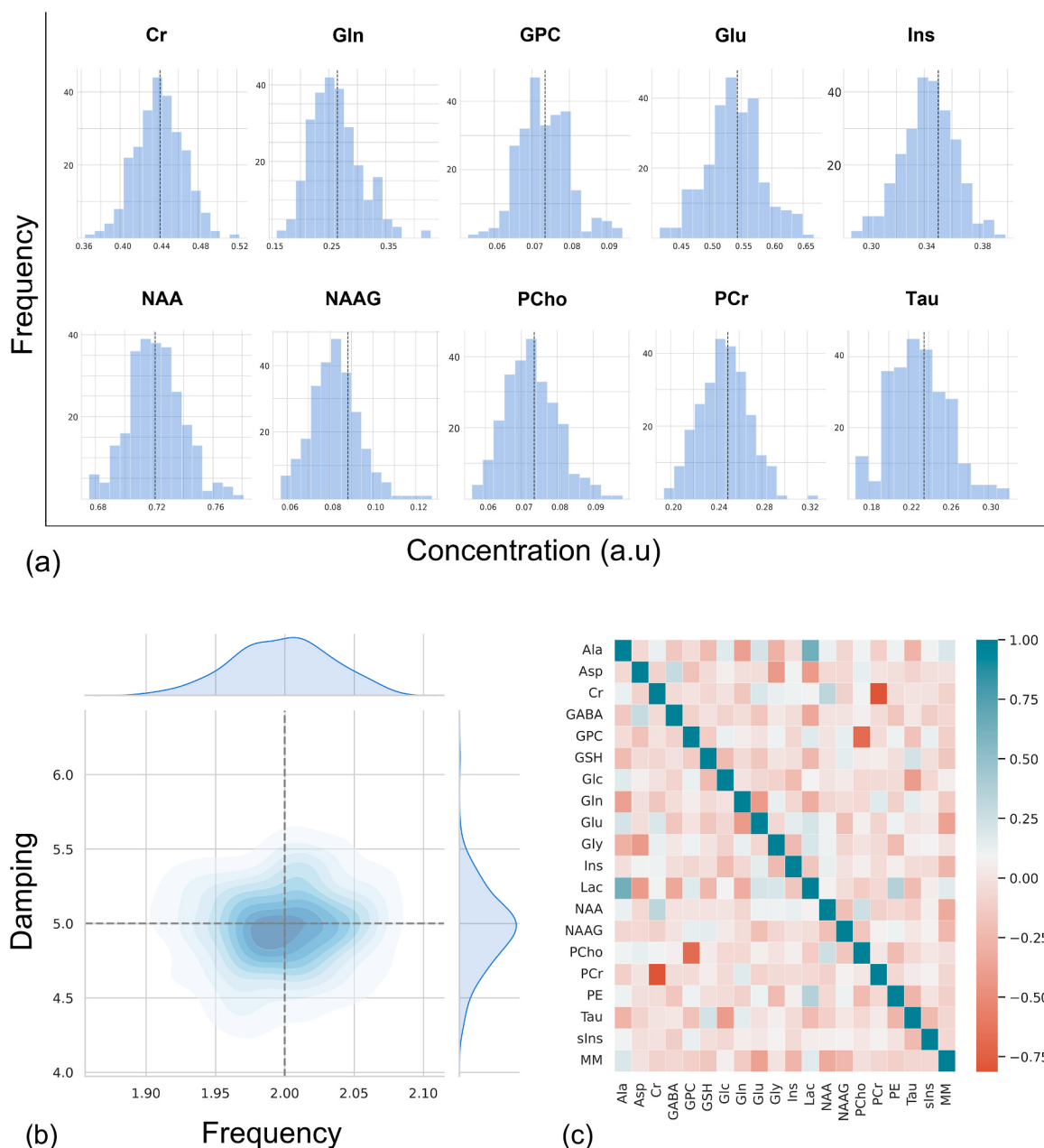
Given the recent agreement on using LCM fitting [3] and several pioneering research highlighting the possible use of DL in MRS [18,25,36,38], this work demonstrates that an LCM fitting-based self-supervised DL approach can be used to quantify both simulated and in-vivo MRS data.

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 Figs. 4 and 8. 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 (Figs. 4 and 8). 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 [5].

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 (Fig. 3, Table 5). 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 (Fig. 3 and Table 5). Noteworthy, DQ-nMM showed remarkable results for metabolites 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 [67].

DQ-rpMM showed better performance than DQ-pMM and QUEST-



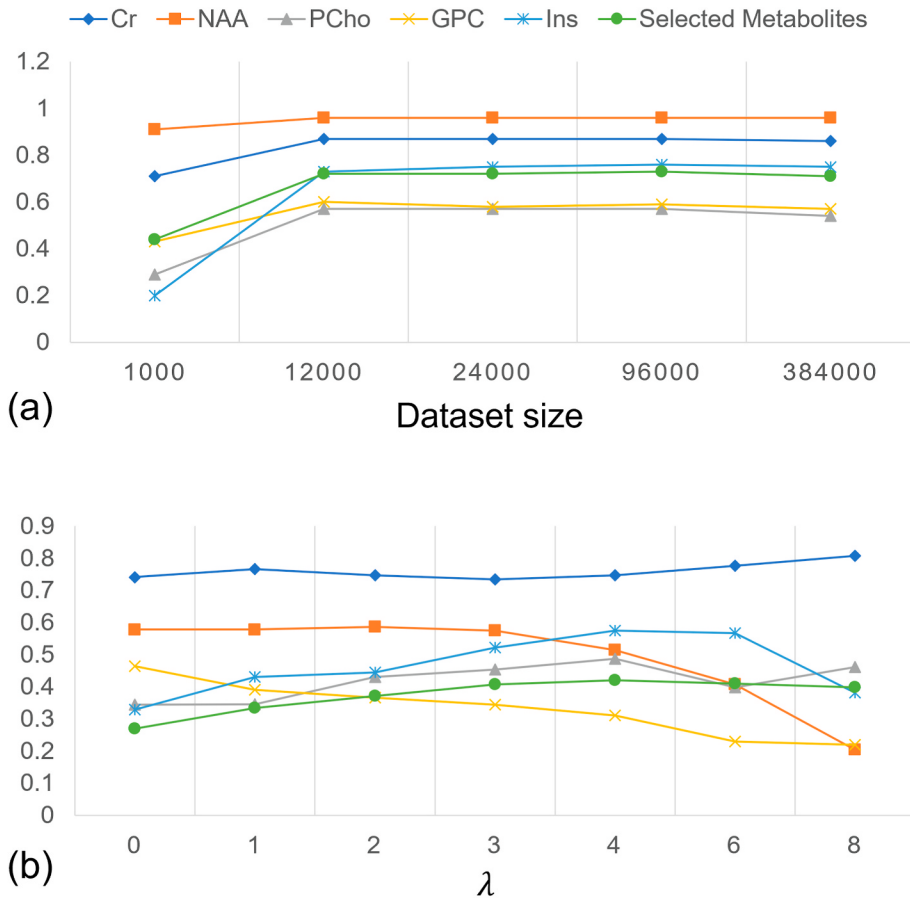
**Fig. 5.** The results of the MC analysis. Spectra are quantified by the DQ-nMM method. (a) Histograms of the estimated concentrations. The vertical lines are true values. (b) A scatter visualization of a joint distribution of estimated frequencies (Hz) and dampings (Hz) of spectra. The dashed vertical 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.

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 5). 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 (Fig. 6).

The main drawback of supervised deep learning quantification of MRS data is (i) the lack of model interpretability and (ii) the need for ground truth data. Introducing prior knowledge and a model function to a deep neural network can provide a self-supervised interpretable deep learning approach to MRS data quantification. Gurbani et al. [38] previously presented a model-informed self-supervised neural network in which the information of peak components was required as prior

knowledge; however, their method depends heavily on prior information of the excessive number of metabolites and peaks per metabolite which is time-consuming in particular for spectra acquired in higher magnetic fields. Our approach incorporates realistic quantum-mechanics simulated metabolite responses, eliminating the need for the information of peak components, e.g., the location of NAA resonance, or any other constraints.

The computational bottleneck in processing volumetric MRSI is accurate quantification, partly because the current LCM approaches rely on iterative algorithms. Since our method showed comparable performance compared with QUEST and FiTAID, once trained, our proposed method can simply be implemented on scanner computers, allowing for fast (10000 spectra in 80 s using a standard computer) and accurate quantification. As reported in Ref. [68], our findings confirm that when



**Fig. 6.** (a) The mean  $R^2$  value calculated across the selected metabolites (NAA, Cr, PCho, GPC, and Ins), and the  $R^2$  value for each of the selected metabolites obtained (a) from the DQ-nMM model trained with datasets with sizes of 1000, 12000, 24000, and 96000, and 384000, (b) from DQ-rpMM trained with different  $\lambda$ . The figure shows that the choice of  $\lambda$  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 PCho, and Cr, and, conversely, a poor performance in the quantification of NAA and GPC. 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;  $\lambda$ , the significance of the regularization term.

a numerical MM pattern is available, methods that are equipped with this type of pattern, such as DQ-nMM, should be utilized due to their good performance.

In general, one of the main weaknesses of DL models is that they cannot perform well on new data acquired using different acquisition parameters. In the case of the same pulse sequence with either longer or shorter TE, it is necessary for our method to be retrained using a new basis set; however, transfer learning methods can mitigate this issue.

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 (Fig. 8 and Table 6).

Our results (Table 6) 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 (Figs. 6 and 7). Our results indicate that the accuracy of our proposed method trained on a dataset with 1000 samples is suboptimal, and a training dataset with a minimum size of 12000 samples is necessary for optimal quantification. As reported by Nakkiran et al. [29], our results verify that increasing the number of training samples reduces test performance. It is interesting since big datasets are rare in MRS applications.

In the field of machine learning, it is typical to estimate the computation cost using FLOPs (floating-point operations per second) [64]. 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 [69]:  $O(n \times m) + k(O(m^3) + O(n \times m) + O(m^3))$  where  $n$  is

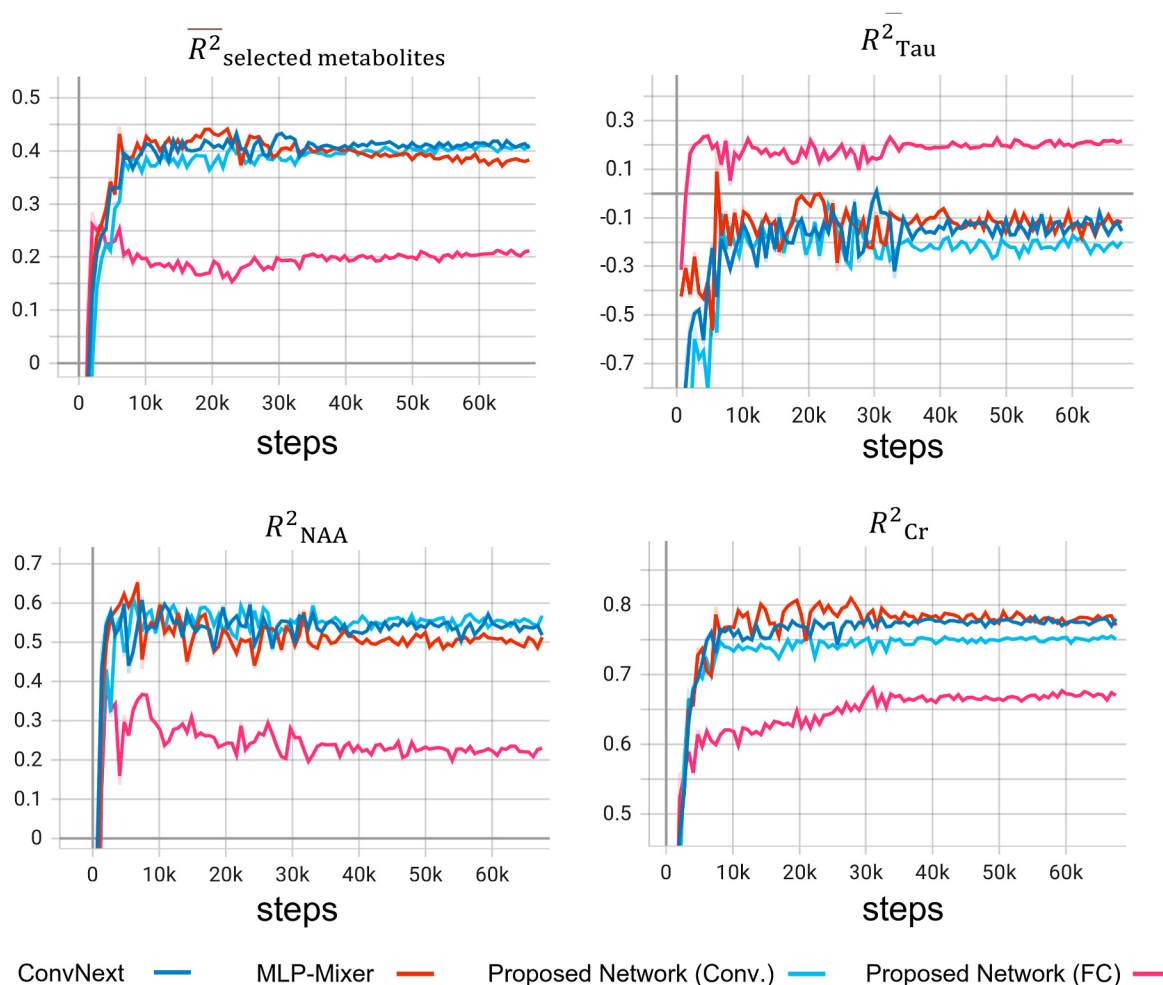
the number of data points and  $m$  is the number of model parameters, and  $k$  is the number of the Levenberg-Marquardt damping parameters that are being used [69]. By setting  $k$  to 1 for the sake of simplicity,  $m$  to 24, and  $n$  to 2048, the total cost of Levenberg-Marquardt at every iteration is 1.4 M FLOPs. For example, FiTAID [53] requires 100 iterations by default, and the computational cost of a fitting procedure using traditional LCM is above 100 MFLOPs.

We investigated the influence of deep neural network design by comparing our proposed network to existing architectures for machine vision tasks. As depicted in Fig. 7, substituting convolutional layers with FC layers resulted in a substantial drop in performance and an increase in computational cost (4.52 MFLOPs). The results (Fig. 7) 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). Our results suggest that (i) networks with simple and shallow design can perform optimal quantification of MRS data and (ii) using convolutional layers are necessary.

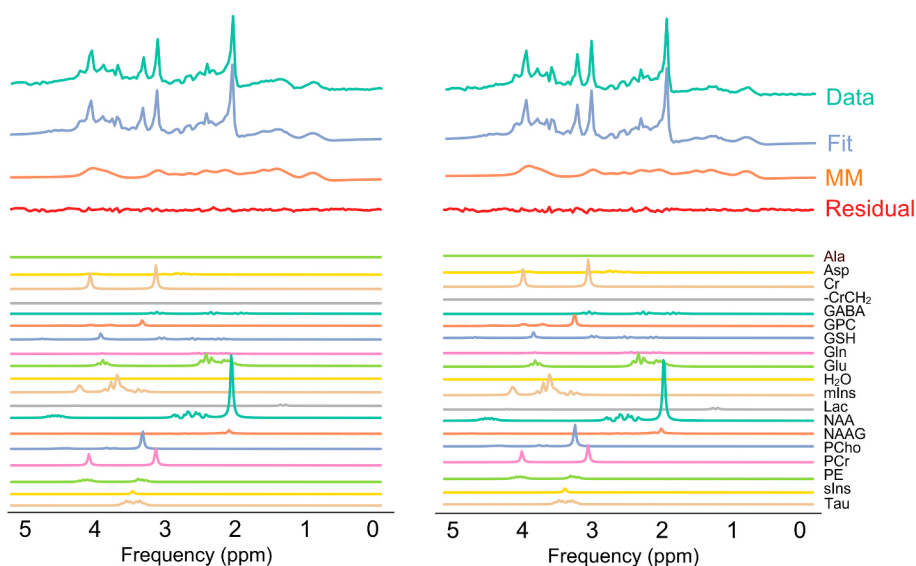
Because of using a model-decoder as the decoder, our proposed method can calculate the Cramér-Rao lower bounds to estimate uncertainty. However, additional work is required to extend the capability of the proposed method for quantifying MRS data with simultaneous uncertainty estimation. Additionally, it is important to evaluate how the distribution of parameters in the training set affects the performance of the network.

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.

While the proposed DL-based method for quantification of MR spectroscopy data shows promising results, there are some limitations to



**Fig. 7.** 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.



**Fig. 8.** Two example spectra 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 and two more example spectra are provided in Supporting Information Figures S4 and S7. GT, ground truth; DQ-rpMM, deep learning-based quantification using parametric MM and a regularization term.



**Table 6**

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.

			Subject #1	Subject #2	Subject #3	Subject #4	Subject #5
[metabolite]/[tCr]	mIns	Mean	1.187	1.078	0.995	1.039	1.1
		SD	0.016	0.032	0.033	0.032	0.028
		CV	0.013	0.03	0.033	0.031	0.026
Tau	Tau	Mean	0.294	0.335	0.377	0.326	0.316
		SD	0.01	0.013	0.02	0.032	0.011
		CV	0.034	0.039	0.052	0.099	0.034
tNAA	tNAA	Mean	1.69	1.69	1.729	1.616	1.656
		SD	0.025	0.017	0.021	0.065	0.043
		CV	0.015	0.01	0.012	0.04	0.026
tCho	tCho	Mean	0.25	0.259	0.205	0.233	0.237
		SD	0.005	0.005	0.022	0.018	0.005
		CV	0.018	0.018	0.106	0.079	0.02
Glx	Glx	Mean	1.28	1.402	1.424	1.288	1.351
		SD	0.016	0.033	0.013	0.073	0.053
		CV	0.012	0.024	0.009	0.057	0.039

consider:

First, the proposed method was validated using simulated and publicly accessible in-vivo human brain MRS data, which may not fully represent the variability and complexity of clinical data. Therefore, further validation with a larger dataset that includes more diverse patient populations is needed to evaluate the generalizability of the method.

Second, the proposed method relies on quantum-mechanics simulated metabolite responses, which may not fully capture the variability of in-vivo metabolite responses. This may result in some inaccuracies in the quantification of metabolite concentrations.

Third, the proposed method is constrained by the LCM model, which may limit its ability to detect and quantify metabolites that are not included in the model. Therefore, the proposed method may not be suitable for the analysis of MRS data that contains metabolites that are not well characterized by the LCM model.

Lastly, the proposed method requires a large amount of training data to optimize the neural network, which may limit its applicability in certain settings where obtaining large datasets is challenging.

## 6. Conclusions

In this study, we presented a self-supervised deep learning technique for accurate quantification of in-vivo and simulated MRS data using linear combination modeling. Our proposed approach eliminates the need for labeled data during training and highlights the contribution of each basis spectrum in the overall spectrum. We compared our method with conventional approaches and observed that it quantified MRS signals with comparable performance but in a significantly shorter amount of time. Moreover, we demonstrated that more complex and deeper architectures did not improve the performance of our shallower architecture.

We proposed a novel approach for MM modeling which increased the performance of our method. Our results indicated that a training dataset with a minimum size of 12000 is required for precise quantification, and larger datasets do not improve quantification accuracy.

In conclusion, our study presents a self-supervised DL-based approach for accurate MRS data quantification, offering a faster alternative to conventional techniques. Our findings could be of significant value in accelerating the quantification of large MRS datasets. Future studies should investigate the clinical relevance of our proposed method.

## 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.

## Data availability statement

The source code is freely available at [<https://github.com/is-i-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/>.

## Declaration of competing interest

The authors have no conflict of interest to declare.

## 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 Skłodowska-Curie grant agreement No 813120 (INSPIRE-MED) and was also supported by institutional support RVO:68081731 - Czech Academy of Sciences, Institute of Scientific Instruments.

This work was supported by the staff and equipment of the ISI-MR facility of the Czech-Biomed infrastructure, supported by grants CZ.02.1.01/0.0/0.0/18\_046/0016045, LM2018129 and LM2023050 of the MEYS CR.

The authors thank Radim Korínek, Ph.D. (Czech Academy of Sciences, Institute of Scientific Instruments, Czech Republic), Reza Goodarzi, Rudy Rizzo, and Mohammad Jafarian for their valuable technical support.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbimed.2023.106837>

## References









- [1] R. Kreis, C. Cudalbu, R.A. De Graaf, C. Gasparovic, A. Heerschap, B. Lanz, J. Near, G. Öz, M. Terpstra, I. Tkáč, M. Wilson, Terminology and Concepts for the Characterization of in Vivo MR Spectroscopy Methods and MR Spectra : Background and Experts ' Consensus Recommendations, 2021, <https://doi.org/10.1002/nbm.4347>.



- [2] Robin A. de Graaf, in: *Vivo NMR Spectroscopy*. Vivo NMR Spectrosc, Wiley Online Books, 2019, <https://doi.org/10.1002/9781119382461.ch1>.
- [3] J. Near, A.D. Harris, C. Juchem, R. Kreis, M. Marjańska, G. Öz, J. Slotboom, M. Wilson, C. Gasparovic, Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations, *NMR Biomed.* e4257 (2020), <https://doi.org/10.1002/nbm.4257>.
- [4] J.B. Poulet, D.M. Sima, S. Van Huffel, MRS signal quantitation: a review of time- and frequency-domain methods, *J. Magn. Reson.* 195 (2008) 134–144, <https://doi.org/10.1016/j.jmr.2008.09.005>.
- [5] C. Stagg, D.L. Rothman, *Magnetic Resonance Spectroscopy: Tools for Neuroscience Research and Emerging Clinical Applications*, Academic Press, 2013.
- [6] J. Near, A.D. Harris, C. Juchem, R. Kreis, M. Marjańska, G. Öz, J. Slotboom, M. Wilson, C. Gasparovic, Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations, *NMR Biomed.* 1–23 (2020), <https://doi.org/10.1002/nbm.4257>.
- [7] K. Landheer, K.M. Swanberg, C. Juchem, Magnetic Resonance Spectrum Simulator (MARSS), a Novel Software Package for Fast and Computationally Efficient Basis Set Simulation 1–13, 2019, <https://doi.org/10.1002/nbm.4129>.
- [8] Z. Starčuk, J. Starčuková, Quantum-mechanical simulations for in vivo MR spectroscopy: principles and possibilities demonstrated with the program NMRScopeB, *Anal. Biochem.* 529 (2017), <https://doi.org/10.1016/j.ab.2016.10.007>.
- [9] Daniel G.Q. Chong, R. Kreis, C.S. Bolliger, C. Boesch, J. Slotboom, 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. Phys. Biol. Med.* 24 (2011) 147–164, <https://doi.org/10.1007/s10334-011-0246-y>.
- [10] W.T. Clarke, C.J. Stagg, S. Jbabdi, FSL-MRS: an end-to-end spectroscopy analysis package, *Magn. Reson. Med.* 85 (2021) 2950–2964, <https://doi.org/10.1002/MRM.28630>.
- [11] G. Oeltzschner, H.J. Zöllner, S.C.N. Hui, M. Mikkelsen, M.G. Saleh, S. Tapper, R.A. Edden, Osprey: open-source processing, reconstruction & estimation of magnetic resonance spectroscopy data, *J. Neurosci. Methods* 343 (2020), <https://doi.org/10.1016/j.jneumeth.2020.108827>.
- [12] S.W. Provencher, Automatic quantitation of localized in vivo 1H spectra with LCModel, *NMR Biomed.* 14 (2001) 260–264, <https://doi.org/10.1002/NBM.698>.
- [13] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt, D. Graveron-Demilly, Time-domain semi-parametric estimation based on a metabolite basis set, *NMR Biomed.* 18 (2005) 1–13, <https://doi.org/10.1002/nbm.895>.
- [14] B.J. Soher, P. Semanchuk, D. Todd, J. Steinberg, K. Young, VesPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis, *J. Magn. Reson., Ser. A* (2011).
- [15] M. Wilson, G. Reynolds, R.A. Kauppinen, T.N. Arvanitis, A.C. Peet, A constrained least-squares approach to the automated quantitation of in vivo 1H magnetic resonance spectroscopy data, *Magn. Reson. Med.* 65 (2011) 1–12, <https://doi.org/10.1002/mrm.22579>.
- [16] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [17] A.S. Lundervold, A. Lundervold, An overview of deep learning in medical imaging focusing on MRI, *Z. Med. Phys.* 29 (2019) 102–127, <https://doi.org/10.1016/j.zemedl.2018.11.002>.
- [18] R. Kreis, S.P. Kyathanahally, Deep learning approaches for detection and removal of ghosting artifacts in, *MR Spectroscopy* 863 (2018) 851–863, <https://doi.org/10.1002/mrm.27096>.
- [19] H. Lee, H.H. Lee, H. Kim, Reconstruction of spectra from truncated free induction decays by deep learning in proton magnetic resonance spectroscopy, *Magn. Reson. Med.* (2020) 1–10, <https://doi.org/10.1002/mrm.28164>.
- [20] P. Klukowski, M. Augoff, M. ZieRba, M. Drwal, A. Gonczarek, M.J. Walczak, NMRNet: a deep learning approach to automated peak picking of protein NMR spectra, *Bioinformatics* 34 (2018) 2590–2597, <https://doi.org/10.1093/bioinformatics/bty134>.
- [21] Z. Iqbal, D. Nguyen, G. Hangel, S. Motyka, W. Bogner, S. Jiang, Super-resolution 1H magnetic resonance spectroscopic imaging utilizing deep learning, *Front. Oncol.* 9 (2019) 1–13, <https://doi.org/10.3389/fonc.2019.01010>.
- [22] Z. Iqbal, D. Nguyen, M.A. Thomas, S. Jiang, Deep learning can accelerate and quantify simulated localized correlated spectroscopy, *Sci. Rep.* 11 (2021) 1–13, <https://doi.org/10.1038/s41598-021-88158-y>.
- [23] A. Shamaei, J. Starčuková, I. Pavlova, Z. Starčuk, Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals, *Magn. Reson. Med.* 89 (2023) 1221–1236, <https://doi.org/10.1002/MRM.29498>.
- [24] S. Tapper, M. Mikkelsen, B.E. Dewey, H.J. Zöllner, S.C.N. Hui, G. Oeltzschner, R.A. Edden, Frequency and phase correction of J-difference edited MR spectra using deep learning, *Magn. Reson. Med.* 85 (2021) 1755–1765, <https://doi.org/10.1002/mrm.28525>.
- [25] S.S. Gurbani, E. Schreibmann, A.A. Maudsley, J.S. Cordova, B.J. Soher, H. Poptani, G. Verma, P.B. Barker, H. Shim, L.A.D. Cooper, A convolutional neural network to filter artifacts in spectroscopic MRI, *Magn. Reson. Med.* 80 (2018) 1765–1775, <https://doi.org/10.1002/mrm.27166>.
- [26] C.F. Higham, D.J. Higham, Deep learning: an introduction for applied mathematicians, *SIAM Rev.* 61 (2019) 860–891, <https://doi.org/10.1137/18M1165748>.
- [27] F. Lam, Y. Li, X. Peng, Constrained magnetic resonance spectroscopic imaging by learning nonlinear low-dimensional models, *IEEE Trans. Med. Imag.* 39 (2020) 545–555, <https://doi.org/10.1109/TMI.2019.2930586>.
- [28] Y. Li, Z. Wang, F. Lam, Separation of metabolite and macromolecule signals for 1 H-mrsi using learned nonlinear models, *Proc. - Int. Symp. Biomed. Imaging* (2020) 1725–1728, <https://doi.org/10.1109/ISBI45749.2020.9098365>, 2020-April.
- [29] P. Nakkiran, G. Kaplun, Y. Bansal, T. Yang, B. Barak, I. Sutskever, Deep double descent: where bigger models and more data hurt, *J. Stat. Mech. Theor. Exp.* 2021 (2019), 124003, <https://doi.org/10.1088/17426596/2021/12/02292>.
- [30] M. Belkin, D. Hsu, S. Ma, S. Mandal, *Reconciling Modern Machine Learning Practice and the Bias-Variance Trade-Off*, 2019.
- [31] M. Mikkelsen, D.L. Rimbault, P.B. Barker, P.K. Bhattacharyya, M.K. Brix, P.F. Buur, K.M. Cecil, K.L. Chan, D.Y.T. Chen, A.R. Craven, K. Cuyppers, M. Dacko, N. W. Duncan, U. Dydak, D.A. Edmondson, G. Ende, L. Erslund, M.A. Forbes, F. Gao, I. Greenhouse, A.D. Harris, N. He, S. Heba, N. Hoggard, T.W. Hsu, J.F.A. Jansen, A. Kangarlu, T. Lange, R.M. Lebel, Y. Li, C.Y.E. Lin, J.K. Liou, J.F. Lirng, F. Liu, J. R. Long, R. Ma, C. Maes, M. Moreno-Ortega, S.O. Murray, S. Noah, R. Noeske, M. D. Noseworthy, G. Oeltzschner, E.C. Porges, J.J. Prisciandaro, N.A.J. Puts, T.P. L. Roberts, M. Sack, N. Sailasuta, M.G. Saleh, M.P. Schallmo, N. Simard, D. Stoffers, S.P. Swinnen, M. Tegenthoff, P. Truong, G. Wang, I.D. Wilkinson, H.J. Wittsack, A. J. Woods, H. Xu, F. Yan, C. Zhang, V. Zipunnikov, H.J. Zöllner, R.A.E. Edden, Big GABA II: water-referenced edited MR spectroscopy at 25 research sites, *Neuroimage* 191 (2019) 537–548, <https://doi.org/10.1016/j.neuroimage.2019.02.059>.
- [32] M. Mikkelsen, P.B. Barker, P.K. Bhattacharyya, M.K. Brix, P.F. Buur, K.M. Cecil, K. L. Chan, D.Y.T. Chen, A.R. Craven, K. Cuyppers, M. Dacko, N.W. Duncan, U. Dydak, D.A. Edmondson, G. Ende, L. Erslund, F. Gao, I. Greenhouse, A.D. Harris, N. He, S. Heba, N. Hoggard, T.W. Hsu, J.F.A. Jansen, A. Kangarlu, T. Lange, R.M. Lebel, Y. Li, C.Y.E. Lin, J.K. Liou, J.F. Lirng, F. Liu, R. Ma, C. Maes, M. Moreno-Ortega, S. O. Murray, S. Noah, R. Noeske, M.D. Noseworthy, G. Oeltzschner, J.J. Prisciandaro, N.A.J. Puts, T.P.L. Roberts, M. Sack, N. Sailasuta, M.G. Saleh, M.P. Schallmo, N. Simard, S.P. Swinnen, M. Tegenthoff, P. Truong, G. Wang, I.D. Wilkinson, H. J. Wittsack, H. Xu, F. Yan, C. Zhang, V. Zipunnikov, H.J. Zöllner, R.A.E. Edden, Big GABA: edited MR spectroscopy at 24 research sites, *Neuroimage* 159 (2017) 32–45, <https://doi.org/10.1016/j.neuroimage.2017.07.021>.
- [33] Y. Hiltunen, J. Kaartinen, J. Pulkkinen, A.M. Häkkinen, N. Lundbom, R. A. Kauppinen, Quantification of human brain metabolites from in vivo 1H NMR magnitude spectra using automated artificial neural network analysis, *J. Magn. Reson.* 154 (2002) 1–5, <https://doi.org/10.1006/JMRE.2001.2457>.
- [34] H. Bhat, B.R. Sajja, P.A. Narayana, Fast quantification of proton magnetic resonance spectroscopic imaging with artificial neural networks, *J. Magn. Reson.* 183 (2006) 110–122, <https://doi.org/10.1016/j.jmr.2006.08.004>.
- [35] N. Hatami, M. Sdika, H. Ratiney, Magnetic resonance spectroscopy quantification using deep learning, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 11070 LNCS (2018) 467–475, [https://doi.org/10.1007/978-3-030-00928-1\\_53](https://doi.org/10.1007/978-3-030-00928-1_53).
- [36] H.H. Lee, H. Kim, Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain, *Magn. Reson. Med.* 82 (2019) 33–48, <https://doi.org/10.1002/mrm.27727>.
- [37] M. Chandler, C. Jenkins, S.M. Shermer, F.C. Langbein, MRSNet: metabolite quantification from edited magnetic resonance spectra with convolutional neural networks, *arXiv* 1–12, <https://doi.org/10.48550/arXiv.1909.03836>, 2019.
- [38] S.S. Gurbani, S. Sheriff, A.A. Maudsley, H. Shim, L.A.D. Cooper, Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting, *Magn. Reson. Med.* 81 (2019) 3346–3357, <https://doi.org/10.1002/mrm.27641>.
- [39] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 80 (313) (2006) 504–507, [https://doi.org/10.1126/SCIENCE.1127647/SUPPL\\_FILE/HINTON.SOM.PDF](https://doi.org/10.1126/SCIENCE.1127647/SUPPL_FILE/HINTON.SOM.PDF).
- [40] A. Shamaei, J. Starčuková, Z.S. Zenon Starčuk, A wavelet scattering convolutional network for magnetic resonance spectroscopy signal quantitation, in: *BIO SIGNALS 2021 - 14th International Conference on Bio-Inspired Systems and Signal Processing*; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021, SciTePress, 2021, pp. 268–275, <https://doi.org/10.5220/0010318502680275>.
- [41] R. Rizzo, M. Dziadosz, S.P. Kyathanahally, A. Shamaei, R. Kreis, Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias, *Magn. Reson. Med.* (2022), <https://doi.org/10.1002/mrm.29561>.
- [42] A. Shamaei, *Deep learning for magnetic resonance spectroscopy quantification: a time-frequency analysis approach*, in: V. Novak (Ed.), *Proceedings II of the 26th Conference STUDENT EEICT 2020: Selected Papers [Online]*, BRNO UNIV TECHNOLOGY, FAC ELECTRICAL ENG & COMMUNICATION, Brno, 2020, pp. 131–135.
- [43] M. Marjańska, D.K. Deelchand, R. Kreis, J.R. Alger, P.J. Bolan, T. Borbath, F. Boumezeur, C.C. Fernandes, E. Coello, B.H. Nagraja, M. Považan, H. Ratiney, D. Sima, J. Starčuková, B.J. Soher, M. Wilson, J.J.A. van Asten, Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM, *Magn. Reson. Med.* 87 (2022) 11–32, <https://doi.org/10.1002/mrm.28942>.
- [44] V. Govindaraju, K. Young, A.A. Maudsley, Proton NMR chemical shifts and coupling constants for brain metabolites, *NMR Biomed.* 13 (2000) 129–153, [https://doi.org/10.1002/1099-1492\(200005\)13:3<129::AID-NBM619>3.0.CO;2-V](https://doi.org/10.1002/1099-1492(200005)13:3<129::AID-NBM619>3.0.CO;2-V).
- [45] J. Near, R. Edden, C.J. Evans, A. Harris, P. Jezard, Frequency and phase drift correction of magnetic resonance spectroscopy data by spectral registration in the time domain, *Magn. Reson. Med.* 50 (2015) 44–50, <https://doi.org/10.1002/mrm.25094>.
- [46] R. Simpson, G.A. Devenyi, P. Jezard, T.J. Hennessy, J. Near, Advanced processing and simulation of MRS data using the FID appliance (FID-A)—an open source, MATLAB-based toolkit, *Magn. Reson. Med.* 77 (2017) 23–33, <https://doi.org/10.1002/mrm.26091>.

- [47] E. Cabanes, Y. Le Fur, G. Simond, P.J. Cozzone, Optimization of residual water signal removal by HLSVD on simulated short echo time proton, MR Spectra of the Human Brain 125 (2001) 116–125, <https://doi.org/10.1006/jmre.2001.2318>.
- [48] K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybern. 36 (1980) 193–202, <https://doi.org/10.1007/BF00344251>.
- [49] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Network. 61 (2015) 85–117, <https://doi.org/10.1016/J.NEUNET.2014.09.003>.
- [50] D. Hendrycks, K. Gimpel, GAUSSIAN ERROR LINEAR UNITS (GELUS), 2020, <https://doi.org/10.48550/arXiv.1606.08415> arXiv 1–9.
- [51] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: an imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019).
- [52] C. Cudalbu, K.L. Behar, P.K. Bhattacharyya, W. Bogner, T. Borbath, R.A. de Graaf, R. Gruetter, A. Henning, C. Juchem, R. Kreis, P. Lee, H. Lei, M. Marjanika, R. Mekle, S. Murali-Manohar, M. Považan, V. Rackayová, D. Simicic, J. Slotboom, B.J. Soher, Z. Starčuk, J. Starčuková, I. Tkáč, S. Williams, M. Wilson, A.M. Wright, L. Xin, V. Mlynárik, Contribution of macromolecules to brain 1H MR spectra: experts' consensus recommendations, NMR Biomed. 34 (2021) 1–24, <https://doi.org/10.1002/nbm.4393>.
- [53] D.G.Q. Chong, R. Kreis, C. Bolliger, C. Boesch, J. Slotboom, 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. Phys. 24 (2011) 147–164.
- [54] P. Lightning, PyTorch lightning [WWW Document]. GitHub. Note, <https://doi.org/10.5281/zenodo.3828935>, 2020.
- [55] S. Falkner, A. Klein, F. Hutter, BOHB: Robust and Efficient Hyperparameter Optimization at Scale, 2018.
- [56] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J.E. Gonzalez, I. Stoica, Tune: A Research Platform for Distributed Model Selection and Training, 2018.
- [57] D.P. Kingma, J.L. Ba, Adam: a method for stochastic optimization, in: 3rd International Conference on Learning Representations. International Conference on Learning Representations, ICLR, 2015.
- [58] I. Loshchilov, F. Hutter, SGDR: stochastic gradient descent with warm restarts, 5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc. (2016), <https://doi.org/10.48550/arXiv.1608.03983>.
- [59] W.T. Clarke, T.K. Bell, U.E. Emir, M. Mikkelsen, G. Oeltzschner, A. Shamaei, B. J. Soher, M. Wilson, NIFTI-MRS: a standard data format for magnetic resonance spectroscopy, Magn. Reson. Med. 88 (2022) 2358–2370, <https://doi.org/10.1002/MRM.29418>.
- [60] L. Vanhamme, A. Van Den Boogaart, S. Van Huffel, Improved method for accurate and efficient quantification of MRS data with use of prior knowledge, J. Magn. Reson. 129 (1997) 35–43, <https://doi.org/10.1006/jmre.1997.1244>.
- [61] H.J. Zöllner, M. Považan, S.C.N. Hui, S. Tapper, R.A.E. Edden, G. Oeltzschner, Comparison of different linear-combination modeling algorithms for short-TE proton spectra, NMR Biomed. 34 (2021), <https://doi.org/10.1002/nbm.4482>.
- [62] R. Birch, A.C. Peet, H. Dehghani, M. Wilson, Influence of macromolecule baseline on 1H MR spectroscopic imaging reproducibility, Magn. Reson. Med. 77 (2017) 34–43, <https://doi.org/10.1002/mrm.26103>.
- [63] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Comput. Sci. 7 (2021) 1–24, <https://doi.org/10.7717/PEERJ-CS.623>.
- [64] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, F.A. Research, A ConvNet for the 2020s, 2022, <https://doi.org/10.48550/arXiv.2201.03545> arXiv.
- [65] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, A. Dosovitskiy, MLP-mixer: an All-MLP Architecture for Vision, 2021, <https://doi.org/10.48550/arXiv.2105.01601> arXiv.
- [66] D. Stefan, F. Di Cesare, A. Andrasescu, E. Popa, A. Lazariev, E. Vescovo, O. Strbak, S. Williams, Z. Starčuk, M. Cabanas, D. Van Ormondt, D. Graveron-Demilly, Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package, Meas. Sci. Technol. 20 (2009), <https://doi.org/10.1088/0957-0233/20/10/104035>.
- [67] R.A.E. Edden, M.G. Pomper, P.B. Barker, In vivo differentiation of N-acetyl aspartyl glutamate from N-acetyl aspartate at 3 Tesla, Magn. Reson. Med. 57 (2007) 977–982, <https://doi.org/10.1002/MRM.21234>.
- [68] D. Simicic, T. Borbath, C. Cudalbu, L. Xin, J. Starčuková, I. Tkáč, B. Lanz, In Vivo Macromolecule Signals in Rat Brain 1 H- MR Spectra, 2021, <https://doi.org/10.1002/mrm.28910> at 9. 4T: Parametrization, spline baseline estimation, and T 2 relaxation times 2384–2401.
- [69] Y. Lin, D. O'Malley, V.V. Vesselinov, A computationally efficient parallel Levenberg-Marquardt algorithm for highly parameterized inverse model analyses, Water Resour. Res. 52 (2016) 6948–6977, <https://doi.org/10.1002/2016WR019028>.

# NIfTI-MRS: A standard data format for magnetic resonance spectroscopy

William T. Clarke<sup>1</sup>   | Tiffany K. Bell<sup>2,3,4</sup> | Uzay E. Emir<sup>5,6</sup>  | Mark Mikkelsen<sup>7</sup>  | Georg Oeltzschner<sup>8,9</sup>  | Amirmohammad Shamaei<sup>10,11</sup>  | Brian J. Soher<sup>12</sup>  | Martin Wilson<sup>13</sup> 

<sup>1</sup>Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

<sup>2</sup>Department of Radiology, University of Calgary, Calgary, Alberta, Canada

<sup>3</sup>Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada

<sup>4</sup>Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada

<sup>5</sup>School of Health Sciences, Purdue University, West Lafayette, Indiana, USA

<sup>6</sup>Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana, USA

<sup>7</sup>Department of Radiology, Weill Cornell Medicine, New York, New York, USA

<sup>8</sup>Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

<sup>9</sup>F. M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, Maryland, USA

<sup>10</sup>Czech Academy of Sciences, Institute of Scientific Instruments, Brno, Czech Republic

<sup>11</sup>Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic

<sup>12</sup>Center for Advanced MR Development, Department of Radiology, Duke University Medical Center, Durham, North Carolina, USA

<sup>13</sup>Centre for Human Brain Health and School of Psychology, University of Birmingham, Birmingham, United Kingdom

## Correspondence

William T. Clarke, Wellcome Centre for Integrative Neuroimaging, FMRIB, University of Oxford, Level 0, John Radcliffe Hospital, Oxford, OX3 9DU, UK. Email: [william.clarke@ndcn.ox.ac.uk](mailto:william.clarke@ndcn.ox.ac.uk)

## Funding information

European Union; Marie Skłodowska-Curie, Grant/Award Number: 813120; National Institutes of Health, Grant/Award Numbers: K99/R00 AG062230, P41 EB015909, P41 EB031771, R01 EB016089, R01 EB023963, R01 EB028259, S10 OD021648; Wellcome Trust, Grant/Award Numbers: 102584/Z/13/Z, 203139/Z/16/Z

**Purpose:** Multiple data formats in the MRS community currently hinder data sharing and integration. NIfTI-MRS is proposed as a standard spectroscopy data format, implemented as an extension to the Neuroimaging informatics technology initiative (NIfTI) format. This standardized format can facilitate data sharing and algorithm development as well as ease integration of MRS analysis alongside other imaging modalities.

**Methods:** A file format using the NIfTI header extension framework incorporates essential spectroscopic metadata and additional encoding dimensions. A detailed description of the specification is provided. An open-source command-line conversion program is implemented to convert single-voxel and spectroscopic imaging data to NIfTI-MRS. Visualization of data in NIfTI-MRS is provided by development of a dedicated plugin for FSLeaves, the FMRIB Software Library (FSL) image viewer.

**Results:** Online documentation and 10 example datasets in the proposed format are provided. Code examples of NIfTI-MRS readers are implemented in common programming languages. Conversion software, *spec2nii*, currently converts 14 formats where data is stored in image-space to NIfTI-MRS, including Digital

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Magnetic Resonance in Medicine* published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.



Imaging and Communications in Medicine (DICOM) and vendor proprietary formats.

**Conclusion:** NIFTI-MRS aims to solve issues arising from multiple data formats being used in the MRS community. Through a single conversion point, processing and analysis of MRS data are simplified, thereby lowering the barrier to use of MRS. Furthermore, it can serve as the basis for open data sharing, collaboration, and interoperability of analysis programs. Greater standardization and harmonization become possible. By aligning with the dominant format in neuroimaging, NIFTI-MRS enables the use of mature tools present in the imaging community, demonstrated in this work by using a dedicated imaging tool, FSLeves, for visualization.

#### KEYWORDS

MRS, MRSI, open data format, spectroscopy, visualization

## 1 | INTRODUCTION

MRS is a highly flexible technique that can generate a wide range of sensitive and specific imaging contrasts complementary to the typically water-derived contrasts of MRI. MRS allows simultaneous measurement of multiple *in vivo* metabolite concentrations that, when combined with additional dynamic contrast-encoding techniques, can be used to derive metabolite concentration time courses, metabolite diffusion properties, or *in vivo* chemical kinetics.<sup>1–3</sup>

*In vivo* MRS data typically requires a complex data storage array containing multiple dimensions, including spectral frequency (equivalently encoded as a time dimension), spatial encoding, and any dimensions required for “dynamic” encoding. Before preprocessing, additional array dimensions may be required, for example, to separately store signals from multi-channel receive coils, or multiple transients.<sup>4</sup> In addition, any information about the acquisition that is required for interpretation needs to be stored, too, including the transmitter frequency and the signal dwell time, as well as information on the spatial dimensions and orientation of the measurement volume.

To date, there is no established standardized data format for communicating MRS and MRSI data. Whereas a Digital Imaging and Communications in Medicine (DICOM) standard for MRS(I) exists,<sup>5</sup> it is not fully implemented by the vendors. Crucially, DICOM does not provide an intuitive way to store spectroscopic data or a standardized method to encode data that requires two or more additional encoding dimensions. As a result, vendors have developed their own separate, proprietary (closed) formats to store raw and processed spectroscopic data. The specific format used is dependent on the scanner software version and local practice. These proprietary formats vary

greatly in the degree in which data has been subjected to inline processing, that is, whether data has been spatially reconstructed, which (and how) metadata is stored, and how the storage is formatted.

The lack of a standard data format hinders the use of MRS in several ways:

- 1 It impedes integration with other imaging modalities.

Without standardized encoding of spatial orientation and position, registration with other modalities requires a per-data-format solution. This hinders both co-analysis of spectroscopic data with other modalities, and leveraging of other modalities in reconstruction and processing of the typically low-signal MRS.

- 2 It impedes the consistent analysis of data.

Any spectroscopic analysis program must implement dedicated interpretation modules for each format. Development of new modules is time consuming and relies on expert knowledge of a format. This results in incomplete coverage of the formats by any one analysis pipeline. Thus, consistent (and comparable) analysis is prevented.

- 3 It impedes the creation of new, discrete, specialized tools.

Without a standard storage format, processing and analysis of spectroscopic data often occurs in a single monolithic process, frequently relying on local analysis pipelines and depending on the local MRS expert to create them. This impedes modular processing, which is important for development and uptake of new acquisition and analysis methods, as well as improving existing ones. The barrier to creating new tools is high if implementation of all processing steps is required. Furthermore, quantifying improvements from modifying a discrete step is difficult if it is inseparable from other steps.

#### 4 It raises the difficulty of sharing data.

Compared to other modalities, sharing MRS data is not straightforward. With the diversity of storage formats, especially across platforms, users cannot reliably read and interpret data received from other users. Data-sharing repositories are required to handle mixed data types, which are processed to varying levels and have varying metadata. This creates enormous sharing friction and discourages researchers from sharing. Anonymization tools required for ethical public sharing of data require per-format implementation.

Combined, these factors significantly raise the barrier to adoption and use of MRS, especially for nonexpert users. Compared to other MRI modalities, MRS analysis workflows remain highly customized and specific and require unique MRS expertise on site.

To address these issues, we propose a single data format based on the Neuroimaging Informatics Technology Initiative (NIfTI) format<sup>6</sup> for storing single-voxel, spectroscopic imaging, and unlocalized spectral data. We call the proposed format *NIfTI-MRS*. The NIfTI file format is the standard format for storing anatomical, functional, diffusion, and quantitative MRI and arterial spin labeling data in the MR neuroimaging research domain. NIfTI has provided a cornerstone for analysis across neuroimaging, allowing integrated analysis of modalities.

The proposed format will:

- provide a simplified pathway from scanner to final analysis;
- enable interoperability and modularity of analysis programs;
- enable easier display and co-interpretation with other modalities, and
- establish a format for data sharing.

In addition, the proposed format is designed to provide a simple anonymization procedure and flexible storage of meta-data, further removing friction from the process of sharing.

To facilitate adoption, this work describes the implementation of an open-source command-line conversion program capable of converting many original formats to the proposed format. The program, *spec2nii*, provides single-point conversion for all spatially reconstructed data (including single-voxel data) from 14 different formats alongside anonymization scripts and manual editing tools for NIfTI-MRS.

By aligning MRS with the most widely used neuroimaging format, NIfTI-MRS will also allow researchers to create comprehensive MRS research designs that incorporate different modalities with ease. To demonstrate this,

we have created an FSleyes (FMRIB Software Library, University of Oxford, Oxford, UK)<sup>7</sup> plugin to visualize multi-dimensional NIfTI-MRS data alongside structural MRI and results from MRS fitting.

## 2 | METHODS

### 2.1 | The NIfTI-MRS data format

A brief description of the proposed standard (version 0.6) is included here. The full standard is provided as Supporting Information, and the latest version can be found in Ref. 8.

#### 2.1.1 | Design

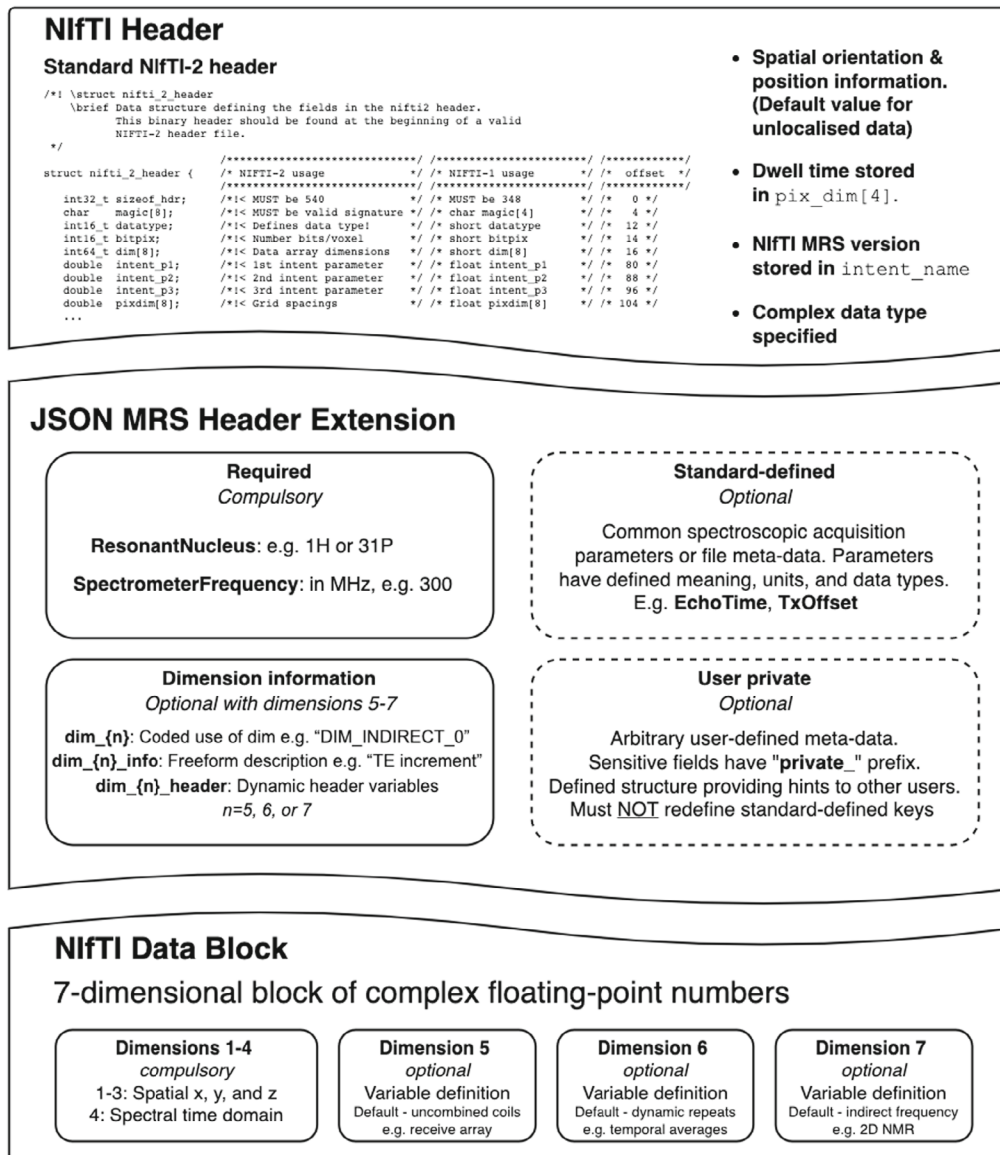
The NIfTI-MRS format is designed to contain single-voxel, contiguous multi-voxel, and unlocalized (or partially localized) time-domain MRS data in up to three spatial dimensions (i.e., MRSI-encoded data). Optionally, NIfTI-MRS can encode up to three additional data dimensions, for example, for arrays of interrelated signals. The standard is designed with low minimum-conformance metadata requirements to simplify adoption, while providing for more complicated metadata requirements in the full format.

The NIfTI format contains three sections: the data header, optional header extensions, and the data block. The proposed NIfTI-MRS format comprises a NIfTI-2-formatted file with a mandatory header extension formatted according to the JavaScript Object Notation (JSON) standard.<sup>9,10</sup>

In NIfTI-MRS (Figure 1), the NIfTI data header<sup>11,12</sup> structure is identical to the one used for structural or functional MRI data, although some values are constrained or re-utilized for spectroscopy-specific purposes. For example, the dwell time of the time domain data is stored in *pixdim*<sup>4</sup> (with units are specified in the *xyz\_t\_units* field). Spatial position encoding, that is, dimensions and orientation of the measurement volume, is implemented as in the NIfTI specification, and a default value of *pixdim* is specified for unlocalized data. Finally, only complex datatypes may be specified in the *datatype* field (e.g., “DT\_COMPLEX”, 32). The NIfTI-MRS standard is versioned, and the version is specified in the *intent\_name* field (in the format *mrs\_v{major}\_{minor}*).

The NIfTI-MRS data block is used to store up to 7D complex time-domain data. The first four dimensions are required for a valid NIfTI-MRS file: the first three dimensions are used for spatial encoding (*x*, *y*, and *z* coordinates), and the fourth dimension is used to store the time-domain





**FIGURE 1** Schematic representation of the proposed NifTI-MRS format. The format utilizes the native NifTI header and data block, whereas using a JSON-formatted NifTI header extension to store additional required metadata. The NifTI data block encodes spatial dimensions (dimensions 1–3), a time domain (dimension 4), and up to three further dimensions (total limited to seven by the NIFTI format definition). By default, the additional dimensions encode uncombined receive coil information (dimension 5), repeated measurements (dimension 6), and an “indirect” frequency dimension, that is, for 2D-NMR (dimension 7). The purpose of these dimensions can be explicitly coded and changed by using the “dim\_...” keys in the header extension. JSON, JavaScript object notation.

FID (or echo). All three spatial dimensions have a size of one for single-voxel data. The NifTI(-MRS) format is only suitable to store spatially reconstructed data; that is, data that has been reconstructed from the acquired  $k(t)$ -space representation. The remaining three dimensions (fifth, sixth, and seventh) are optional. They flexibly encode different dynamic aspects, with the specific purpose and interpretation of each dimension being documented in the header extension. NifTI-MRS is limited to seven dimensions by the definition of the parent NifTI format. Modification of the data storage format to extend

the number of dimensions would break compatibility with existing imaging software tools.

The header extension has the official NifTI identification code 44, “NIFTI\_ECODE\_MRS.” It comprises key-value pairs formatted according to the JSON standard, which can be arbitrarily nested, if necessary. The header extension contains the minimum necessary metadata required for meaningful interpretation of the spectroscopic data, additional information about the optional higher encoding dimensions, and further MRS-specific metadata.

### 2.1.2 | Header extension metadata

The header extension contains 4 types of metadata key:

- 1 The two mandatory keys *SpectrometerFrequency* and *ResonantNucleus*. These must be present in all NIfTI-MRS formatted files because they are a necessary requirement for correct reconstruction and interpretation of the time-domain data.
- 2 Higher encoding dimension information. Three keys are defined per additional encoding dimension ( $n = 5, 6, 7$ ): *dim\_{n}*, *dim\_{n}\_info*, and *dim\_{n}\_header*. The first, *dim\_{n}*, is mandatory if the dimension is used. *Dim\_{n}* takes the value of a list of predefined tags (e.g., *DIM\_COIL* or *DIM\_EDIT*) to identify the purpose of the dimension. A *DIM\_USER\_{0-2}* tag allows for user-specified purposes, which can be described using the *dim\_{n}\_info* field. The *dim\_{n}\_header* tag enables each element of a dimension to be associated with different values of metadata keys.
- 3 Standard-defined metadata. These keys correspond to well-defined and frequently used sequence, hardware, or subject data. They are defined in the NIfTI-MRS standard and may not be redefined. These keys are optional.
- 4 User-defined metadata. Keys can be arbitrarily defined by users to store unusual metadata not foreseen in the standard or for which no fixed format or recommendations exist. User-defined JSON metadata permits additional fields for user-defined keys, encouraging in-place documentation to aid interpretation of the keys. These keys are optional.

### 2.1.3 | Spatial information

Orientation, position, and voxel size information are stored in the NIfTI-MRS header in accordance with the NIfTI standard. NIfTI encodes the spatial information using one, or both, of two ( $4 \times 4$ ) affine matrices: *qform* and *sform*. The former is stored using a series of header keys (*pixdim*, *quatern\_b*, *quatern\_c*, *quatern\_d*, *qoffset\_x*, *qoffset\_y*, *qoffset\_z*, *qfac*), whereas the latter is stored directly. These two affine matrices each encode the relationship between the data stored in the 3D spatial grid and a meaningful coordinate system. The most common coordinate system is that of scanner-anatomical (*qsform\_code* = 1), that is, relative to the position of the subject in the scanner; however, it can also be relative to just the physical scanner, the scanner table, or a defined standard interpretation space (*qsform\_code* = 4), for example, “MNI\_152.”<sup>13</sup> Two different coordinate systems can be encoded by using each affine matrix. NIfTI-MRS follows the original NIfTI standard and emphasizes the use of *qform* as the default.

As such, for NIfTI-MRS, conformance is achieved in the header either when:

- 1 *qform\_code* is set  $> 0$ ,
  - 2 The second to fourth elements of *pixdim* are set to the appropriate voxel dimensions, or to a default of 10 m,
  - 3 *quatern\_b*, *quatern\_c*, *quatern\_d*, *qoffset\_x*, *qoffset\_y*, *qoffset\_z* are set,
  - 4 A valid value of *qfac* is set.
- Or:
- 1 *qform\_code* is set = 0 (*NIfTI\_XFORM\_UNKNOWN*),
  - 2 The second to fourth elements of *pixdim* are set to the appropriate voxel dimensions or to a default of 10 m.

The former *pixdim* option is suitable for data that has a meaningful spatial position, and the latter for data with no real-world position, for example, simulated data. Use of the default *pixdim* size (10 m) indicates a dimension has no unlocalization, or that it has poorly defined extent (e.g., unlocalization provided only by the limited extent of coil sensitivity).

The NIfTI format in general (and therefore also NIfTI-MRS) cannot store spatially noncontiguous data (i.e., data with a gap between voxels or slices) in a single file. Distinct contiguous volumes need to be stored separately.

### 2.1.4 | Anonymization of protected health information

To ease the process of anonymization, all standard-defined metadata keys are marked as privacy-sensitive or not privacy-sensitive in the standard. User-defined metadata may be self-marked as privacy-sensitive by appending “*private\_*” to the start of the key or any nested key within the definition.

Anonymization of data stored as NIfTI-MRS is simplified through two features:

- 1 Only a subset of metadata is retained in the conversion to NIfTI-MRS. This metadata is selected and only incorporates that which is well defined.
- 2 Anonymization tools acting on NIfTI-MRS can reliably identify sensitive fields that have been converted using their definition in the standard.

### 2.1.5 | Processing provenance

Preprocessing steps applied to the data can be optionally recorded in the header extension. The type of preprocessing applied, the program used, the program version, and

any additional information provided by the preprocessing algorithm can be stored sequentially in the “ProcessingApplied” field. Provenance is not provided by the NIfTI-MRS data standard itself but either requires adequate implementation in relevant software packages or manual addition by users.

### 2.1.6 | Phase and frequency conventions

The NIfTI-MRS standard defines a strict frequency and phase convention for data. This convention follows the conventions of Levitt.<sup>14</sup> In this convention, the absolute frequency scale should increase from left to right. For nuclei with a gyromagnetic ratio greater than zero, this corresponds to resonances from nuclei with less shielding (more deshielding), which therefore experience a higher magnetic field, appearing on the left; that is, they have more negative (higher magnitude) Larmor frequencies, noting  $\omega = -\gamma B_0$ . This results in a typically displayed chemical shift (“ppm”) axis increasing from right to left. A description of this convention in the spectral time domain is provided in the specification.

## 2.2 | Software implementation

To promote the use of NIfTI-MRS, the standard has been implemented into software for conversion, visualization, and data input–output (I/O). Functions for loading, writing, and visualizing NIfTI-MRS data have been created for common programming languages and integrated into open-source analysis packages.

### 2.2.1 | Conversion to NIfTI-MRS

An open-source conversion program *spec2nii* has been created. *Spec2nii* reads vendor-proprietary, DICOM, and processing toolbox formats and generates NIfTI-MRS-formatted files. The program can also inspect, edit, and anonymize existing NIfTI-MRS files. The program operates on the command line; is written in Python; and is developed as a public, open-source resource

### 2.2.2 | Visualization using FSLeys

To enable visualization of the multi-dimension NIfTI-MRS format, an FSLeys plugin has been created. FSLeys is the free, open-source FMRI Software Library (FSL) image viewer. The developed plugin extends FSLeys to interpret

the NIfTI-MRS format and enables display of single and multi-voxel spectroscopy stored in the format.

## 3 | RESULTS

### 3.1 | The NIfTI-MRS data format

The specification for the NIfTI-MRS data format is available with this document as Supporting Information and online at Ref. 8. To assist in the interpretation of the standard, both online explanatory documentation ([wtclarke.github.io/mrs\\_nifti\\_standard/](https://wtclarke.github.io/mrs_nifti_standard/)) and example data<sup>15</sup> have been created. Figure 2 shows extracts of NIfTI-MRS header extensions from 4 of the 10 example datasets. These four examples demonstrate the structure of the header extension, the flagging of privacy-sensitive data, dynamic header values describing different conditions in one of the additional data dimensions (here spectral editing), and records of processing provenance. Additional examples are given in the specification and in the example data (at Ref. 15), including an example of “Fingerprinting” (example\_09.nii.gz) where four different changing acquisition parameters would exceed the number of available data dimensions and do not form a densely filled rectilinear grid. In this case, a single additional dimension (dimension 5) is used, and arrays of each changing parameter are stored in the *dim\_5\_header* field.

### 3.2 | Software implementation

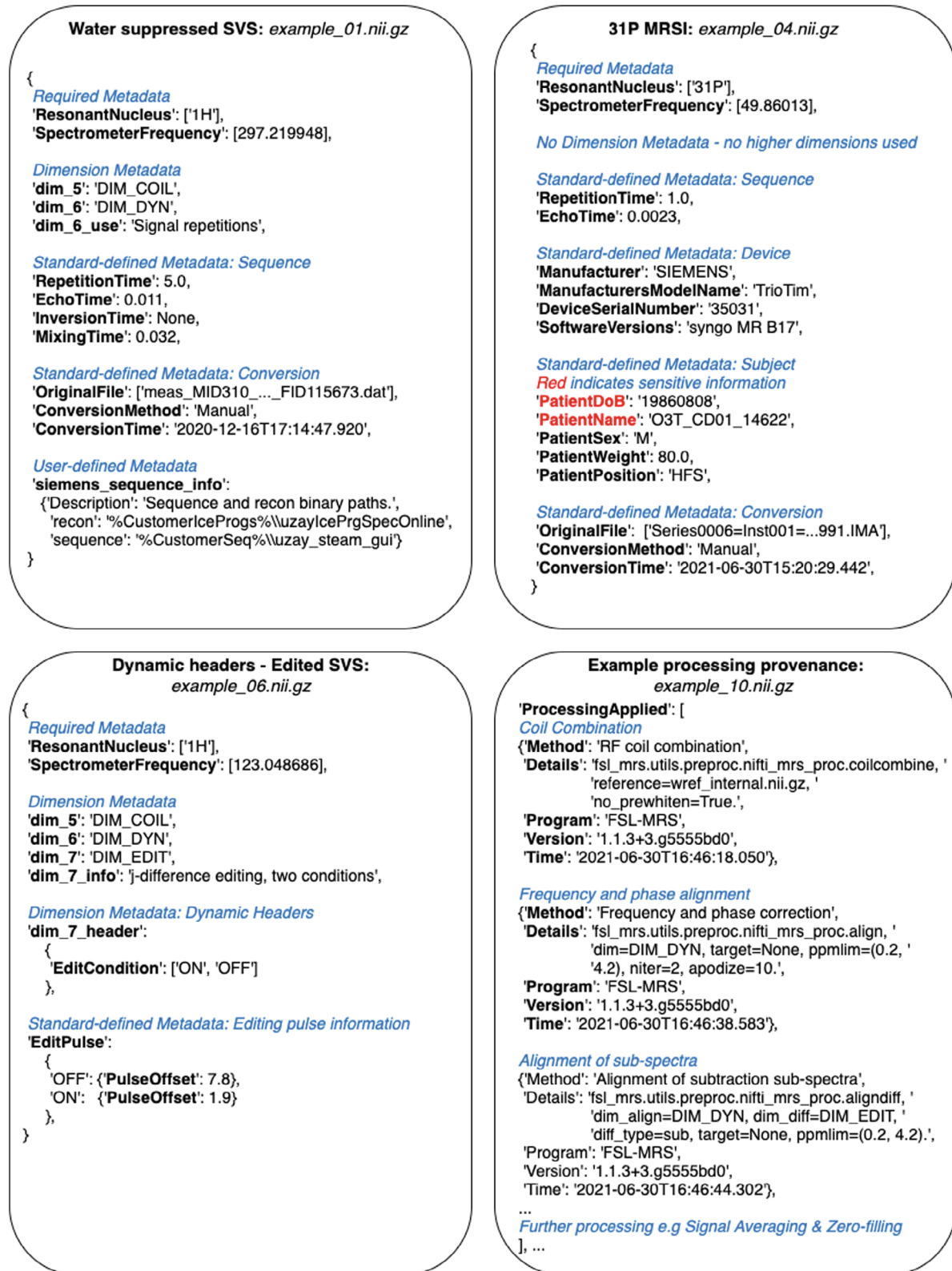
#### 3.2.1 | Conversion to NIfTI-MRS

The conversion program *spec2nii* has been created and released as a Python package. The package is developed online (<https://github.com/wtclarke/spec2nii>) and is available from the PyPI and Conda package managers. *Spec2nii* provides automatic or semi-automatic conversion of 14 data formats (vendor-proprietary and DICOM) to NIfTI-MRS (Table 1).

Figure 3 outlines the proposed use cases of *spec2nii* and NIfTI-MRS enabled by this work:

- 1 Varied input data is passed to *spec2nii*.
- 2 *spec2nii* identifies the appropriate reconstruction and conversion pathway.
- 3 Conversion is carried out, and a NIfTI-MRS file is returned.
- 4 Preprocessing is carried out using NIfTI-MRS as an intermediary format.
- 5 Preprocessed data is stored or further analyzed (fitted).





**FIGURE 2** Extracts of NIfTI-MRS JSON-formatted header extensions for 4 different pieces of example data. The full example data is available from Ref.15. Each example demonstrates a different aspect of the header extension format. Figure annotations are shown as blue italicized text. (A) Structure of a header extension of  $^1\text{H}$  single-voxel data before preprocessing. (B) Header extension for processed  $^{31}\text{P}$  MRSI, including fields that are marked for anonymization (red). (C) Example of dynamic header fields indicating an editing condition stored in the seventh dimension. (D) Extract of the processing provenance in a MEGA-PRESS<sup>16</sup> sequence preprocessed using FSL-MRS<sup>17</sup>.  $^1\text{H}$ , hydrogen-1;  $^{31}\text{P}$ , phosphorus-31; FSL, FMRIB Software Library; MEGA-PRESS, Mescher-Garwood point resolved spectroscopy.

TABLE 1 List of *spec2nii* supported formats in version 0.3.4

Vendor/ Software	Format	File Extension	Spectroscopy Formats Handled			Automatic Orientation	Notes
			SVS	MRSI	FID		
Siemens	Twix	.dat	✓	×	✓	✓	VB and VE baselines
Siemens	DICOM	.ima	✓	✓	✓	✓	
Philips	SPAR/SDAT	.SPAR/.SDAT	✓	×	✓	✓	
Philips	data-list	.data/.list	✓	×	✓	✓	
Philips	DICOM	.dcm	✓	×	✓	WIP	
GE	pfile	.7	✓	✓	✓	✓	Per-sequence mapping required
UIH	DICOM	.dcm	✓	✓	✓	✓	
Bruker	2dseq	–	✓	✓	✓	✓	PV 5.1, 6.0, 6.0.1, 7.0.0, 360
Bruker	fid	–	✓	×	✓	WIP	PV 5.1, 6.0, 6.0.1, 7.0.0, 360
Varian	fid	–	✓	×	✓	WIP	
LCModel	RAW/H2O	.RAW/.H2O	✓	×	✓	×	
jMRUI	Text	.txt	✓	×	✓	×	
jMRUI	MRUI	.mrui	✓	×	✓	×	
–	ASCII	.txt	✓	×	✓	×	

✓ = included fully; FID, free induction decay; WIP = work in progress, X = not yet included (inclusion dependent on the availability of test data and information on interpretation of the data-format). SVS, single voxel spectroscopy

The data at any stage after conversion (stage 3) can be shared, whether that is the unprocessed converted data, fully preprocessed, or partially processed data.

For this work, only the conversion of spatially reconstructed data has been implemented in *spec2nii*. Data that are stored in a k-space representation cannot currently be converted. In the future, some spatially unreconstructed data from standard vendor-supplied MRSI sequences will be handled by *spec2nii*. For data requiring specialist reconstruction, for example, those with non-Cartesian trajectories, we propose a future pathway incorporating conversion to International Society for Magnetic Resonance in Medicine raw data format (ISMRMRD)<sup>18</sup> and third-party reconstruction provided by the sequence developers. To ease use of NIfTI-MRS with custom or “offline” reconstruction routines, we have provided minimal code examples of NIfTI-MRS input/output (I/O) (see next section), which could be used to output NIfTI-MRS directly.

The formats supported by *spec2nii* are summarized in Table 1. *Spec2nii* carries out automatic spatial orientation calculations for 7 of the 14 supported formats.

### 3.2.2 | NIfTI-MRS I/O and compatibility

In Ref.19, minimal examples of NIfTI-MRS file readers have been provided in 4 common programming

languages (Java, MatLab [MathWorks, Natick, MA], Python, and R). The examples exploit the availability of robust NIfTI I/O libraries in those programming languages.

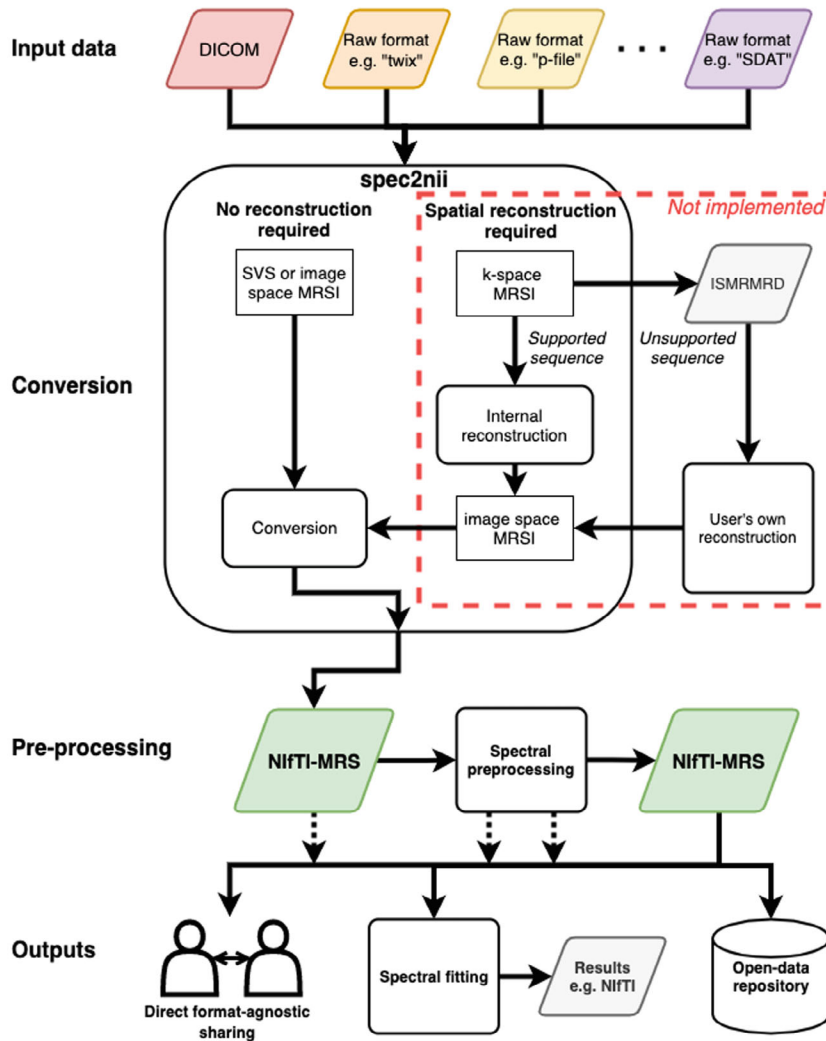
Support for NIfTI-MRS has been established in five open-source spectroscopy analysis packages: FSL-MRS,<sup>17</sup> Osprey,<sup>20</sup> Spant,<sup>21,22</sup> Vespa<sup>23</sup> (<https://github.com/vespa-mrs/vespa>), and FID-A.<sup>24</sup> Each package has I/O compatibility with the standard. Example short-TE STEAM and water reference data at 7 tesla, provided in the NIfTI-MRS format and processed in FSL-MRS, Spant, and Osprey (with Osprey using FID-A as foundation), are shown in Figure 4. The spatial location of the single voxel is shown overlaid on a NIfTI structural image in each package. This demonstrates the benefit for interoperability of different software solutions, and the potential to form the foundation of a mutually compatible, multi-language analysis software ecosystem.

### 3.2.3 | Visualization in FSLeaves

Visualization of NIfTI-MRS-formatted data is possible using FSLeaves and the NIfTI-MRS plugin for FSLeaves. The plugin implements:

- a pannable and zoomable display for spectra;





**FIGURE 3** Proposed MRS and MRSI processing pipelines using NifTI-MRS and incorporating conversion in spec2nii. In the proposed pipeline, raw data from a variety of formats (e.g., DICOM, Siemens “Twix.dat”, GE “p-file,” or Philips “SDAT/SPAR”) are converted to NifTI-MRS using spec2nii. Subsequently preprocessing can be applied, with both the input and results stored in NifTI-MRS. Data can be shared with other users or a data repository in a format-agnostic way at any stage of the preprocessing pipeline. The preprocessed NifTI-MRS file can then be passed on to modeling software. Spec2nii can convert unlocalized, single-voxel, and spatially reconstructed MRSI. In the future, spec2nii will also convert MRSI stored in k-space for certain common supported sequences. For other sequences (e.g., those with non-Cartesian trajectories), a pipeline incorporating the ISMRMRD format<sup>18</sup> and third-party reconstruction is proposed. The red box indicates software yet to be implemented. DICOM, Digital Imaging and Communications in Medicine.

- display of NifTI-MRS headers;
- automatic calculation of the chemical shift axis;
- display of individual spectra stored in the higher dimensions (5th–7th dimensions);
- interactive zeroth and first-order phasing of spectra, and
- easy comparison of spectra from different voxels.

The plugin is maintained at <https://git.fmrib.ox.ac.uk/wclarke/fsleyes-plugin-mrs> and is available from the Pypi and Conda package managers. It operates using the FSleyes interface, is written in python, is open source, and is available under the BSD 3-clause license. FSleyes is available under the Apache License, version 2.0.<sup>7</sup>

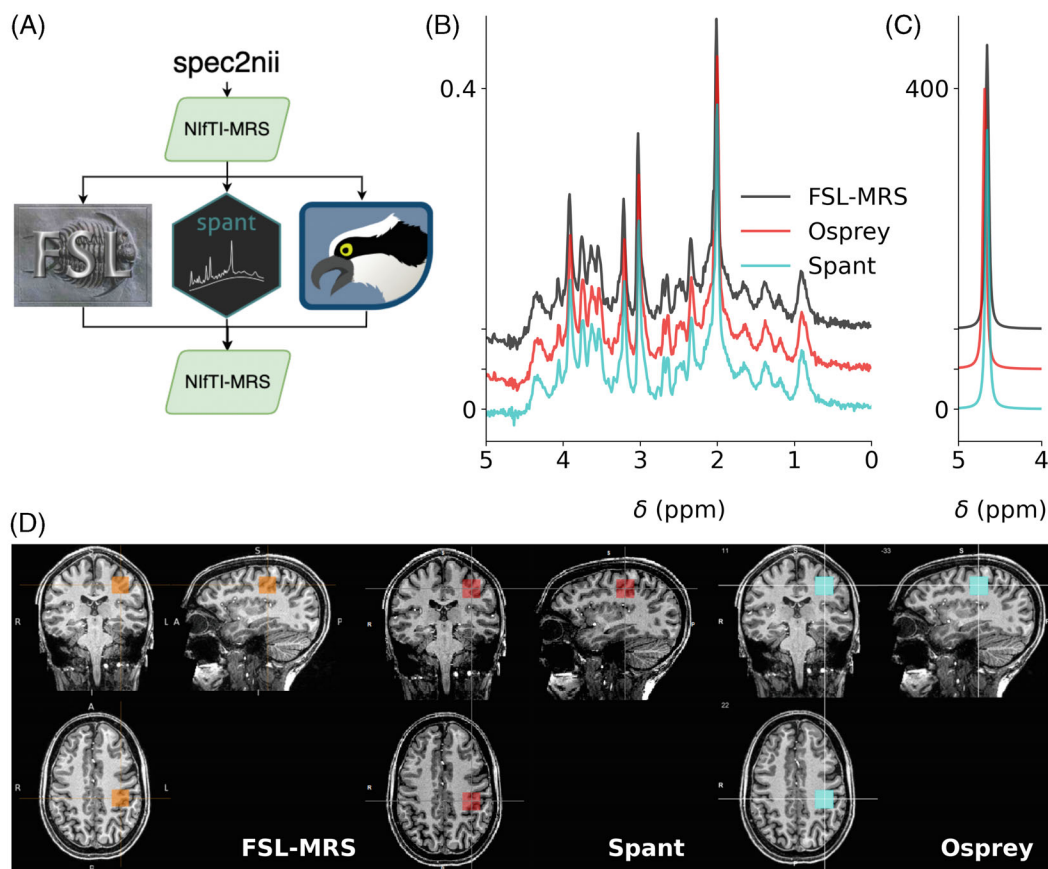
This work enables NifTI-MRS formatted data to be displayed alongside arbitrary MRI and other modality imaging data formatted as NifTI. It also enables the visualization of multidimensional NifTI-MRS data. Figure 5 shows two examples of NifTI-MRS data displayed in FSleyes using the plugin.

## 4 | DISCUSSION

In this work, we have proposed NifTI-MRS as a new standard data storage format for MRS and MRSI. The standard was initially developed and agreed by a group of MRS physicists and researchers with experience in MRS analysis software use and development. Following a draft specification, feedback was sought from MRS software developers, experts in the field, and the wider technical MRS community. The standard and associated files are maintained online at Ref.19.

NifTI-MRS can provide a simplified analysis pathway and a path for interoperability of analysis programs, and greatly simplifies display and co-interpretation of MRS data alongside other modalities. In addition, the standard fulfills several other desirable objectives:

- A standard data storage format is a prerequisite for incorporating MRS data into standardized data structuring schemes, such as in the Brain Imaging Data Structure (BIDS).<sup>25</sup>



**FIGURE 4** NIfTI-MRS single voxel spectroscopy data processed in 3 spectroscopy toolboxes supporting input and output of NIfTI-MRS data. (A) Raw data converted to NIfTI-MRS format by *spec2nii* was loaded and preprocessed in each of the toolboxes before being written back out to NIfTI-MRS. Comparison of water-suppressed (B) and water-reference (C) data processed in each toolbox is simple using the pipeline in (A). The output NIfTI-MRS data is easy to present alongside structural data stored in NIfTI format using a variety of MRS analysis software (D)

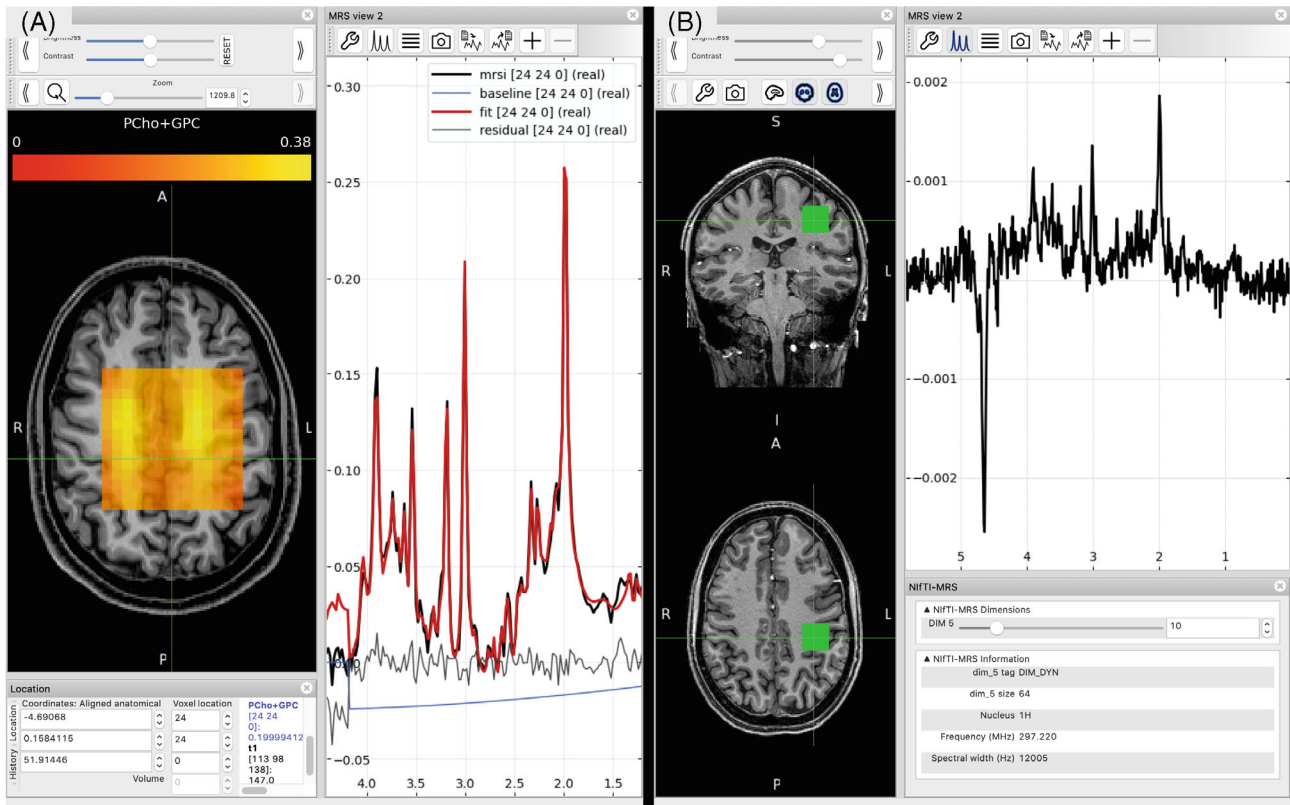
- Similarly, a standard format allows for extensive and easy data archiving and sharing in open science databases, such as OpenNEURO<sup>26,27</sup> or MRS-specific databases.<sup>28</sup> One study<sup>29</sup> has already utilized NIfTI-MRS for the purpose of releasing study data on XNAT central (<https://central.xnat.org/data/projects/PN21>). And another has recently submitted data formatted as NIfTI-MRS to the National Institute of Mental Health Data Archive (National Institutes of Health, Rockville, MD).<sup>30</sup>
- Standard header definitions and processing provenance make it easier for users to comply with minimum reporting standards consensus statements,<sup>31</sup> and to reproduce and recreate data analysis workflows simply by extracting the provenance information from a NIfTI-MRS file users received

The developers of NIfTI-MRS have sought to make the standard as accessible as possible. *Spec2nii* implements a mature open-source program suitable for both one-off and batch conversion tasks. NIfTI I/O libraries are available in all common programming languages, and minimal

examples of readers have been provided. Support for NIfTI-MRS is already available in five analysis packages (with further packages implementing the standard currently). This represents a critical mass of actively developed processing tools.

To date, *spec2nii* handles 14 different formats; however, complete coverage of sequences across such a diverse range of formats cannot be guaranteed. Currently, data must be in a spatially reconstructed format before conversion. Development of *spec2nii* has relied on a community-led model, with users contributing examples, test data, and code contributions to the developers.

In this work, FSLeyes has been extended with a plugin to create an NIfTI-MRS compatible data viewer. The authors consider this just an example of the existing tools present in the MR imaging community that can be easily leveraged for MRS and MRSI by using NIfTI-MRS. NIfTI-MRS mitigates the difficulty of aligning and registering MRSI data with MRI (and other modality) data, easing the simultaneous use of imaging and spectroscopic data to further methodological and analysis techniques in both fields.



**FIGURE 5** (A) MRSI data displayed in FSleyes using the NIFTI-MRS plugin. In addition to the NIFTI-MRS data, results from spectral fitting (also stored in NIFTI format) are displayed overlaid on the spectrum, to the left a metabolite map of total choline is displayed overlaid on a  $T_1$ w structural image. The spectrum display automatically displays voxel-wise spectrum and fits as the cursor is moved over the orthographic display. (B) Partially preprocessed single voxel data displayed in FSleyes alongside corresponding structural data. Beneath the main spectral view panel, an additional panel “NIFTI-MRS” (bottom-right) displays a summary of the MRS specific header information contained in the file, and a slider UI element to allow the user to view each spectrum stored in the higher (5th–7th) dimensions. Here the tenth (of 64) temporal averages stored in dimension 5 is displayed

The NIFTI-MRS standard is versioned, enabling backward-compatible future extension of the standard. The authors propose that the International Society for Magnetic Resonance in Medicine MRS Study Group Committee for Code and Data Sharing ([www.mrshub.org/](http://www.mrshub.org/)), established in 2020 as a permanent standing committee with rotating members, helps maintain oversight of the standard and any future development alongside the original authors.

Establishing a standardized data format for MRS is a requirement for MRS to be used for large scale cross-center, or population, studies. The NIFTI format provided this for neuroimaging studies such as the Human Connectome Project.<sup>32</sup> Furthermore, the uptake of the NIFTI-MRS standard shows the possibility of establishing an ecosystem of open-source analysis toolboxes similar to that which has benefitted functional neuroimaging (e.g., AFNI, FSL, SPM).<sup>33–35</sup> Nonetheless, this work does not tackle other key aspects such as communication of MRS fitting results.

## 5 | CONCLUSION

NIFTI-MRS has been designed to both promote the use of MRS in biomedical research and ease the technical development of MRS analysis in the wider field of biomedical imaging. The standard and associated tools are developed in an open-research context. Further use and development of NIFTI-MRS and associated tools are dependent on the expertise and contributions of the community, including MR hardware vendors. To date, the community has enthusiastically done so, enabling the rapid development of the standard and its inclusion in multiple software tools.

## ACKNOWLEDGMENT

The authors thank all contributors of test data and source code to the *spec2nii* program— especially Jack Miller and Tomas Psorn —and Paul McCarthy, the FSleyes developer, for his extensive advice in creating the NIFTI-MRS visualization plugin.




## FUNDING INFORMATION


Support provided by funding from the Wellcome Trust and the Royal Society, grant 102584/Z/13/Z (William T. Clarke.); the National Institutes of Health (NIH), grants K99/R00 AG062230, S10 OD021648, P41 EB031771, P41 EB015909, R01 EB016089, R01 EB023963, and R01 EB028259 (Georg Oeltzschner); and funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 813120 (Amir M. Shamaei.). The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust, grant 203139/Z/16/Z.

## DATA AVAILABILITY STATEMENT


All source code and data examples can be found in the related GitHub repository, [https://github.com/wtclarke/mrs\\_nifti\\_standard](https://github.com/wtclarke/mrs_nifti_standard), release version 0.6.0 at time of publication.

## ORCID

William T. Clarke  <https://orcid.org/0000-0001-7159-7025>

Uzay E. Emir  <https://orcid.org/0000-0001-5376-0431>

Mark Mikkelsen  <https://orcid.org/0000-0002-0349-3782>

Georg Oeltzschner  <https://orcid.org/0000-0003-3083-9811>

Amirmohammad Shamaei  <https://orcid.org/0000-0001-8342-3284>

Brian J. Soher  <https://orcid.org/0000-0003-2750-7277>

Martin Wilson  <https://orcid.org/0000-0002-2089-3956>

## TWITTER

William T. Clarke  @wt\_clarke

## REFERENCES

- Bednařík P, Tkáč I, Giove F, et al. Neurochemical and BOLD responses during neuronal activation measured in the human visual cortex at 7 Tesla. *J Cereb Blood Flow Metab.* 2015;35:601-610.
- Palombo M, Shemesh N, Ronen I, Valette J. Insights into brain microstructure from in vivo DW-MRS. *Neuroimage.* 2018;182:97-116.
- Spencer RG, Balschi JA, Leigh JS, Ingwall JS. ATP synthesis and degradation rates in the perfused rat heart. 31P-nuclear magnetic resonance double saturation transfer measurements. *Biophys J.* 1988;54:921-929.
- Near J, Harris AD, Juchem C, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR Biomed.* 2021;34:e4257.
- C.8.14 MR Spectroscopy Modules. Available at: [http://dicom.nema.org/medical/dicom/2017c/output/chtml/part03/sect\\_C.8.14.html](http://dicom.nema.org/medical/dicom/2017c/output/chtml/part03/sect_C.8.14.html). Accessed December 16, 2020.
- Cox RW, Ashburner J, Breman H, et al. A (sort of) new image data format standard: NiFTI-1. In Proceedings of the 10th Annual Meeting of the Organization for Human Brain Mapping (OHBM), Budapest, Hungary, 2004.
- McCarthy P. FSLEyes. 2020. Available at: <https://10.5281/zenodo.3937147>.
- Clarke WT, Mikkelsen M, Oeltzschner G, et al. NiFTI-MRS Format Specification. 2021. Available at: <https://10.5281/zenodo.5084789>.
- Pezoa F, Reutter JL, Suarez F, Ugarte M, Vrgoč D. Foundations of JSON schema. *Proceedings of the 25th International Conference on World Wide Web.* International World Wide Web Conferences Steering Committee; 2016:263-273.
- Introducing JSON: ECMA-404 The JSON Data Interchange Standard. Available at: <https://www.json.org/json-en.html>. Accessed December 16, 2020.
- Cox RW. Official definition of the nifti1 header. Available at: <https://nifti.nimh.nih.gov/pub/dist/src/niftilib/nifti1.h>. Accessed July 9, 2021.
- Reynolds R. Official definition of the nifti2 header. Available at: <https://nifti.nimh.nih.gov/pub/dist/doc/nifti2.h>. Accessed July 9, 2021.
- Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J. A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage.* 1995;2:89-101.
- Levitt MH. The signs of frequencies and phases in NMR. *J Magn Reson.* 1997;126:164-182.
- Clarke WT. NiFTI-MRS Example Data; 2021. Available at: <http://10.5281/zenodo.5085449>.
- Mescher M, Merkle H, Kirsch J, Garwood M, Gruetter R. Simultaneous in vivo spectral editing and water suppression. *NMR Biomed.* 1998;11:266-272.
- Clarke WT, Stagg CJ, Jbabdi S. FSL-MRS: an end-to-end spectroscopy analysis package. *Magn Reson Med.* 2021;85:2950-2964.
- Inati SJ, Naegele JD, Zwart NR, et al. ISMRM raw data format: a proposed standard for MRI raw datasets. *Magn Reson Med.* 2017;77:411-421.
- Clarke W. [http://github.com/wtclarke/mrs\\_nifti\\_standard](http://github.com/wtclarke/mrs_nifti_standard). 2020. (Accessed June 21, 2022)
- Oeltzschner G, Zöllner HJ, Hui SCN, et al. Osprey: open-source processing, reconstruction & estimation of magnetic resonance spectroscopy data. *J Neurosci Methods.* 2020;343:108827.
- Wilson M. Adaptive baseline fitting for MR spectroscopy analysis. *Magn Reson Med.* 2021;85:13-29.
- Wilson M. Spant: an R package for magnetic resonance spectroscopy analysis. *Joss.* 2021;6:3646.
- Soher B, Semanchuk P, Todd D, Steinberg J, Young K. VeSPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis. In Proceedings of the 19th Annual Meeting of ISMRM, Montréal, Québec, Canada, 2011. p. 1410.
- Simpson R, Devenyi GA, Jezzard P, Hennessy TJ, Near J. Advanced processing and simulation of MRS data using the FID appliance (FID-A)—an open source, MATLAB-based toolkit. *Magn Reson Med.* 2017;77:23-33.
- Gorgolewski KJ, Auer T, Calhoun VD, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data.* 2016;3:160044.
- Poldrack RA, Gorgolewski KJ. OpenfMRI: open sharing of task fMRI data. *Neuroimage.* 2017;144:259-261.

27. Markiewicz CJ, Gorgolewski KJ, Feingold F, et al. OpenNeuro: an open resource for sharing of neuroimaging data. *eLife* 2021;10:e71774.
28. Jiang SH, Coello E, Louis MS, Breedlove KM, Lin AP. MRSDB: a scalable multisite data library for clinical and machine learning applications of magnetic resonance spectroscopy. In Proceedings of the 29th Annual Meeting of ISMRM, Vancouver, British Columbia, Canada, 2021. p. 3764.
29. Zacharopoulos G, Sella F, Kadosh RC. The impact of a lack of mathematical education on brain development and future attainment. *Proc Natl Acad Sci U S A*. 2021;118:e2013155118.
30. Ashizawa T, Öz G, Paulson HL. Spinocerebellar ataxias: prospects and challenges for therapy development. *Nat Rev Neurol*. 2018;14:590-605.
31. Lin A, Andronesi O, Bogner W, et al. Minimum reporting standards for in vivo magnetic resonance spectroscopy (MRSin-MRS): Experts' consensus recommendations. *NMR Biomed*. 2021;34:e4484.
32. Glasser MF, Smith SM, Marcus DS, et al. The Human Connectome Project's neuroimaging approach. *Nat Neurosci*. 2016;19:1175-1187.
33. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996;29:162-173.
34. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *Neuroimage*. 2012;62:782-790.
35. Friston KJ, ed. *Statistical Parametric Mapping: the Analysis of Functional Brain Images*. 1st ed. Elsevier/Academic Press; 2007.

**How to cite this article:** Clarke WT, Bell TK, Emir UE, et al. NifTI-MRS: A standard data format for magnetic resonance spectroscopy. *Magn Reson Med*. 2022;88:2358-2370. doi: 10.1002/mrm.29418



---

# AMIR SHAMAEI

[223363@vut.cz](mailto:223363@vut.cz), [amirshamaei@isibrno.cz](mailto:amirshamaei@isibrno.cz),

Institute of Scientific Instruments of the Czech Academy of Sciences  
Room 401, +420 541 514 394  
aKralovopolska 147, 612 64 Brno, Czech  
Republic

[LinkedIn](#) | [GitHub](#) | [Twitter](#) | [Researchgate](#)

## CURRENT POSITION

Marie Curie Fellow at the INSPiRE-MED consortium\* Brno, Czech Republic  
Institute of Scientific Instruments of the Czech Academy of Sciences,  
Magnetic Resonance Group 2019-Present  
\* received funding from the European Union's Horizon 2020 research  
and innovation program under the Marie Skłodowska-Curie grant  
agreement No 813120

## EDUCATION

Brno University of Technology, Department of Biomedical Engineering Brno, Czech Republic  
Ph.D. | Biomedical Engineering  
Dissertation: Development of new artificial intelligence functionalities  
for magnetic resonance spectroscopic signals processing 2019-Present  
Supervisor: Dr. Jana Starčuková  
Co-Supervisors: Dr. Zenon Starčuk, Dr. Radovan Jiřík

Amirkabir University of Technology (Tehran Polytechnic) Tehran, Iran  
M.Sc. | Biomedical Engineering  
Dissertation: Design and fabrication of a Lab-on-chip device for  
assessment of cancer cells viability during electrical stimulation 2015-2018  
Supervisor: Dr. Saviz  
Co-Supervisors: Dr. Solouk, Dr. Abdullahad

Zanjan University Zanjan, Iran  
B.Sc. | Electrical Engineering  
Dissertation: Wireless System Network Simulator Software 2010-2015  
Supervisor: Dr. Babazadeh

## RESEARCH INTERESTS

Deep learning, Magnetic Resonance in Medicine, Self-Supervised Machine Learning

## HONORS & AWARDS

Received a full scholarship from the Marie Skłodowska-Curie Actions (a set of major research fellowships created by the European Union/European Commission to support research in the European Research Area)

Ranked 1<sup>st</sup> in the student competition conference at the Brno University of Technology

Awarded Thermo Fisher Scientific Corporation's 2020 prize

---

Ranked 272<sup>nd</sup> out of nearly 30,000 participants in the Iranian university entrance exam for master's degree

Received a full scholarship from Amirkabir University of Technology

Received a full scholarship from Zanjan University

## CONFERENCE PRESENTATIONS

**Shamaei, A.M.**, Starčuková J., Radim Kořínek, and Starcuk Jr. Z., Magnetic Resonance Spectroscopic Imaging Data Denoising by Manifold Learning: An Unsupervised Deep Learning Approach, Poster presentation delivered in person at *ISMRMB 2022*, May, 2022.

**Shamaei, A.M.**, Starčuková J. and Starcuk Jr. Z., Frequency and Phase Shift Correction of MR Spectra Using Deep Learning in Time Domain, Poster presentation delivered virtually at *ESMRMB 2021*, October, 2021.

Clarke, W., Bell, T., Emir, U., Mikkelsen, M., Oeltzschner, G., Rowland, B., **Shamaei, A.M.**, Soher, B, Tapper, S., and Wilson, M, NIfTI MRS: A standard format for spectroscopic data, Poster presentation delivered in person at *ISMRMB 2021*, May, 2021.

**Shamaei, A.M.**, Starčuková J. and Starcuk Jr. Z., A Wavelet Scattering Convolution Network for Magnetic Resonance Spectroscopy Signal Quantitation, Oral Presentation delivered virtually at *the 14th International Joint Conference on Biomedical Engineering Systems and Technologies*, Vienna, February, 2021, 10.5220/0010318502680275.

**Shamaei, A.M.**, Jiřík R. and Starčuková J., Deep Learning For Magnetic Resonance Spectroscopy Quantification: A Time-Frequency Analysis Approach, Presented in *The 26th Conference Student EEICT, Brno University of Technology*, March, 2020, WOS:000598376500032.

**Shamaei, A.M.** and Saviz, M. Voltage Transfer Functions for In-Vitro Cell Stimulation: A Computational Study, Poster presentation delivered at *IEEE Conference on Biomedical Engineering, Amirkabir University of Technology*, October, 2017, 10.1109/ICBME.2017.8430245.

## PEER-REVIEWED PUBLICATION

**Shamaei, A**, Saviz, M., Solouk, A. et al. An In Vitro Electric Field Exposure Device with Real-Time Cell Impedance Sensing. *Iran J Sci Technol Trans Sci* 44, 575–585 (2020). [doi.org/10.1007/s40995-020-00861-z](https://doi.org/10.1007/s40995-020-00861-z)

**Shamaei, A**, Starcukova, J, Pavlova, I, Starcuk, Z. Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals. *Magn Reson Med.* 2023; 89: 1221– 1236. [doi:10.1002/mrm.29498](https://doi.org/10.1002/mrm.29498)

Clarke, W, Mikkelsen, M, Oeltzschner, G, Bell T.K., **Shamaei, A**, Soher, B.J., Emir, U, Wilson, W. A standard data format for magnetic resonance spectroscopy. *Magnetic Resonance in Medicine* 2022; 1- 13. [doi:10.1002/mrm.29418](https://doi.org/10.1002/mrm.29418),

Rizzo, R, Dziadosz, M, Kyathanahally, SP, **Shamaei, A**, Kreis, R. Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias. *Magn Reson Med.* 2022; 1- 21. [doi:10.1002/mrm.29561](https://doi.org/10.1002/mrm.29561)

**Shamaei, A**, Starcukova, J, Starcuk, Z. Physics-informed Deep Learning Approach to Quantification of Human Brain Metabolites from Magnetic Resonance Spectroscopy Data.

---

Computers in Biology and Medicine. 2023; 158: 106837. doi: 10.1016/j.combiomed.2023.106837.

## CERTIFICATIONS

qMRI workshop, Max Planck Institute for Human Cognitive and Brain Sciences, 2023

Cambridge Ellis Machine Learning Summer School, University of Cambridge, 2022

INSPIRATION Workshop (Business case game for developing medical devices), Icometrix, Leuven, Belgium, 2022

Simultaneous PET-MR workshop, The University of Manchester, 2022

Workshop on Open-Access and Ethics, The University of Manchester, 2022

Soft skills workshop on communication, EPFL, 2021

Deep Layers workshop, Czech Academy of Science, 2021

Machine learning: from basics to big data and deep learning in MRI/MRS, KU Leuven, 2021

jMRUI Software Package workshop, Czech Academy of Science, 2020

Introduction to clinical MRS, MRS quantitation methods, University of Bern, 2020

Biomedical Signal Processing and Programming (SignalPlant) Workshop, Czech Academy of Science, 2019

NeuroImaging - Mapping the function and structure of the brain, CEITEC Masaryk University, 2019

Multiphysics Analysis of bio-model Using COMSOL workshop, 2017

Introduction to computing in Python Course, 2014

Matlab Programming Course, 2010

## OPEN SOURCE PROJECTS

[Model-Constrained Deep Learning Approach to the Quantification of Magnetic Resonance Spectroscopy Data](#)

[Model-Informed unsupervised Deep Learning Approach to Frequency and Phase Correction of MRS Signals](#)

[BrukerEyes: a cross-platform open-source package for Bruker Datasets management](#)

[A Cross-Platform Open-Source tool for Interrelated MRS\(NifTi\) Datasets visualization](#)

[Frequency and Phase Correction of MRS signals Using Cross-Correlation](#)

[A JAVA package for converting Bruker datasets, including MRS and MRSI, to Nifti format](#)

[HLSVDPro JAVA implementation](#)

[A JAVA package providing Reading/Writing interface for NifTi MRS format](#)

[A Java library for reading and writing NifTI \(1&2\) image volumes](#)

---

## SKILLS

**Programming language:** Java, Python, C

**Software:** MatLab, jMRUI, Slicer, LC MODEL, FiTAID, Adobe Illustrator, Microsoft Office, Altium Designer, CST, LATEX

**Language:** Persian (native), Azeri (native), English (Professional working proficiency), Czech (Elementary proficiency), and Arabic (Elementary proficiency)

## FEATURED GRADUATE COURSES

<b>Image Processing, 16/20</b> AmirKabir University of Technology	Fall, 2015
<b>Bioelectromagnetism 17.5/20</b> AmirKabir University of Technology	Spring, 2016
<b>Advanced Digital Signal Processing, 17.25/20</b> AmirKabir University of Technology	Fall, 2016
<b>Cognition &amp; Brain Physiology, 18.9/20</b> AmirKabir University of Technology	Spring, 2017
<b>New trends in the analysis and classification of biomedical data, 95/100</b> Brno University of Technology	Spring, 2020

## INTERESTS

Artificial Intelligence, History of Humankind, Arduino & Raspberry Pi, GYM, Traveling, Bayesian statistics, Books/Non-fiction

## TEACHING & WORK EXPERIENCE

<b>TEACHING ASSISTANT</b> Amirkabir University of Technology Principles of Bio-Electronics	2016-2017
<b>INTERNSHIP</b> Swiss Institute for Translational and Entrepreneurial Medicine	Winter 2020
<b>INTERNSHIP</b> The University of Bern – Departments of Radiology, Neuroradiology and Nuclear Medicine, INSEL, Switzerland	Winter 2020
<b>INTERNSHIP</b> Tesla Dynamic Coils, Netherlands	Summer 2022

