의학박사 학위논문

Deep Learning-based Metabolite Quantification in Proton Magnetic Resonance Spectroscopy of the Brain

양성자 자기공명분광법을 사용한
딥러닝 기반
두뇌 대사체 정량화 기법

2021 년 10 월

서울대학교 대학원
의과학과

이 형 훈

Thesis for the Degree of
Doctor of Philosophy in Medical Science

Deep Learning-based Metabolite Quantification in Proton Magnetic Resonance Spectroscopy of the Brain

양성자 자기공명분광법을 사용한
딥러닝 기반
두뇌 대사체 정량화 기법

February 2022

Seoul National University Graduate School
Major in Biomedical Sciences

Hyeong Hun Lee

양성자 자기공명분광법을 사용한 딥러닝 기반 두뇌 대사체 정량화 기법

지도 교수 김 현 진

이 논문을 의학박사 학위논문으로 제출함

2021 년 10 월

서울대학교 대학원

의과학과

이 형 훈

이형훈의 의학박사 학위논문을 인준함

2022 년 1 월

위 원 장 _____ **(인)**

부위원장 _____ **(인)**

위 원 _____ **(인)**

위 원 _____ **(인)**

위 원 _____ **(인)**

Deep Learning-based Metabolite Quantification in Proton Magnetic Resonance Spectroscopy of the Brain

Examiner Hyeonjin Kim

Submitting a Ph.D. Dissertation of Public Administration
October 2021

Seoul National University Graduate School
Major in Biomedical Sciences
Hyeong Hun Lee

Confirming the Ph.D. Dissertation written by Hyeong Hun Lee
January 2022

Chair _____(Seal)

Vice Chair _____(Seal)

Examiner _____(Seal)

Examiner _____(Seal)

Examiner _____(Seal)

Abstract

Deep Learning-based Metabolite Quantification in Proton Magnetic Resonance Spectroscopy of the Brain

Hyeong Hun Lee
Major in Biomedical Sciences
Department of Biomedical Sciences
Seoul National University Graduate School

Nonlinear-least-squares-fitting (NLSF) is widely used in proton magnetic resonance spectroscopy (MRS) for quantification of brain metabolites. However, it is known to subject to variability in the quantitative results depending on the prior knowledge. NLSF-based metabolite quantification is also sensitive to the quality of spectra. In combination with NLSF, Cramer-Rao lower Bounds (CRLB) are used as representing lower bounds of fit errors rather than actual errors. Consequently, a careful interpretation is required to avoid potential statistical bias.

The purpose of this study was to develop more robust methods for metabolite quantification and uncertainty estimation in MRS by employing deep learning that has demonstrated its potential in a variety of different tasks including medical imaging.

To achieve this goal, first, a convolutional neural network (CNN) was developed. It maps typical brain spectra that are degraded with noise, line-

broadening and unknown baseline into noise-free, line-narrowed, baseline-removed spectra. Then, metabolites are quantified from the CNN-predicted spectra by a simple linear regression with more robustness against spectral degradation. Second, a CNN was developed that can isolate each individual metabolite signals from a typical brain spectrum. The CNN output is used not only for quantification but also for calculating signal-to-background-ratio (SBR) for each metabolite. Then, the SBR in combination with big training data are used for estimating measurement uncertainty heuristically. Finally, a Bayesian deep learning approach was employed for theory-oriented uncertainty estimation. In this approach, Monte Carlo dropout is performed for simultaneous estimation of metabolite content and associated uncertainty. These proposed methods were all tested on in vivo data and compared with the conventional approach based on NLSF and CRLB.

The methods developed in this study should be tested more thoroughly on a larger amount of in vivo data. Nonetheless, the current results suggest that they may facilitate the applicability of MRS.

Keywords: Magnetic Resonance Spectroscopy, Deep learning, Metabolites

Student Number: 2014-22018

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Magnetic Resonance Spectroscopy

Spectroscopy is the study of the interaction between particles and electromagnetic radiation or the response of particles to radiation as a function of the frequency of radiation. The types of spectroscopy are divided according to the wavelength of the electromagnetic wave and the properties of materials corresponding thereto. Representative examples include infrared (IR) spectroscopy, ultraviolet (UV) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy. Among them, NMR spectroscopy uses radio frequencies in the range of 10 MHz – 1 GHz for the observation of magnetic properties and energy states of nuclei. [1, 2]

In vivo NMR is one of the major applications of modern NMR [3-5], and its representative examples include magnetic resonance imaging (MRI) and magnetic resonance spectroscopy (MRS). These methods obtain spatial encoded or voxel-localized signal from a sample placed in a static external magnetic field using RF pulses and magnetic field gradient pulses. MRS is a technique that observes metabolic activities such as oxidative phosphorylation and glycolysis in specific tissues in vivo by non-invasively measuring the resonance signals of metabolites in anatomically localized voxels.

This Section introduces the basic principles of MRS, those metabolites detectable in vivo, and measurement methods.

1.1.1. Nuclear Spin

Spin is the intrinsic angular momentum of subatomic particles (i.e., protons, neutrons, and electrons), exhibiting the property of being quantized as discrete values in a magnetic field, [6-10] and can be classified into nuclear and electron spins. Since MRS mainly uses nuclear spin, which is the intrinsic angular momentum of the nucleus, the following descriptions are based on nuclear spin.

The magnitude of the nuclear spin angular momentum **L** can be written as

$$L = \hbar\sqrt{I(I+1)}. \tag{1.1}$$

where I is the nuclear spin quantum number [8, 9] and is characterized by the following properties of subatomic particles. [11]

- The number of neutrons and the protons are both even, then the I = 0.
- The number of neutrons plus the number of protons is odd, then the I = half-integer. (i.e., 1/2, 3/2, 5/2)
- The number of neutrons and the number of protons is both odd, then the I = integer. (i.e., 1, 2, 3)

Since this study deals only with hydrogen (proton), the following contents will be explained on the basis of I = 1/2.

In the presence of an external magnetic field $B_0 = B_z$, a spin angular momentum component along z can only have a certain value of L_z, which can be written as

$$L_Z = \hbar m \tag{1.2}$$

$$m = -I, -I+1, \dots, 0, , \dots I-1, I \tag{1.3}$$

where m is the nuclear spin projection quantum number, and the possible values for proton (I=1/2) in a magnetic field are +1/2 or –1/2 (called spin-up state (α, ground energy state) and spin-down state (β, excited energy state), respectively). Meanwhile, the angular momentum **L** in classical mechanics is also expressed as

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \tag{1.4}$$

$$\mathbf{p} = m\mathbf{v}. \tag{1.5}$$

Next, the magnetic dipole moment $\boldsymbol{\mu}$ is a vector indicating the degree to which a particle receives a rotational force in response to a magnetic field, and is expressed as a product of a current *i* and the area of the current loop **A**. That is,

$$\boldsymbol{\mu} = i \times \mathbf{A} = \left(e\frac{\mathbf{v}}{2\pi\mathbf{r}}\right)\pi\mathbf{r}^2 \tag{1.6}$$

where *e* is the electric charge, **r** is the distance between the particle and the center of rotation, *i* is the value obtained by dividing the amount of charge of the electron by the period of the circular motion, and the period of the circular motion is expressed as $2\pi\mathbf{r}/\mathbf{v}$. [12] Substituting the $\mathbf{L}$ derived in Equation (1.4) into the $\boldsymbol{\mu}$, it is written as

$$\boldsymbol{\mu} = \left(\frac{e}{2m}\right)\mathbf{L} = \gamma\mathbf{L} \tag{1.7}$$

where γ is the gyromagnetic ratio, which represents the ratio of the magnetic moment to the angular momentum of the nuclei and represents a unique value for each nucleus as shown in the Table 1-1. This means that each nucleus can be

distinguished even in an environment with the same magnetic field strength.

Nucleus	Spin (I)	Gyromagnetic Ratio (10^7 rad $T^{-1}s^{-1}$)	Gyromagnetic Ratio (MHz T^{-1})	Natural Abundance (%)
^{1}H	1/2	26.752	42.576	99.985
^{2}H	1	4.107	6.536	0.015
^{13}C	1/2	6.728	10.705	1.108
^{17}O	5/2	-3.628	-5.772	0.0001
^{19}F	1/2	25.181	40.059	100
^{31}P	1/2	10.841	17.235	100
^{23}Na	3/2	7.080	11.262	100

Table 1-1. Gyromagnetic ratios, spin quantum number and natural abundance of selected nuclei.

As in Equation (1.7), a nucleus with a magnetic moment exhibits angular momentum, and if there is an external magnetic field ($\mathbf{B_0}$), a rotational force (torque; **T**) is applied to the magnetic moment and the spin precession occurs around the direction of the $\mathbf{B_0}$.

$$\mathbf{T} = \mathbf{r} \times \mathbf{F} = \left(\frac{d\mathbf{L}}{dt}\right) = \boldsymbol{\mu} \times \mathbf{B_0} \tag{1.8}$$

Rewriting Equation (1.8) for $\boldsymbol{\mu}$ as

$$\left(\frac{d\boldsymbol{\mu}}{dt}\right) = \gamma\boldsymbol{\mu} \times \mathbf{B_0} = \boldsymbol{\mu} \times \boldsymbol{\omega_0}. \tag{1.9}$$

$$\boldsymbol{\omega_0} = \gamma \mathbf{B_0}. \tag{1.10}$$

Equation (1.10) is called the Larmor Equation and $\boldsymbol{\omega_0}$ represents the precession frequency of the proton (or other nuclei) in the external magnetic field, and is directly proportional to the external magnetic field ($\mathbf{B_0}$).

Meanwhile, since the magnetic potential energy is also related to magnetic moment $\boldsymbol{\mu}$ and external magnetic field $\mathbf{B_0}$, it can be written as

$$E = -\boldsymbol{\mu} \cdot \mathbf{B_0}. \tag{1.11}$$

Assuming that the direction of $\mathbf{B_0}$ is along the z-axis, Equation (1.11) can be rewritten as

$$E = -\mu_z B_0 = -(\gamma L_Z) B_0 = -\gamma \hbar m B_0. \tag{1.12}$$

Therefore, the energy difference ΔE between the spin-up and spin-down states can be written as follows.

$$\Delta E = \gamma \hbar B_0 = \hbar \omega_0 \tag{1.13}$$

$$\omega_0 = \gamma B_0 \tag{1.14}$$

where ω_0 is also called the Larmor frequency and has the identical concept as $\boldsymbol{\omega_0}$ in Equation (1.10). This is the most fundamental concept in MRS that is utilized in the generation of the resonance signal through the RF excitation as described in Section 1.1.3.

In conjunction with Equation (1.3) and (1.11), if an external magnetic field is applied along the z-axis, the μ_z of the proton is aligned in a direction parallel (spin-up state) to or antiparallel (spin-down state) to the magnetic field direction according to the spin state. (Figure 1-1) There exists a population difference between the two states, which is descripted in detail in the next Section.

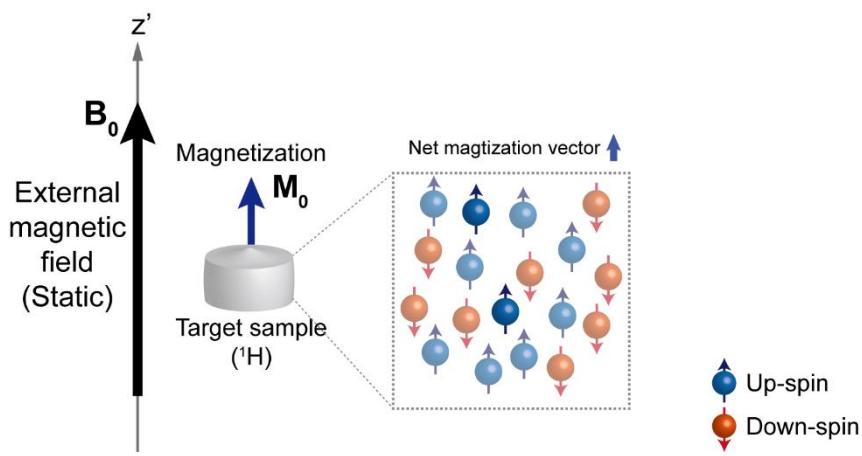


Figure 1-1. Graphical representation of magnetization.

1.1.2. Magnetization

In a sample at thermal equilibrium in the presence of an external magnetic field, the spins in spin-up state (α) slightly outnumbers those in spin-down state (β), and consequently, a magnetization is generated (Figure 1-1). The magnetization **M** of an *n*-spin system is expressed as

$$\mathbf{M} = \sum_{i}^{n} \mu_i = \mathrm{n}_\alpha \mu_\alpha + \mathrm{n}_\beta \mu_\beta. \tag{1.15}$$

where μ_α and μ_β are the magnetic moments in the spin-up and spin-down states of the proton, respectively, and n_α and n_β are the populations for each spin state. Since μ_α and μ_β contain 1/2 and $-1/2$ of the value of m (Equation (1.2) and (1.7)), respectively, it can be rewritten as follows.

$$\mathbf{M} = \left(\frac{1}{2}\right)\frac{\gamma h}{2\pi}\mathrm{n}_\alpha + \left(-\frac{1}{2}\right)\frac{\gamma h}{2\pi}\mathrm{n}_\beta = \frac{\gamma h}{4\pi}\left(\mathrm{n}_\alpha - \mathrm{n}_\beta\right) \tag{1.16}$$

On the other hand, according to the Boltzmann statistics, the population ratio of n_α to n_β at thermal equilibrium can be written as

$$\frac{\mathrm{n}_\alpha}{\mathrm{n}_\beta} = \mathrm{e}^{\Delta \mathrm{E}/k\mathrm{T}} = \mathrm{e}^{h\upsilon_0/k\mathrm{T}}. \tag{1.17}$$

where *k* is the Boltzmann constant and the T is the absolute temperature [13]. Using a first order approximation of the Taylor series, Equation (1.17) becomes

$$\frac{\mathrm{n}_\alpha}{\mathrm{n}_\beta} \simeq 1 + \left(\frac{h\upsilon_0}{k\mathrm{T}}\right). \tag{1.18}$$

where $h\upsilon_0/k\mathrm{T} \ll 1$. Then, the difference of the population between the spin-up and spin-downstate can also be approximated as below for the *n*-spin system.

$$\mathrm{n}_\alpha - \mathrm{n}_\beta \simeq \frac{\mathrm{n}}{2}\left(\frac{h\upsilon_0}{k\mathrm{T}}\right) \tag{1.19}$$

Therefore, by plugging Equation (1.13) and (1.1) into Equation (1.15) and assuming the direction of $\mathbf{B_0}$ is along the z-axis, the magnitude of the macroscopic magnetization at the thermal equilibrium can be written as

$$\mathrm{M}_0 = \frac{\gamma h^2}{4\pi}\left(\frac{\mathrm{n}\upsilon_0}{2k\mathrm{T}}\right) = \gamma^2 \hbar^2 \left(\frac{\mathrm{n}\mathrm{B}_0}{4k\mathrm{T}}\right). \tag{1.19}$$

Equation (1.19) intuitively represents several factors that determine the magnitude of the magnetization that is associated with the sensitivity of the MRS signal. First, M_0 is directly proportional to the strength of B_0. Therefore, the signal sensitivity is improved in a higher magnetic field. Second, due to the fact that M_0 is proportional to the square of γ, ^{1}H has a much higher sensitivity than other nuclei. (e.g., according to Table 1-1, the γ of ^{1}H is ~2.47 times and ~3.98 times higher than those of ^{31}P and ^{13}C, respectively, and therefore, when all other variables in Equation (1.19) remain the same, M_0 of ^{1}H is about ~6.1 times and ~15.8 times larger than those of

^{31}P and ^{13}C, respectively.) Combined with its natural abundance in human body, ^{1}H is considered the nucleus of choice in MRS. Third, M_0 is directly proportional to n. Therefore, the larger the sample size (voxel size in MRS), the stronger the signal.

1.1.3. MRS Signal

At thermal equilibrium, **M** is a static vector along the z-direction. That is, defining the z-component of the magnetization as longitudinal magnetization (M_z), $M_z = M_0$, and as no transverse magnetization (M_{xy}) exists, there is no signal detectable.

$$\frac{dM_z(t)}{dt} = 0 \tag{1.20}$$

In order to generate an MRS signal, an additional, time-varying magnetic field ($\mathbf{B_1}$) must be applied to the sample in the direction perpendicular to $\mathbf{B_0}$ to flip M_z onto the transverse plane.

In MRS, RF pulses are utilized as the $\mathbf{B_1}$ field that is turned on and off with a limited duration (<100 ms on 3T human scanner). The magnetic field component of an RF pulse applied in the +x direction can be expressed as

$$\mathbf{B_1}(t) = 2B_1^e(t)\cos(\omega t + \varphi)\,\hat{\mathbf{x}}. \tag{1.21}$$

where B_1^e is the envelope of $\mathbf{B_1}(t)$, φ is the initial phase of $\mathbf{B_1}(t)$ (ignored from here on for simplicity), and ω is the carrier frequency of the RF pulse. Therefore, while an RF pulse is being applied, magnetization exhibits precession around both $\mathbf{B_0}$ and $\mathbf{B_1}$. Due to the application of an RF pulse, magnetization exhibits complex motion in three-dimensional space. More details about the motion of the magnetization are described in Section 1.1.7. On the other hand, since the RF pulse is linearly polarized according to Equation (1.21), it can be decomposed into two circularly polarized fields.

$$\begin{aligned}\mathbf{B_1}(t) = {} & B_1^e(t)[\cos(\omega t)\,\hat{\mathbf{x}} + \sin(\omega t)\,\hat{\mathbf{y}}] \\ & + B_1^e(t)[\cos(\omega t)\,\hat{\mathbf{x}} - \sin(\omega t)\,\hat{\mathbf{y}}]\end{aligned} \tag{1.22}$$

where $B_1^e(t)[\cos(\omega t)\,\hat{\mathbf{x}} + \sin(\omega t)\,\hat{\mathbf{y}}]$ in Equation (1.22) represents the circularly polarized field rotating counterclockwise, of which the influence is usually negligible. [14]

Given that a torque is exerted on the magnetic moment μ in the magnetic field **B**, and that the magnetization is the sum over all magnetic moments (Equation (1.14)), Equation (1.9) can be generalized to the macroscopic magnetization and is

expressed as [1]

$$\frac{d\mathbf{M}(t)}{dt} = \gamma \mathbf{M}(t) \times \mathbf{B}(t). \tag{1.23}$$

The total magnetic field **B**(t) considering both the approximated Equation for $\mathbf{B_1}$ and B_0 can be expressed as

$$\mathbf{B}(t) = B_1 \cos(\omega t)\hat{\mathbf{x}} - B_1 \sin(\omega t)\hat{\mathbf{y}} + B_0 \hat{\mathbf{z}}. \tag{1.24}$$

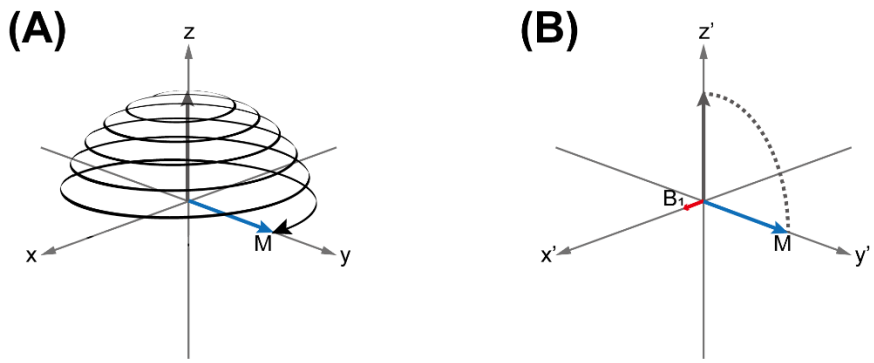


Figure 1-2. Change of magnetization by RF pulse. (A) Magnetization behavior in laboratory frame (B) Magnetization behavior in rotating frame.

Due to the RF pulse, M_z experiences a torque, which causes a rotation towards the transverse plane. (Figure 1-2(A)) Since the two magnetic fields act on the magnetization simultaneously, Equation (1.23) can be written for each component as

$$\frac{dM_x(t)}{dt} = \gamma(M_y B_0 + M_z B_1 \sin(\omega t)) \tag{1.25}$$

$$\frac{dM_y(t)}{dt} = \gamma(-M_x B_0 + M_z B_1 \cos(\omega t)) \tag{1.26}$$

$$\frac{dM_z(t)}{dt} = \gamma(-M_x B_1 \sin(\omega t) + M_y B_1 \cos(\omega t)). \tag{1.27}$$

At this point, it is more convenient to describe the behavior of the magnetization in a frame of reference that also rotates at ω around the z-axis. (Figure 1-2(B)) The magnetization in the rotating frame is defined as

$$M'_x(t) = M_x \cos(\omega t) - M_y \sin(\omega t) \tag{1.28}$$

$$M'_y(t) = M_x \sin(\omega t) + M_y \cos(\omega t) \tag{1.29}$$

$$M'_z(t) = M_z(t). \tag{1.30}$$

Using Equation (1.28) – (1.30), Equation (1.25) – (1.27) become as follows.

$$\frac{dM'_x(t)}{dt} = (\gamma B_0 - \omega)M'_y \tag{1.31}$$

$$\frac{dM'_y(t)}{dt} = -(\gamma B_0 - \omega)M'_x + \gamma B_1 M'_z \tag{1.32}$$

$$\frac{dM'_z(t)}{dt} = -\gamma B_1 M'_y \tag{1.33}$$

Equation (1.31) – (1.33) can be expressed in the form of Equation (1.23) as

$$\frac{d\mathbf{M}'(t)}{dt} = \gamma \mathbf{M}'(t) \times \mathbf{B_{eff}}(t). \tag{1.34}$$

where $\mathbf{B_{eff}}$ is defined as the effective magnetic field, which can be expressed as follows.

$$\mathbf{B_{eff}} = B_1\widehat{\mathbf{x}'} + \left(B_0 - \frac{\omega}{\gamma}\right)\widehat{\mathbf{z}'} \tag{1.35}$$

$$B_{eff} = |\mathbf{B_{eff}}| = \sqrt{B_1^2 + \left(B_0 - \frac{\omega}{\gamma}\right)^2} \tag{1.36}$$

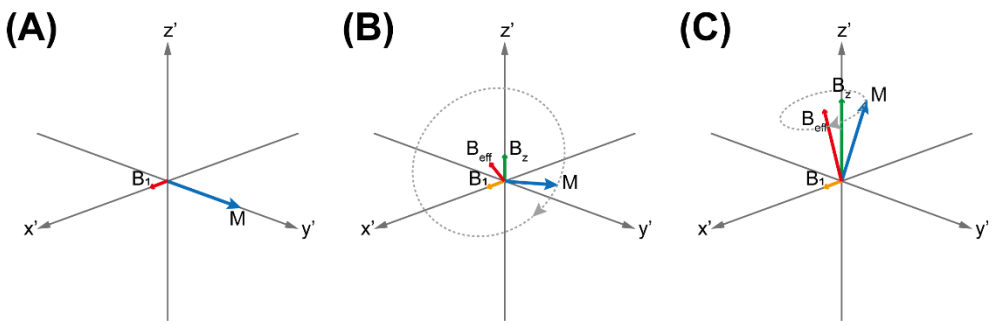


Figure 1-3. Vector representation of on/off resonance by RF pulse in rotating frame. (A) on resonance, (B) In the presence of a relatively weak or (C) strong effective magnetic field.

Figure 1-3(A) shows the on-resonance case. That is, when ω of the applied RF pulse is equal to Larmor frequency of the spin system (ω_0), Equation (1.35) becomes $\mathbf{B_{eff}} = B_1\widehat{\mathbf{x}'}$, and the magnetization is entirely placed on the transverse plane, as shown in Equations (1.31) – (1.32). In this process, the amount of rotation experienced by the net magnetization (**M**) due to an RF pulse is defined as the flip angle and can be expressed as

$$\alpha = \gamma \int_0^T B_1^e(t)dt. \tag{1.37}$$

If the shape of the RF pulse is a simple square, the flip angle can be simply expressed

as $\alpha = \gamma B_1 T$, and α=90° should be applied for maximum transverse component under the ideal on-resonance condition.

For an off-resonance case (i.e., $\omega \neq \omega_0$), the direction of $\mathbf{B_{eff}}$ lies between the transverse plane and the z-axis (Figure 1-3(B)-(C)). Therefore, the magnetization precesses around that direction of $\mathbf{B_{eff}}$ and induces a more complex rotation when compared to the on-resonance case.

The off-resonance effect in MRS can deteriorate spatial localization performance, and interrupt acquisition of the signal from the desired region. [15] The off-resonance effects can be reduced by optimizing RF pulses [16] and minimizing B_0 inhomogeneity through high-dimensional shimming. [17] For the rest of this chapter, off-resonance effects are ignored.

Immediately after an excitation RF pulse, transverse magnetization appears as an observable signal by inducing an electromotive force in the receiving coil through Faraday's law. According to Equations (1.23), (1.31) – (1.33), the resulting MRS signal will be detected forever. Therefore, Equation (1.23) requires an additional term that serves to cause the net transverse and longitudinal magnetization to return to their thermal equilibrium values over time. This is called relaxation. It is classified into transverse and longitudinal relaxation.

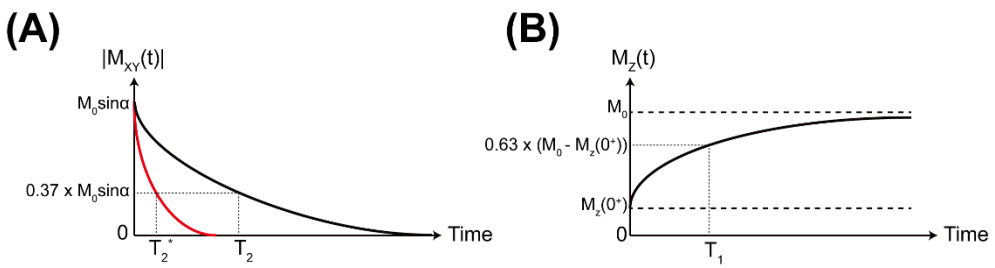


Figure 1-4. Transverse and longitudinal magnetization component with time by T_2 (black), T_2^* (red) relaxation (A) and T_1 relaxation (B).

First, longitudinal relaxation (or T_1 relaxation) is a process in which the M_z of the spin system excited by an RF pulse is restored to the thermal equilibrium value ($M_z = M_0$) through the interaction of the spin system with its environment. (i.e., the process by which the ratio of the spin populations between the spin-up and spin-down states recovers back to its original value at thermal equilibrium as in Equation (1.17)). The environment surrounding the spin system is collectively referred to as

lattice, and longitudinal relaxation is also termed as spin-lattice relaxation. The amount of change of M_z over time in longitudinal relaxation can be expressed as the difference between M_0 and M_z at a certain time (t) as

$$\frac{dM_z(t)}{dt} = \frac{1}{T_1}\left(M_0 - M_z(t)\right). \quad (1.38)$$

Solving Equation (1.38) for M_z,

$$M_Z(t) = M_0\left(1 - e^{-\frac{t}{T_1}}\right) + M_Z(0)e^{-\frac{t}{T_1}}. \quad (1.39)$$

In Equations (1.38) and (1.39), T_1 represents the time constant required for the M_z component to recover to $0.63M_0$, (Figure 1-4(A)). In MRS, different metabolites can have different T_1 depending on the characteristics of the lattice (e.g., viscosity and size of molecules). [18] In addition, T_1 tends to increase as B_0 increases in the range of clinical field strength (1.5 – 3.0T). [19, 20] In order for $M_z(t)$ to have recovered to M_0 at each excitation, the repetition time (TR) must be at least $5T_1$. [21], which is the main cause of the long scan time. [22]

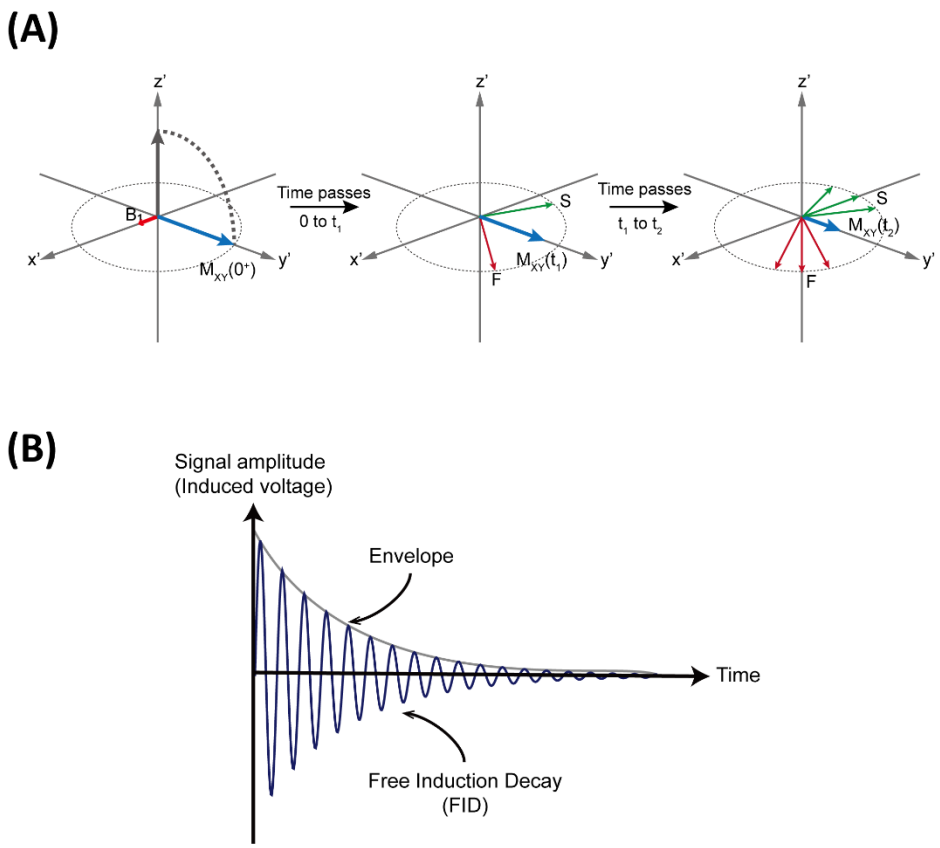


Figure 1-5. Graphical representation of spin dephasing (A) and consequent free induction decay (B).

The second relaxation term, transverse relaxation, also known as spin-spin relaxation, is caused by perturbation of local magnetic field at the site of neighboring spins. The effectiveness of this perturbation depends on the microscopic motion of the spins. As a result of the perturbation, the relative phase of the spins is gradually lost. (Figure 1-5(A)) Similar to longitudinal relaxation, the time-dependent change in the M_{xy} component due to transverse relaxation may be expressed as below.

$$\frac{dM_{xy}(t)}{dt} = \frac{1}{T_2}\left(0 - M_{xy}(t)\right) \tag{1.40}$$

Unlike M_z, the initial net magnetization of M_{xy} is 0 due to lack of phase coherence at thermal equilibrium. Solving Equation (1.40) for M_{xy},

$$M_{xy}(t) = M_{xy}(0)e^{-\frac{t}{T_2}}. \tag{1.41}$$

In Equation (1.40) and (1.41), T_2 represents the time constant required for the M_{xy} component to decrease to $0.37M_0$, (Figure 1-4(B)) and, like T_1, different metabolites can have different T_2. [18] In addition, as field strength (1.5 – 3.0T) increases, T_2 tends to be shortened. [19] The dephasing of the spins causing transverse relaxation is manifested as a decaying signal during data collection known as free induction decay (FID) (Figure 1-5(B)). The signal attenuation of FID is often expressed as an exponential function as in Equation (1.41). In practice, in addition to T_2 relaxation due to spin-spin interactions, the spin dephasing is much faster due to imperfect magnets and/or susceptibility effects. Therefore, the decay of FID is expressed in terms of T_2^*. Considering the decay constant, T_2', due to the local magnetic field inhomogeneity,

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}. \tag{1.42}$$

Although the initial signal decay after RF excitation obeys T_2^* decay, the effect of the imperfect magnet (if it is time-independent) can be compensated by refocusing of the spin phase in a spin echo experiment. [23] Spin echo is a signal generated by applying a 180° RF pulse after a 90° RF pulse, which refocuses the dephased spins. Therefore, the 180° RF pulse is often referred to as a refocusing RF pulse. The time delay between the first 90° pulse the spin echo is defined as echo time (TE). (Figure 1-6)

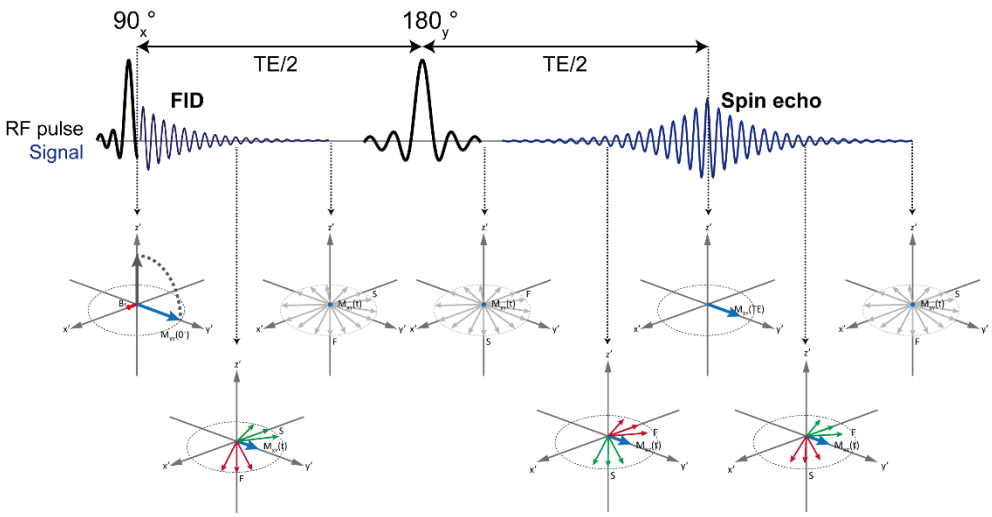


Figure 1-6. Spin echo process by refocusing RF pulse.

Finally, the Equation (1.23) can be rewritten by simultaneously considering RF field and relaxation as follows. [1]

$$\frac{\mathrm{d}\mathbf{M}(\mathrm{t})}{\mathrm{dt}} = \gamma\mathbf{M}(t) \times \mathbf{B}(\mathrm{t}) - \mathrm{R}\{\mathbf{M}(\mathrm{t}) - \mathbf{M}_0\} \tag{1.43}$$

$$\mathbf{B}(\mathrm{t}) = \mathbf{B}_0 + \mathbf{B}_1(\mathrm{t}) \tag{1.44}$$

where R is a relaxation matrix given as

$$\mathrm{R} = \begin{bmatrix} 1/T_2 & 0 & 0 \\ 0 & 1/T_2 & 0 \\ 0 & 0 & 1/T_1 \end{bmatrix}. \tag{1.45}$$

In MRS, the Equation (1.43) – (1.45) are used to analyze only the simple behavior of the magnetization vector under the application of RF pulses and longitudinal / transverse relaxation. For more general analysis of the behavior of the spin systems found in human brain, a quantum mechanical approach is required due to J-coupling (Section 1.1.5).

1.1.4. Chemical Shift

In B_0, the electron cloud around a nucleus is known to shield the magnetic field at the nucleus site [24, 25]. Therefore, the actual strength of the magnetic field experienced by the nucleus slightly differs from that of the static $\mathbf{B_0}$. Consequently, the resonance frequency of the nucleus would change accordingly. (Figure 1-7 (A) – (B)) As those protons detected in MRS reside in molecules with multiple nuclei, the

amount of change in resonance frequency of the protons would depend on their locations in a molecule and the structure of the molecule. Such a change in resonance frequency of a nucleus in a molecule due to its different chemical environment is called chemical shift. The shielded, effective magnetic field at a nucleus is expressed as

$$\mathbf{B} = \mathbf{B_0}(1 - \sigma). \tag{1.46}$$

where σ is a shielding constant (dimensionless) and depend on the chemical environment of the nucleus. Then, Equation (1.14) combined with Equation (1.46) yields

$$\upsilon = \frac{\gamma}{2\pi} B_0(1 - \sigma). \tag{1.47}$$

Chemical shift is expressed in part per million (ppm) as a relative frequency (δ) to a reference resonance frequency, such that it does not depend on B_0. (Figure 1-7 (C))

$$\delta = \frac{\upsilon - \upsilon_{\text{ref}}}{\upsilon_{\text{ref}}} \times 10^6. \tag{1.48}$$

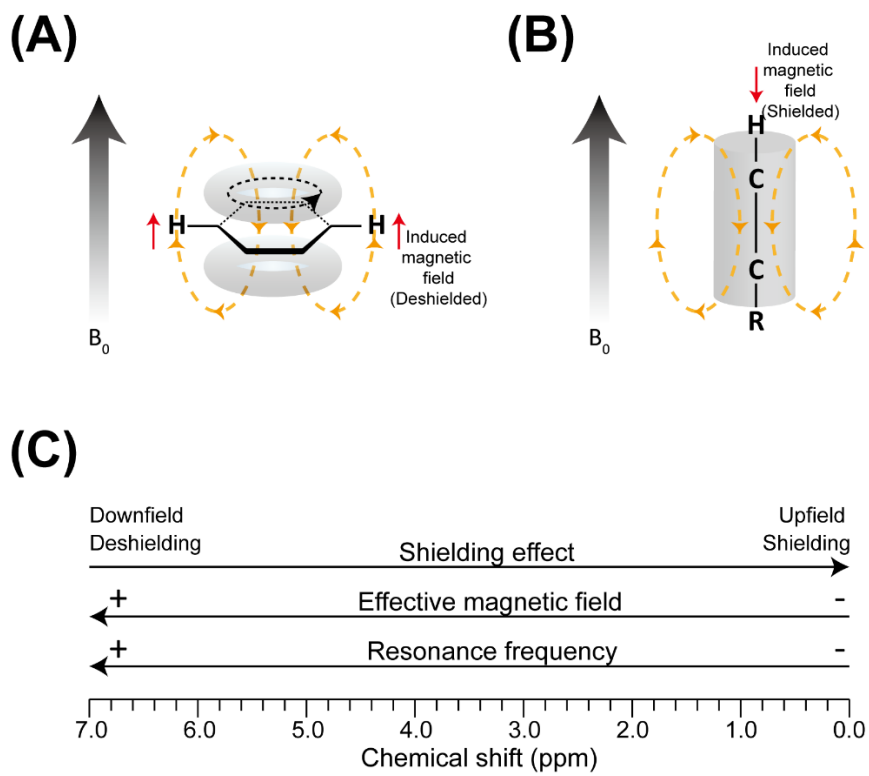


Figure 1-7. Graphical description of chemical shift. (A) deshielding (B) shielding effect (C) chemical shift change pattern according to the degree of shielding

In ^{1}H-MRS, the reference chemical shift ($\delta = 0$) is set to the resonance frequency (υ_{ref}) of the protons in the three methyl groups (in total, 9 protons per molecule) of Sodium trimethylsilyl propanesulfonate (DSS) or Trimethylsilylpropanoic acid (TSP).[26] However, since it does not exist in vivo (especially in the brain), the chemical shift of the protons in the methyl group of N-Acetyl-Aspartate (2.01 ppm) is typically used as an internal reference.

1.1.5. Indirect Spin-Spin Coupling

Indirect spin-spin coupling (also called J-coupling and scalar coupling) is a phenomenon that occurs as a result of an interaction between two nuclear spins in a molecule through a chemical bond (thus indirect) [27, 28].

More specifically, it is an interaction between two nuclear spins via covalent bonds, in which the electrons interact with the nuclear spins (hyperfine or Fermi contact interactions) [29]. Thus, the indirect spin-spin coupling contains information about the relative bonding distance and angle between nuclei, which could be extracted from MR spectra [30]. The strength of the indirect spin-spin coupling is expressed as a coupling constant J (Hz). Denoting $\Delta\upsilon$ as the chemical shift difference between two J-coupled spins, the so-called degree of coupling can be classified as follows [31, 32]:

- If $\Delta\upsilon/\mathrm{J} \gg 1$, the spin system is called week coupled. By considering the relative proximity of alphabet letters, they are denoted as, for instance, AX, A_2X_2, AX_3, … where the subscript corresponds to the number of magnetically equivalent spins, which exhibit exactly the same interactions (e.g. same J-coupling) with other nuclei in the molecule. Therefore, magnetically equivalent nuclei have the same resonance frequency.
- If $\Delta\upsilon/\mathrm{J} \leq 1$, the spin system is called strongly coupled, and denoted as AB, A_2B_2, …
- For those intermediately coupled spin systems, they are denoted as AMX, AMNPQ, …

The J-coupling constant decreases rapidly as the number of chemical bonds between the coupled spins increases. The indirect spin-spin coupling involving 4 or more of chemical bonds is generally ignored.

Summarizing in terms of changes in the energy state of nuclei, a nucleus in a

magnetic field exhibit a single peak (single resonance frequency) in the spectrum as the energy state of the spin is split according to Equation (1.9) – (1.11) due to the Zeeman effect (Figure 1-8 (A), (B)). For a two-spin system with J = 0 between the spins, the chemical shift effect results in two singlets (i.e., two distinguishable peaks due to changes in resonance frequency) (Figure 1-8 (B), (D)). Assuming that the two spins are now indirectly coupled, the energies of the spin states slightly change due to the J-coupling, exhibiting two doublets (Figure 1-8 (C), (D)).

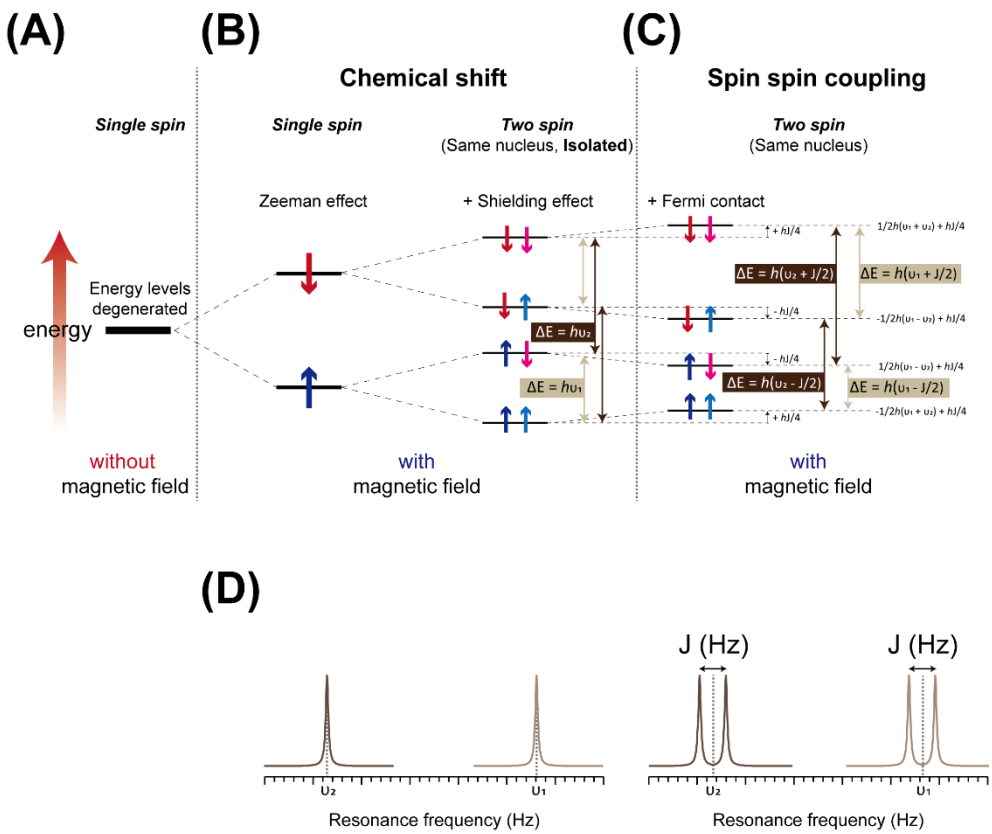


Figure 1-8. Change in energy state of spin due to spin-spin coupling. (A) without a magnetic field (B) with a magnetic field and without coupling state (C) spin-spin coupling considered (D) Resonance signal change form by spin-spin coupling

1.1.6. in vivo Metabolites

MRS can simultaneously acquire signals from multiple metabolites non-invasively. For instance, ^{1}H-MRS can detect various neurotransmitters in the brain, and ^{13}C-MRS can measure fluxes of metabolites non-invasively. ^{31}P-MRS can detect

metabolites involved in energy metabolism.

Since protons have high sensitivity compared to other nuclei, in vivo ^{1}H-MRS is a powerful technique that can distinguish, measure, and quantify a number of important biological compounds.

Figure 1-9 is a ^{1}H-MRS spectrum of a human brain at 3.0 T (healthy volunteer). In general, the number of metabolites that can be observed in vivo using ^{1}H-MRS is more than 20, and the following metabolites can be quantified in human brain: Acetate (Ace), Adenosine Triphosphate (ATP), Alanine (Ala), Ascorbic acid (Asc), Aspartate (Asp), Creatine (Cr), γ-aminobutyric acid (GABA), Glucose (Glc), Glutamate (Glu), Glutamine (Gln), Glutathione (GSH), Glycerophosphorylcholine (GPC), Glycine (Gly), Lactate (Lac), myo-Inositol (mI), N-Acetyl-Aspartate (NAA), N-Acetyl-Aspartyl-Glutamate (NAAG), Phosphocreatine (PCr), Phosphorylcholine (PCh), Phosphorylethanolamine (PE), Pyruvate (Pyr), scyllo-Inositol (sI), Serine (Ser), Succinate (Suc), Taurine (Tau), Threonine (Thr), Valine (Val), 2-Hydorxyglutarate (2-HG).

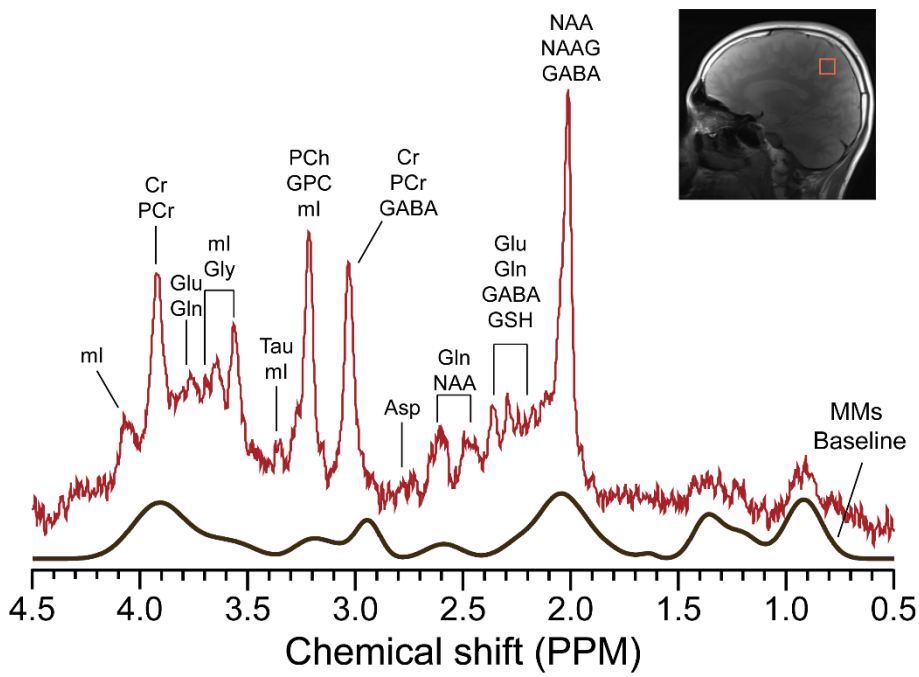


Figure 1-9. ^{1}H-MR spectrum acquired at 3.0 T of healthy subject.

These metabolites are functionally specialized in organs and regions therein. Their concentrations can be altered due to disease-specific changes in metabolite

pathways. For example, GABA and Glu are representative neurotransmitters in the brain, and their concentration changes are largely linked to psychiatric disorders. [33] The chemical shift and J-coupling constant values of a number of metabolites have already been well established. [34] This information is also used to build a metabolite basis set that is essential for metabolite quantification as described in Section 1.1.10.

Macromolecules (MMs) are composed mainly of mobile/immobile proteins (> 3kD), which are also important elements in ^{1}H-MRS of the brain in addition to metabolites. (Figure 1-9) MMs signal are distributed over the entire spectra region in ^{1}H-MRS of the brain, and therefore, significantly hinders metabolite quantification. Unlike metabolites, the spectral characteristics of MMs signal has not been well defined. [35, 36] The influence of the pathological condition of patients on the MMs signal is also yet to be clarified. [37, 38] The variability of the MMs signal has a non-negligible effect on metabolite quantification [39] and ultimately lowers the reliability and accuracy of the metabolite quantification in vivo. To address this issue, metabolite-nulled spectra can be acquired by taking advantage of the large T_1 difference between MMs and metabolites, and included in the metabolite quantification process. [40, 41] However, it is very time consuming to obtain metabolite-nulled spectra, which limits its clinical application. This will be dealt with in more detail in Chapter 2.

1.1.7. RF Pulses and Gradients

As described in the previous Section, RF pulses are used to manipulate magnetization. They can be classified into the following types in terms of their function in the data acquisition:

- Excitation pulse: It serves to flip the longitudinal magnetization onto the xy-plane, and can be further divided into non-self-refocusing and self-refocusing pulses depending on the phase distribution of the excited spins immediately after the pulse. The phase distribution of the spins generated by the RF pulses can effectively be removed for the latter such that the minimum delay required before the onset of data collection can be shorted.
- Saturation pulse: A saturation pulse is used to suppress signal from a specific region of the sample. For instance, MRS signal could also be contributed from an outer-volume (outside a voxel) due to imperfect voxel

localization. To minimize such a voxel contamination, saturation pulses (in combination with gradients) are irradiated to those regions outside the defined voxel.

- Inversion pulse: An inversion pulse aims to invert longitudinal magnetization. ($+M_z \rightarrow -M_z$). For instance, it is used for the inversion recovery-based water suppression by taking advantage of the large difference in T_1 between water and metabolites. In this case, the inversion delay between the inversion pulse and an excitation pulse is tuned at a temporal point such that the amount of the transverse magnetization of water generated by the excitation pulse is negligible compared with that of metabolites. It is this technique that is frequently used for the acquisition of the above-mentioned metabolite-nulled spectra.
- Refocusing pulse: This pulse flips transverse magnetization (M_{xy}) by 180° to induce realignment of the dephased spins and is an essential RF pulse in the spin echo experiment.

Another essential component of MRS is magnetic field gradient. Slice selection is achieved by applying a one-dimensional constant gradient while simultaneously applying a frequency selective RF pulse. MRS signal is detected from three orthogonal intersecting slices (voxel), (Figure 1-10), which will be discussed in detail in Section 1.1.9.

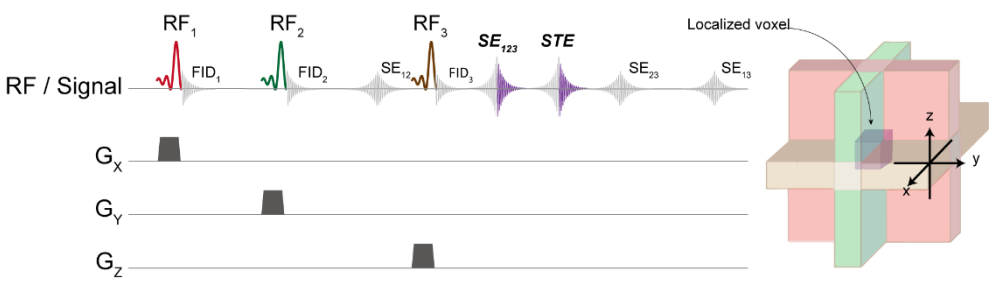


Figure 1-10. MRS signal acquisition process for a specific voxel. It is possible to form an MRS signal for a specific voxel (purple) from three RF pulses and gradient pairs. That is, by controlling the flip angle of RF pulses, timing between RF pulses, etc., the remaining signals can be suppressed while maximizing the desired signal type. (spin echo (SE), or stimulated echo (STE); purple)

The application of a gradient can result in eddy current. It can severely degrade the quality of MRS spectra. [42, 43] Therefore, additional postprocessing may be

performed if necessary. [42, 43]

1.1.8. Water Suppression

Water exhibits the strongest signal and can influence the entire spectral range of ^{1}H-MRS spectra. Therefore, effective suppression of water signal is the most important prerequisite to detecting other metabolite signal. Water signal can be suppressed during data collection using frequency-selective RF pulses and gradients . [44, 45] (Figure 1-11(A)-(B)). It can also be reduced to a certain extent by data postprocessing such as singular value decomposition (SVD). [46] (Figure 1-12(A)-(B))

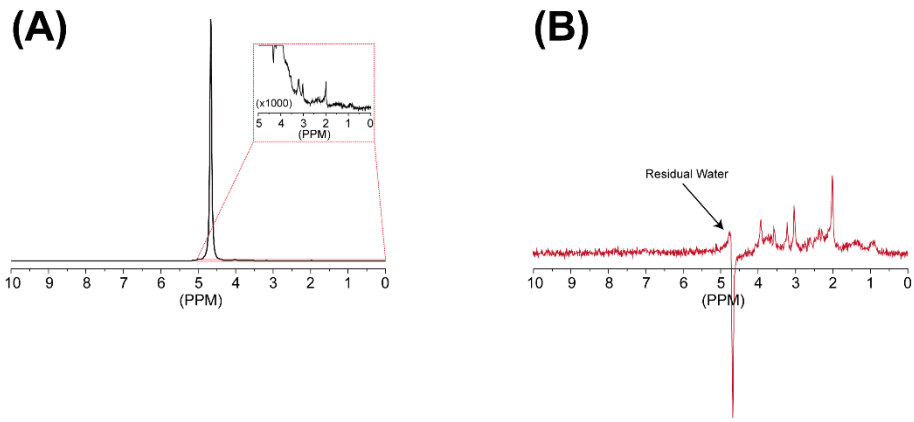


Figure 1-11. MRS signal difference according to whether water suppression module using RF pulses and gradients is applied. (A) water suppressed MRS signal (B) water unsuppressed MRS signal

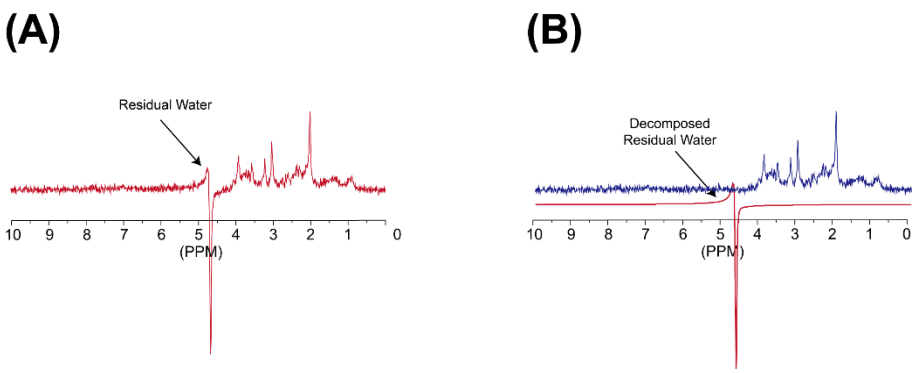


Figure 1-12. MRS signal difference according to whether post-processing method (HLSVD) is applied. (A) MRS signal before HLSVD application (B) MRS signal obtained by decomposing residual water signal using HLSVD

Both approaches have limitations. The performance of water suppression could also be subject to hardware imperfections (e.g., poor shimming and miscalibrated RF pulse power). The removal of water signal by postprocessing depends on the quality of the acquired MR spectrum (e.g., linewidth and signal-to-noise ratio (SNR)). Ineffective water suppression or removal can substantially influence result in the quantitative outcome.

1.1.9 Spatial Localization Methods in Single Voxel MRS

Spatial localization techniques allow for the extraction of region-specific metabolomic information. The performance of the techniques determines the integrity of the signal from the region of interest (ROI).

In order to selectively obtain an MRS signal from a voxel at a specific location of the human body, at least three frequency-selective RF pulse are required along with the current gradient pulses. Such a sequential collection of RF and gradient pulses is called a pulse sequence. The pulse sequences used for single voxel MRS are continuously being developed. Among them, the most frequently used pulse sequences at clinical field strength are as follows:

- Image Selected In vivo Spectroscopy (ISIS) [47]
- Stimulated Echo Acquisition Mode (STEAM) [48]
- Point Resolved Spectroscopy (PRESS) [49, 50]
- Spin-echo Full-intensity Acquired Localized Spectroscopy (SPECIAL) [51]

The ISIS localization method requires eight scans for 3D localization. (Figure 1-13(A)) That is, by turning the three inversion pulses on and off for each axis and switching the position of the receiver channel, the areas to be localized for each scan are defined and a signal is obtained using an excitation pulse (FID type). After performing the 8 scans, the 3D localized spectrum for the desired voxel is obtained by adding and subtracting the resulting spectra. The most important issue of ISIS localization is that a complete, spatially localized spectrum can be obtained only after all eight data are completely acquired. Thus, the performance of spatial localization could be very sensitive to the movement of the patient and the slice profile of the inversion pulses.

STEAM, PRESS, and SPECIAL (Figure 1-13 (B)-(D)) pulse sequences use only three RF pulses and slice selective gradients, and the types of RF pulses are

configured as follows (the resulting types of echo are denoted in parentheses):

- STEAM: Three excitation pulses (Stimulated echo)
- PRESS: One excitation and two refocusing pulses in order (Spin echo)
- SPECIAL: One inversion, excitation and refocusing pulses in order (Spin echo)

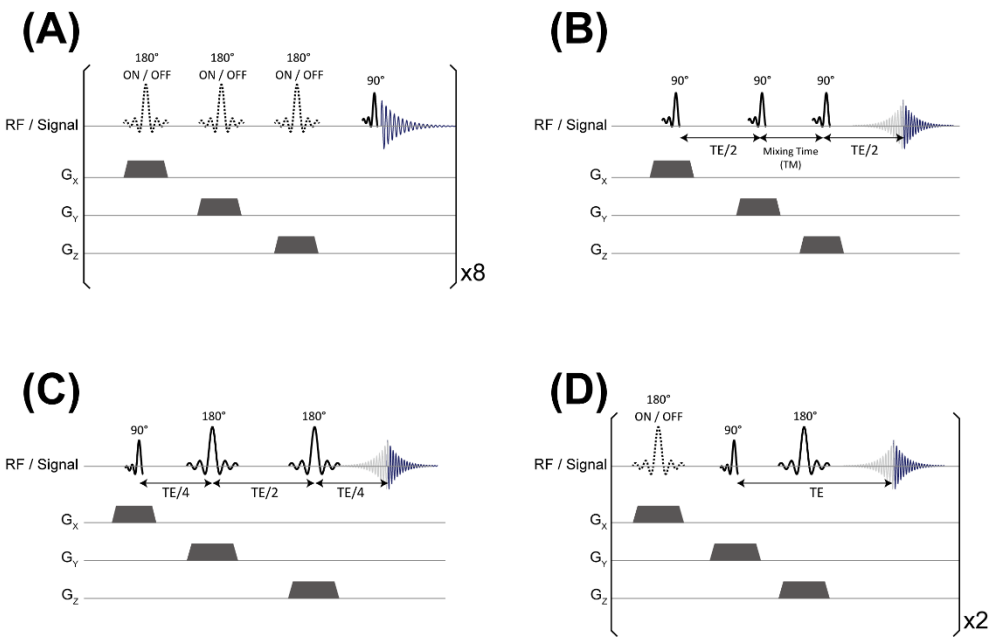


Figure 1-13. Simplified diagram of the (A) ISIS, (B) STEAM, (C) PRESS and (D) SPECIAL single voxel pulse sequence for MRS.

Theoretically, spin echo type pulse sequences (PRESS, SPECIAL) provide twice as high a signal yield as the stimulated echo type pulse sequences (STEAM). [48, 52] However, in general, the smaller the number of RF pulses, the better the localization performance. For instance, in the case of the SPECIAL sequence, since the ISIS localization scheme is employed along one direction, it also has the sensitivity to patient movement. Therefore, STEAM and PRESS pulse sequences that can localize on a single shot are typically used in clinical application.

All pulse sequences are subject to voxel contamination. The signal outside the ROI can sneak into the detected signal due to imperfect pulse profile and hardware. For example, if the ROI is close to the extracranial lipid (e.g., located in the prefrontal cortex), the lipid signal can almost completely ruin the MR spectrum. The most commonly used technique to minimize such unwanted signal is outer volume

suppression (OVS), which exploits the difference in spatial information inside and outside the ROI. [53] OVS can be placed together with the water suppression module in front of the spatial localization module. By placing at least 6 OVS slices around the ROI, MR signal with negligible contamination could be obtained.

1.1.10. Metabolite Quantification

Typical in vivo brain spectra suffer from severe spectral overlap between metabolite resonances and between metabolite and MMs signal due to the limited spectral dispersion, particularly at the clinical field strength. The spectral quality is exacerbated by the limited SNR due to scan time constraint.

Therefore, for quantification of individual metabolites from in vivo brain spectra, conventional spectral analysis methods such as peak area integration or simple line fitting without a prior knowledge is almost of no use. For this reason, advanced spectral fitting methods have been continuously developed in the form of software package such as QUEST [54], LCModel [55], and others. [56, 57]

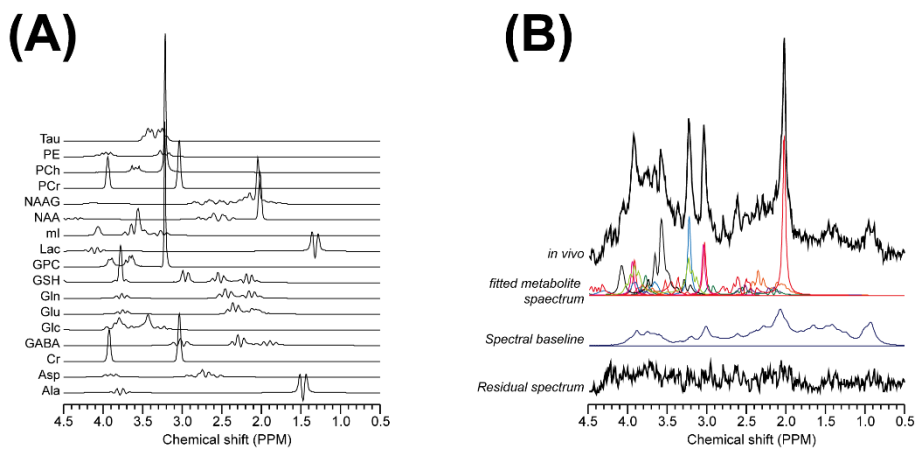


Figure 1-14. Representative spectra of a metabolite basis set and quantification process. A metabolite basis set (A) incorporates chemical shifts, spin-spin coupling, T_2 relaxation times, and line shapes of expected metabolites. The basis set is used to metabolite quantification from raw in vivo MR spectrum (B).

All of these methods employ nonlinear least squares fitting (NLSF) using a prior knowledge – namely the metabolite basis set, which means a collection of each individual metabolite signals either in time-domain (FID) or in frequency domain (spectrum). (Figure 1-14(A)) They are typically prepared either by actual

experiments in phantom or by quantum mechanical simulation incorporating J-coupling.

More specifically, for the measured basis set approach, a high concentration of an aqueous solution (50 mM <) is prepared for each metabolite and spectra are collected with the same sequence and sequence parameters that is to be used in the in vivo experiment. [55] To approximate the in vivo condition, the temperature (310 K) and pH (7.2 - 7.4) should be adjusted. Each metabolite signal should have high SNR. The measured basis set approach has a great advantage in that it largely reflects the actual experimental condition such as hardware imperfection (e.g., eddy current) and non-ideal spatial localization. [58]

The simulation approach requires solving the Equation of motion of density matrix for the given pulse sequence [59] taking account of each of the sequence elements such as RF pulses, gradients and inter-pulse delays. Due to computational complexity, approximations are frequently used such as hard pulses instead of actual shaped pulses and weak coupling even for strongly coupled spins. Next, an in vivo MRS signal (S_{invivo}) can be modeled approximately as

$$S_{invivo} \simeq \sum_{i=1}^{N} \widehat{M}_i(\theta) + b(\varphi) + \varepsilon. \tag{1.49}$$

where $\widehat{M}_i$ is the i-th model metabolite basis set signal of N metabolites, b is the background baseline signals such as MMs, and ε is the Gaussian noise including physiological noise (e.g., breathing, eye movement). Lastly, θ and φ are variables for fine tuning of each modeled signal, which have control over the amplitude of metabolite basis set signal, line-shape (Gaussian or Lorentzian), frequency, and phase. Equation (1.49) is iteratively solved in a nonlinear least squares sense to minimize the cost function given as

$$\underset{\theta,\varphi}{\operatorname{argmin}} \left\| S_{invivo} - \left(\sum_{i=1}^{N} \widehat{M}_i(\theta) + b(\varphi) \right) \right\|^2. \tag{1.50}$$

That is, it is a modification of the Levenberg–Marquardt algorithm **[60]** to find the minimal distance between the S_{invivo} and model functions ($\widehat{M}_i(\theta) + b(\varphi)$). (Figure 1-14(B))

Metabolite quantitation errors are caused by inadequate modeling of the metabolite and/or background signal overlapped with the measurement noise. The

error is usually calculated by estimating the Cramér-Rao lower bound (CRLB). [61] CRLB is a lower bound on the variance (σ^2) of estimators such as metabolite concentrations (amplitudes of the individual metabolite bases), linewidth and phase shifts, and calculated as the inverse of the Fisher information matrix (F). [62]

$$\sigma^2 \geq \mathrm{CRLB} = \mathrm{F}^{-1}. \quad (1.51)$$

CRLB is commonly expressed as %SD (σ, %). In general, the fitting results with a CRLB greater than 30% for quantified metabolite concentrations are excluded from the analysis. [63]

Finally, the amplitude of metabolites obtained through the fitting algorithm are often normalized to water and converted into a (semi) absolute concentration (mmol/kg_{ww}), [64] or expressed as a relative concentration to a reference metabolite (e.g. total creatine (= Cr+PCr)) that is relatively insensitive to physiological changes. [65]

1.2 Deep Learning

Artificial intelligence (AI) originated from the field of computer science in the 1950s and can be defined as a research activity aimed at automating intelligent tasks performed by ordinary humans.[66-69] Therefore, AI encompasses machine learning and deep learning, and could include many other methods that have no learning process at all. Early chess programs, for example, had only hard-coded rules, which were not accepted as machine learning then.

Symbolic AI,[70, 71] an approach to solving problems by making enough explicit rules, was developed, and became the dominant paradigm in the field of AI until the 1980s. Since then, attempts have been made to develop techniques to find a clear way to solve more complex and entangled problems such as image classification and speech recognition, and advanced methods using machine learning have emerged.

Machine learning is a research field that develops algorithms that allow machines to learn and execute unspecified actions from data, [72]. Typically, the data needs to be expressed such that the machine can process it, which is called feature transformation. Deep learning is a sub-concept of machine learning, which was developed as an efficient means for expression learning.[73, 74]

In deep learning, meaningful expressions can be learned progressively from successive layers. In this context, deep learning is also called 'layered representations learning' or 'hierarchical representations learning'. Deep learning has become the de facto standard for processing big data and complex high-dimensional data such as images, texts, and audio signals upon the use of high-performance hardware.

This Section deals with the basic principles of deep learning and the essential concepts that make up neural networks.

1.2.1. Training for Regression Model

A regression is a process of finding a solution from a model function that predicts a ground-truth (GT) value **y** given a corresponding input value **x**. In a linear regression, the model function is expressed as a linear combination of input variables. For instance, for an input vector **X**.

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w} + \mathrm{b} \tag{1.52}$$

where b is a bias, and **w** is the weight vector. $\hat{\mathbf{y}}$ represents the predicted vector, and the difference between **y** and $\hat{\mathbf{y}}$ may vary depending on **w** and b. Therefore, to find the optimal **w** and b, 1) a method for measuring the quality of the current model (loss function) and 2) a method for improving the model quality (optimization algorithm) are required.

First, in order to measure the quality of the model, we need to quantify the error in $\hat{\mathbf{y}}$. The loss for measuring the quality of the model for the entire dataset (n) can be expressed as follows.

$$\begin{aligned} \boldsymbol{L}(\mathbf{w}, \mathrm{b}) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}(\hat{\mathbf{y}}_{\boldsymbol{i}} - \mathbf{y}_{\boldsymbol{i}})\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} l_i(\mathbf{w}, \mathrm{b}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\left(\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_{\boldsymbol{i}} + \mathrm{b}\right) - \mathbf{y}_{\boldsymbol{i}}\right)\right)^2 \end{aligned} \tag{1.53}$$

where $\boldsymbol{L}$, l_i $\mathbf{x}_{\boldsymbol{i}}$, $\mathbf{y}_{\boldsymbol{i}}$ and $\hat{\mathbf{y}}_{\boldsymbol{i}}$ are loss for entire dataset, i-th loss, input data, GT data, and predicted data from the model, respectively. The function used to calculate the loss is the mean squared error (MSE), which is most widely used in regression

models.

Second, the optimization algorithm is usually a stochastic gradient descent algorithm with mini-batch applied. [75] Before optimization, w and b constituting the model are initialized with random numbers. More details on weight and bias initialization are described in Section 1.2.5.

After that, the optimization is performed iteratively in the direction of negative gradient of the loss function over the entire training dataset. [76] When the number of training data is very large, a mini-batch $\boldsymbol{B}$ is constructed by selecting a predetermined number of samples randomly and uniformly, and the average loss over the elements of $\boldsymbol{B}$ is used in the optimization. In each step of optimization in the gradient descent method, the gradient is multiplied by a predefined step size, namely a learning rate (η; $\eta>0$). That is,

$$(\mathbf{w}, \mathrm{b}) \leftarrow (\mathbf{w}, \mathrm{b}) - \frac{\eta}{\boldsymbol{B}} \sum_{i \in \boldsymbol{B}} \partial_{\mathbf{w},\mathrm{b}}\, l_i(\mathbf{w}, \mathrm{b}). \tag{1.54}$$

The $\boldsymbol{B}$ and η are not the values found through model training, but are values that the user must select manually. Therefore, these values are called hyperparameters. Hyperparameter tuning refers to optimizing these values, and the tuning process can be performed by using various methods. [77] The training process of a neural network consists of forward propagation, which calculates and stores variables from the input layer to the output layer, and back propagation, [78, 79] which calculates gradients for the neural network parameters.

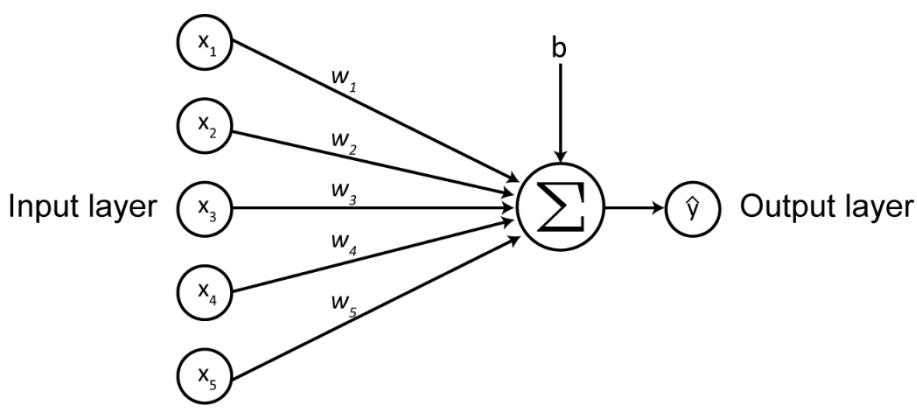


Figure 1-15. Graphical representation of a simple regression process.

Model optimization is accomplished by finding the optimal values of the

neural network parameters while forward propagation and backpropagation are alternately performed. The technical details of forward propagation and backpropagation are provided in Section 1.2.3. After training, the estimated $\hat{\mathbf{w}}$ and $\hat{b}$ are stored. Then, the learned linear regression model is used for inference (prediction) on unseen data $\mathbf{x}$ that is not in the training dataset.

The dimension of the input in the neural network is called a feature dimension. For example, Figure 1-15 shows a neural network with a feature dimension of 5 and an output dimension of 1. The total number of layers in this network is one, and since all inputs are connected to all outputs, this layer is called a fully connected layer or dense layer.

1.2.2. Training for Classification Model

Unlike regression, which predicts successive values, predicting the probability of belonging to a specific category is called classification. Therefore, the model presented in the previous subsection is difficult to apply to the classification problem.

There are two ways to represent GT labels in the classification problems. The first way is to define class with $y \in \{1,2,3\ldots\}$. Although this method can be efficiently stored in the computer, it is more suitable for regression problems, and the order of these numbers is meaningless in problems of classification. The second way is to label classes in the form of $y \in \{(1,0,0,\ldots),(0,1,0,\ldots),(0,0,1,\ldots)\}$ through one-hot-encoding. That is, when the number of classes is n, each class is labeled as an n-dimensional vector.

For classification, the number of output neurons is set equal to the number of classes. The values of output neurons specify the likelihood of the input belonging to the respective classes of the neurons. Therefore, they need to be normalized properly.

Since classification problems require discrete prediction results, it is necessary to express the confidence level for each class. Therefore, unlike linear regression in Equation (1.52), a softmax logistic regression process is added in a multiclass classification model. The softmax logistic regression result is subject to nonlinearity such that the sum of all results becomes 1, and each result has a value

between 0 and 1. To avoid confusion in the expressions, rewrite Equation (1.52) as follows.

$$\mathbf{o} = \mathbf{X}\mathbf{w} + \mathrm{b} \tag{1.55}$$

where $\mathbf{o}$ is the output vector calculated with weight and bias as $\hat{\mathbf{y}}$ in Equation (1.52). Then, softmax logistic regression is expressed as

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathbf{o})$$
$$\hat{\mathrm{y}}_i = \frac{\exp(\mathrm{o}_i)}{\sum_{i=1}^{n} \exp(\mathrm{o}_i)}. \tag{1.56}$$

where $\hat{\mathbf{y}}$ is a vector of probability for each class (Figure 1-16) and o_i and $\hat{\mathrm{y}}_i$ is the i-th output value and probability value, respectively. Therefore, $0 \le \hat{\mathrm{y}}_i \le 1$ for all classes, and the class with the highest probability for the input vector X can be found simply as $\underset{i}{\mathrm{argmax}}\, \hat{\mathrm{y}}_i$.

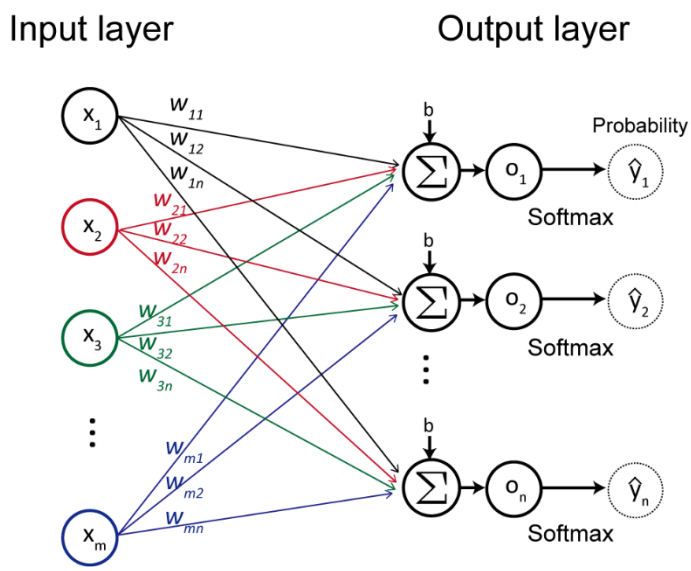


Figure 1-16. Graphical representation of the logistic regression process.

To measure how accurate the output probability is, a loss function is needed as in the linear regression. In this softmax regression, the negative log-likelihood (NLL) estimation is typically used for that purpose, which is given as

$$-\log p(\mathrm{Y}|\mathrm{X}) = -\sum_{i=1}^{n} \log\, p(\mathrm{y}_i|\mathrm{x}_i). \tag{1.57}$$

Using Equation (1.57), the loss function for n-class problem is defined as

$$\boldsymbol{L}(\mathbf{w}, \mathrm{b}) = -\log p(\mathrm{Y}|\mathrm{X}) = -\sum_{i=1}^{n} \mathrm{y}_i \log \hat{\mathrm{y}}_i. \tag{1.58}$$

Equation (1.58) is known as a cross-entropy loss. This is the most commonly used loss in classification problems and is a concept derived from information theory. [80, 81] The training process of the classification model using softmax regression is not different from the contents described in Section 1.2.1. However, classification accuracy is used as a method of evaluating the performance of the model, which is equal to the ratio of the number of correct predictions to the number of predictions for the entire dataset.

Finally, the softmax function is one of the representative non-linear activation functions constituting an artificial neural network. More details such as the types of activation functions are described in Section 1.2.3.

1.2.3. Multilayer Perceptron

Multilayer perceptron (MLP) refers to a class of neural networks that consist of at least three layers of nodes (neurons), namely an input layer, a hidden layer, and an output layer, and multiple layers can be stacked within a hidden layer compartment. (Figure 1-17). The output of each layer is fed into the next layer.

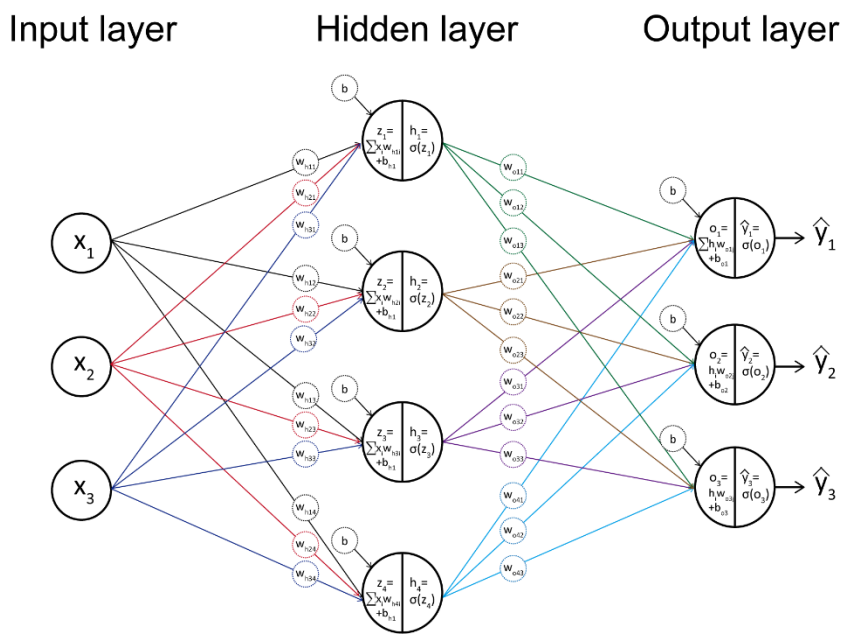


Figure 1-17. Graphical representation of multilayer perceptron with single hidden layer.

The main characteristic of the MLP model is that the layers are not connected with a linear relationship except for the input layer due to the non-linear operations that are performed on the output of every layer.

An MLP for classification including only one hidden layer can be expressed as follows.

$$\mathbf{z} = \mathbf{x}\mathbf{W_1} + \mathrm{b}_1 \tag{1.59}$$

$$\mathbf{h} = \sigma(\mathbf{z}) \tag{1.60}$$

$$\mathbf{o} = \mathbf{h}\mathbf{W_2} + \mathrm{b}_2 \tag{1.61}$$

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathbf{o}) \tag{1.62}$$

where **z** and **o** represent variables before nonlinear operation. Also, $\mathbf{W_1}$, b_1 are the weight vectors and bias between the input layer and the hidden layer, $\mathbf{W_2}$, b_2 are the weight vectors and bias between the hidden layer and the output layer, respectively, and σ represents the nonlinear activation function. As mentioned earlier, the final output $\hat{\mathbf{y}}$ cannot be described as a linear combination of the weight vectors and biases as a result of the application of the nonlinear activation functions expressed in Equation (1.60) and (1.62). Also, Equation (1.59) – (1.62) can simply be expressed in the form of a layer as shown below.

$$\mathbf{h} = \sigma(\mathbf{x}\mathbf{W_1} + \mathrm{b}_1) \tag{1.63}$$

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathbf{h}\mathbf{W_2} + \mathrm{b}_2) \tag{1.64}$$

Widely used nonlinear activation functions include rectified linear unit (ReLU), Sigmoid and Tanh, and since the output characteristics of these functions are different, they can be selectively used according to the purpose. [82]

MLP is known to act as a universal approximator if enough nodes and precise weights are provided. [83] MLP is also called a feed-forward deep neural network (FFDNN) because it operates in the forward direction.

On the other hand, in the process of optimizing artificial neural networks including MLP, backpropagation is required as mentioned in Section 1.2.1. Backpropagation is performed to update the trainable parameters after finding the gradient of the loss function moving in the reverse direction from the output to the input layer according to the chain rule. In more detail, first, the gradient $\partial \boldsymbol{L}/\partial \mathbf{W_2}$ of the loss function for the model parameter closest to the output layer (W_2; bias is ignored for simplicity) is calculated by applying the chain rule.

$$\frac{\partial L}{\partial \mathbf{W_2}} = \frac{\partial L}{\partial \mathbf{o}} \frac{\partial \mathbf{o}}{\partial \mathbf{W_2}} = \frac{\partial L}{\partial \mathbf{o}} \mathbf{h}^{\top} \tag{1.65}$$

As in Equation (1.65), this process multiple operations are performed such as transposition (e.g., $\mathbf{h}^{\top}$ in Equation (1.65)) and multiplication. The above principle applies equally to all chain rule processes below. Next, the gradient of the loss function $\partial L/\partial \mathbf{h}$ for the hidden layer variable **h** is performed as below

$$\frac{\partial L}{\partial \mathbf{h}} = \frac{\partial L}{\partial \mathbf{o}} \frac{\partial \mathbf{o}}{\partial \mathbf{h}} = \mathbf{W_2}^{\top} \frac{\partial L}{\partial \mathbf{o}}. \tag{1.66}$$

Since the activation function is applied on a per-element basis, the gradient for the intermediate variable **z** is

$$\frac{\partial L}{\partial \mathbf{z}} = \frac{\partial L}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{z}} = \frac{\partial L}{\partial \mathbf{h}} \odot \sigma'(\mathbf{z}). \tag{1.67}$$

In order to calculate Equation (1.67), unlike Equations (1.65) – (1.66), it is necessary to perform an element-wise multiplication (operator ($\odot$)). Finally, the gradient of the model parameter ($\mathbf{W}_1$) closest to the input layer is calculated as follows.

$$\frac{\partial L}{\partial \mathbf{W_1}} = \frac{\partial L}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{W_1}} = \frac{\partial L}{\partial \mathbf{z}} \mathbf{x}^{\top} \tag{1.68}$$

The gradient descent algorithm described in Section 1.2.1 is a method of updating the model parameters until the gradients calculated from Equation (1.65) – (1.68) are minimized. That is, as in Equation (1.54), parameter update is performed by subtracting the gradient from the current parameter values during the backpropagation process. There are also other optimization algorithms utilizing gradient values in various ways. [84, 85]

Backpropagation reuses intermediate values (e.g., Equation (1.59)) stored during forward propagation to avoid duplicate calculations. This means that the intermediate values must be held until backpropagation is complete. This is one of the reasons why backpropagation requires much more computing memory compared to forward propagation. Also, the size of these intermediate values is roughly proportional to the number of layers and the batch size. Therefore, training a deeper network using a larger batch size runs the risk of encountering out-of-memory errors.

Data applied as an input to MLP is processed in a one-dimensional form. In other words, all values in the data are considered as equally important. In order to process high-dimensional data such as medical images, MLP can still be used, but there is a critical limitation. Since MLP needs to transform a high-dimensional

feature into a one-dimensional tensor, it only learns information with lack of the spatial structure of the input. Therefore, its performance could be subject to variability even upon a small change in the element of the input data (e.g., a few pixel values). Also, as the size of the data increases, the size of the neural network should also increase accordingly, resulting in a problem of high computational cost. [86] Therefore, artificial neural networks that reduce the inefficiency of the training process and improve the performance have been continuously developed. [87-89] A convolutional neural network (CNN) is one of them [88, 90], which is described in detail in 1.2.6.

1.2.4. Model Evaluation and Selection

The goal of deep learning is to discover common patterns in the data it is dealing with. However, the data used for training the model may not represent the entire data characteristics. Considering the limited sample size, those features contained in the data not participating in the training could be missed. Therefore, the model can overfit to the distribution of the training samples rather than to the true distribution of the data. It can also underfit to the samples, for instance, because the model is too simple. These two problems should always be addressed simultaneously whenever possible, and overfitting is usually the main reason for the degradation of the performance of the model.

The methods frequently used to reduce overfitting are as follows: [91]

- Increase the number of data
- Reduce the complexity of the model
- Apply weight regularization
- Add dropout layer [92]

When the number of data is small, the model easily memorizes a specific pattern or noise present in the data set, so the possibility of overfitting increases. The number of data can artificially be increased by data augmentation.

The range of the types of the functions that a model can approximate is called model capacity, which is known to increase as the numbers of hidden layers and neurons increase. To prevent overfitting, the capacity of artificial neural networks should be reduced.

Weight regularization is used for model optimization along with the loss

functions as in Equation (1.53) to avoid overfitting. For weight regularization, l_1 or l_2-regularization are most widely used.

- l_1 Regularization: inclusion of the sum of the absolute values of all weights in the loss function.

$$\boldsymbol{J} = \boldsymbol{L}(\mathbf{w}, \mathrm{b}) + \lambda\|\mathbf{w}\|_{\mathbf{1}} \tag{1.69}$$

- l_2 Regularization: inclusion of the sum of squares of all weights in the loss function.

$$\boldsymbol{J} = \boldsymbol{L}(\mathbf{w}, \mathrm{b}) + \lambda\|\mathbf{w}\|_{\mathbf{2}} \tag{1.71}$$

where $\boldsymbol{J}$ is an objective function, which is a function used for model optimization problems, and is more commonly used in the model training than the loss function described above. A hyperparameter λ is employed to control the relative contribution of the regularization term. If λ is large, the model prioritizes reducing the loss resulting from the regularization term.

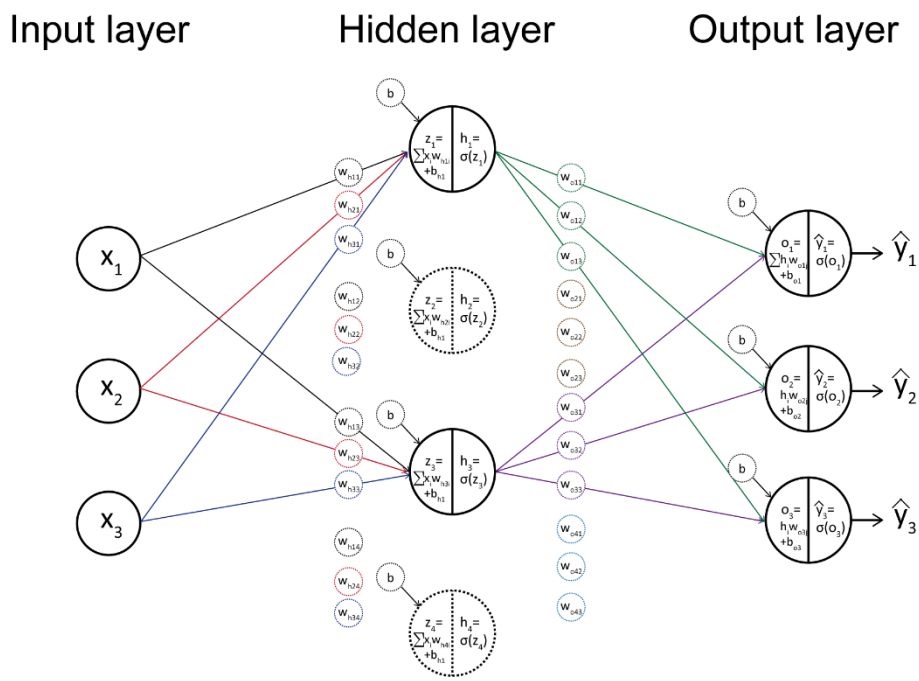


Figure 1-18. Graphical representation of drop out (dotted circles in hidden layer).

In dropout operation, [92] a part of neurons are not used (dropped out) during training. (Figure 1-18) the neurons to be dropped out is selected randomly, and their proportion is predetermined by dropout rate that is also a hyperparameter. Since randomly selected neurons are excluded from the training process each time,

overfitting could be prevented as in ensemble modeling.[93] It is common to use dropout only during training. However, it could also be used at test time as well for Bayesian inference. [94]

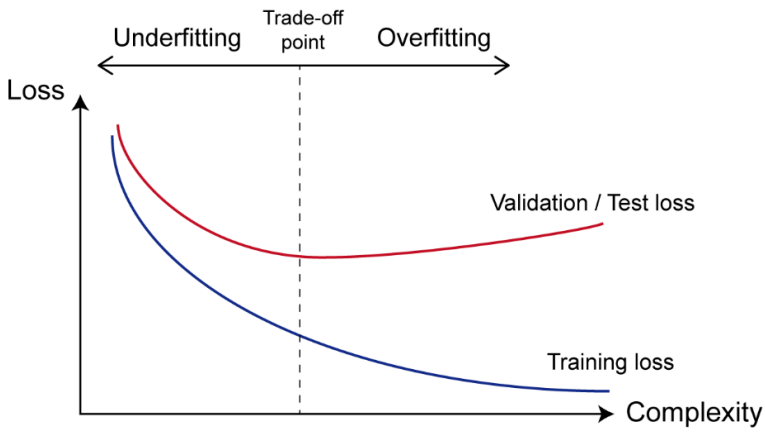


Figure 1-19. Change in loss value according to model complexity.

Overfitting and underfitting problems can usually be identified by estimating the error. (Figure 1-19) Errors in deep learning are divided into 1) training errors and 2) generalization errors. (validation and test errors) A training error is an error that occurs in the training process using a training dataset. A generalization error refers to an error that occurs on previously unseen data such as a test data set.

In the training process, a model is usually selected after evaluating the performance of several candidate models. This process is called model selection. Candidate models can simply be a group of models with different values of hyper-parameters and different types of activation functions. Model selection should not depend solely on the training data because the generalized error can never be expected to be similar to that of the training error. It is possible to check overfitting or underfitting continuously during training using an additionally secured data set other than the training and test data (i.e., validation data set).

The optimal model could also be selected using K-fold cross validation. (Figure 1-20) [95] In K-fold cross validation, the original data is divided into K non-overlapping partial datasets. The model is trained on K-1 datasets and validated on the rest one dataset in turn. In this way, all data are involved in the validation. The optimal model is selected by checking the training and validation errors of the K models.

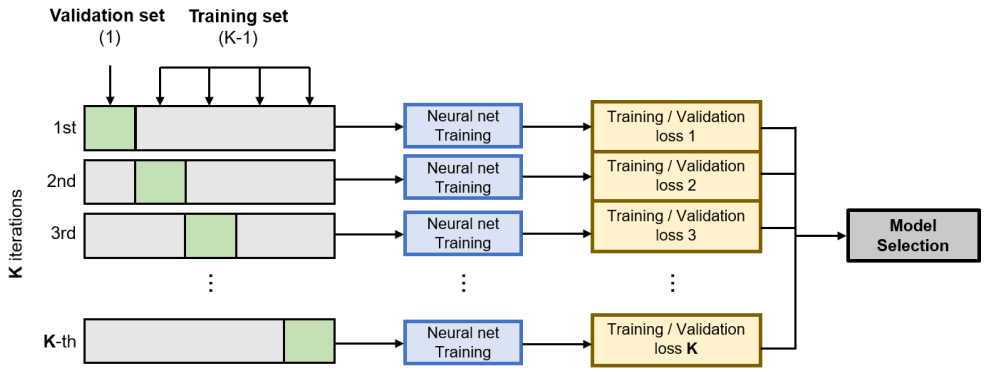


Figure 1-20. Graphical representation of K-fold cross validation.

1.2.5. Training Stability and Initialization

During the training phase of neural networks, a phenomenon in which the gradient gradually decreases as the backpropagation progresses towards the input layer may occur, which is called gradient vanishing. [96, 97] Conversely, the gradient may increase gradually, and the weights become abnormally large and eventually diverge, which is called gradient exploding. [96, 97] Both problems adversely affect training and lead to the inability to find an optimal model. The simplest way to alleviate gradient vanishing is to use ReLU, Leaky ReLU, or Exponential Linear Unit (ELU) as the activation function of hidden layers. Under this condition, a training stability can be achieved by minimizing the saturation effect in which the output value of the activation function converges to 0 or 1 (i.e. the gradient becomes 0). Gradient clipping [98] can be used to prevent gradient exploding by limiting the gradient value not to exceed a threshold value.

The gradient vanishing/exploding problems can also be alleviated by initializing weights and biases. Even if the same model is trained, the training results may vary depending on the initial values of the weights and biases. In general, biases are all initialized to 0, and weights are usually initialized by LeCun [99], Xavier [100], or He [101] initialization

Another way to prevent gradient vanishing/exploding is to use batch normalization (BN). [102] BN is known to improve the gradient vanishing and overfitting problem by making the learning process efficient by normalizing the input of each layer of the neural network with the mean and variance. [102]

1.2.6. Convolutional Neural Networks

In the visual cortex, only a part of neurons residing in a specific area activate in response to external stimuli, and the specific area is called receptive field. **[103]** In other words, it refers to the area of the sensory periphery where stimuli can affect the electrical activity of cells. Receptive fields can overlap each other, and overlapping receptive fields can respond to more complex information (objects) by combining low-level patterns (surfaces, edges). Based on this concept, a CNN was developed. [88, 90]

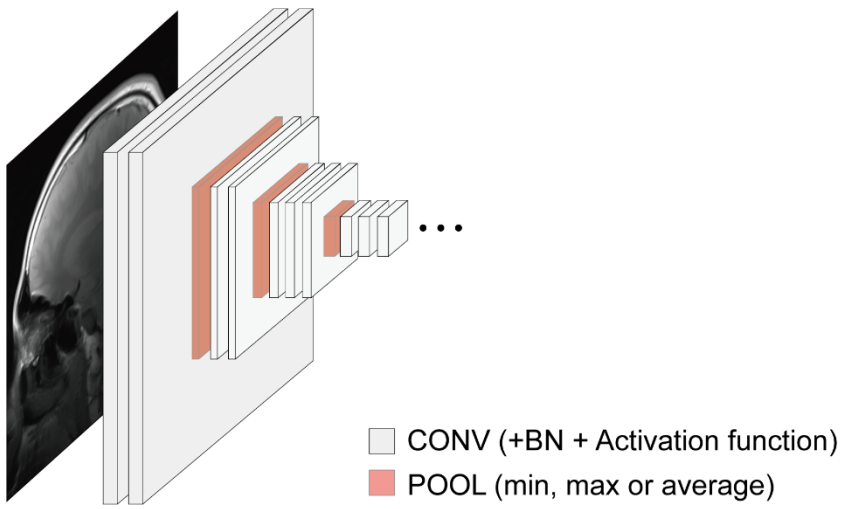


Figure 1-21. Part of the hierarchical structure of a typical convolutional neural network.

As discussed in Section 1.2.3, MLP is not capable of learning the high-order information such as correlated pixels within the data. On the other hand, CNN can learn such correlations between adjacent signals, and even extract the strongest (or weakest) feature. Among the layers constituting CNN, the former is achieved by the convolutional layers, and the latter is obtained by the pooling layer. [74] In Figure 1-21, CONV means a convolution operation. The result of the convolution operation passes through a nonlinear activation function such as ReLU, which is also included in the convolution layer. (BN is also included between convolution and activation [102]) After that, a pooling operation (POOL) is performed.

The convolution layer extracts data features. That is, the filter (kernel) scans over the entire data by multiplying the filter with the elements of the data and resulting in a map of activations. (Figure 1-22 (A)) Therefore, the result of convolution operation using filters is called a feature map. One can set the size and

movement range of the filter. The movement range of the filter is defined as the stride. Therefore, each convolution operation does not necessarily use all data dimensions. As a result, CNN requires a much smaller number of weights than MLP and yet is capable of preserving the spatial structural information of the input. If multiple convolutional layers are stacked, the resulting feature map will be much smaller than the initial input. To keep the size of the feature map the same as the size of the input after the convolution operation, padding is used, which is a function that adds a specified number of rows and columns of zeroes at the edge of the input. (Figure 1-22 (B))

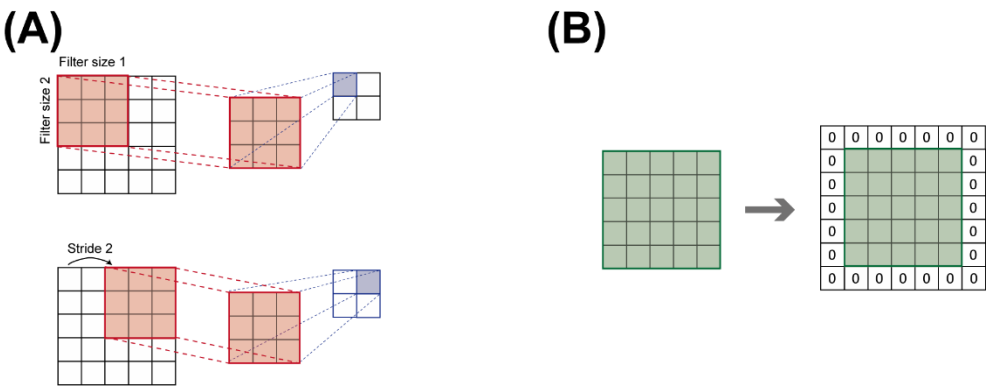


Figure 1-22. Functional elements composing the convolutional layer. (A) filter stride (B) zero padding

In general, a pooling layer is added after a convolutional layer. Typically, max pooling and average pooling are used for pooling operations. [104] The pooling layer downsamples the feature map to reduce its size. Therefore, the use of pooling has the effect of reducing the number of parameters of the model.

CNN can be designed much deeper than MLP. The resulting capacity of the model can also be extended accordingly. Therefore, the former can outperform the latter.

In a preliminary study (data not shown), the potential application of MLP in deep learning-based ^{1}H-MRS was investigated. However, the results were not satisfactory due to the difficulty of training the neural network as the number of hidden layers increases in an effort of improving its performance. The convolutional layer, which is an essential component of a CNN effectively extracts local patterns or features in the data. As such, it is known to work better than an MLP for data types such as audio and electroencephalogram (EEG). [105, 106] In ^{1}H-MRS, each

metabolite has its unique resonance frequency and lineshape (i.e., local pattern in one-dimensional space) depending on the chemical shift and J-coupling of the protons residing in that metabolite. In this research, therefore, CNNs are used exclusively. It will be shown in the subsequent chapters that CNN has a great potential in deep learning-based MRS by efficiently handling the complexity of the spectra through partial connection and weight sharing.

1.3 Purpose of the Research

Figure 1-23 summarizes the development of the metabolite quantification methods in the field of ^{1}H-MRS of the brain in chronological order along with the achievement and limitation of the individual methods.[56, 107-121], [122] Given the limitation of the current methods such as LCModel [55] and jMRUI (software package including HLSVD [118], AMARES [119], QUEST [115] in Figure 1-23) [114], the purpose of this study is to develop deep learning-based methods that address the following issues in ^{1}H-MRS.

(1) Sensitivity of the performance of the existing ^{1}H-MRS spectrum analysis software to the initial setting of the software and the quality of the spectra

(2) Limited information as to the uncertainty in the quantitative results of metabolite content. Currently, CRLB is most widely used as a measure of the reliability of spectral fitting. However, it provides only the minimum possible errors, which are not necessarily highly correlated with the actual errors. The current data inclusion criteria based on a cutoff CRLB value is also known to be subject to statistical bias [123].

Chapter 2 demonstrate the potential of a deep learning-based ^{1}H-MRS. It is shown that the metabolite content estimated from in vivo human brain spectra by using the deep learning-based method is comparable to that with the existing nonlinear least squares fitting method. Furthermore, the proposed method is more robust against degradation of the spectral quality.

Chapters 3 and 4 are dedicated to developing deep learning-based methods that provide not only the metabolite content but also the uncertainty therein. In Chapter 3, a heuristic, big data-driven method is introduced for quantitative uncertainty estimation. In Chapter 4, a theory-oriented, [94, 124] Bayesian deep

learning-based method is proposed.

Finally, Chapter 5 summarizes the achievement from the studies and discusses the follow-up research that can potentially further advance the metabolite quantification methods in ^{1}H-MRS.

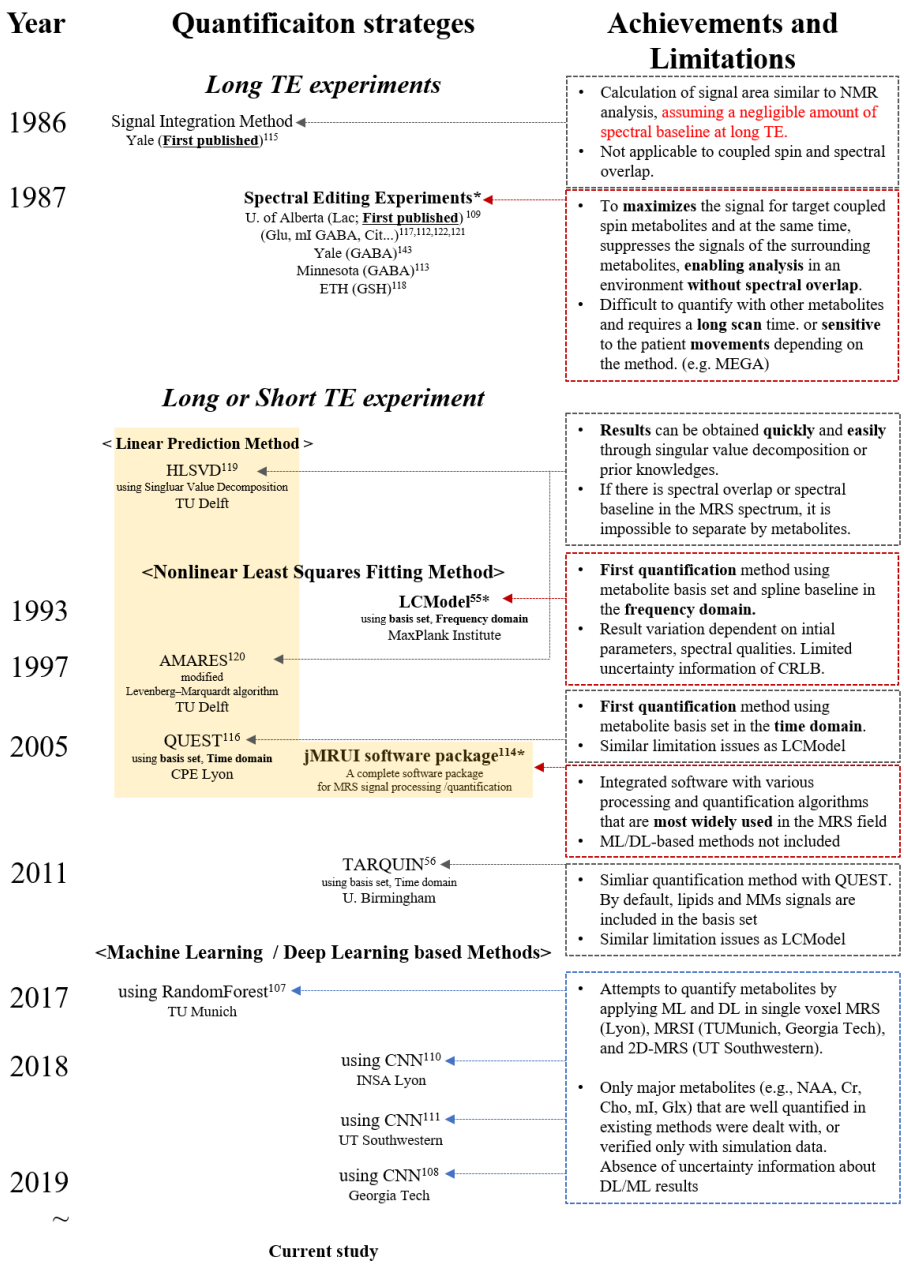


Figure 1-23. Brief development status of metabolite quantification methods in ^{1}H-MRS of the brain so far. Quantification methods that are frequently used up to now are marked with *.

1.4 Preparation of MRS Spectra and Their Usage

In this research, various types of spectra such as phantom spectra, simulated spectra, and in vivo spectra (both animal and human) are used for specific purposes. The preparation methods and usage of in vivo and phantom spectra are summarized in Table 1-2 and Table 1-3, respectively. The preparation methods and usage of simulated brain spectra are summarized in Table 1-4. Finally, Table 1-5 provides the information about the rest of the spectra that are also used in this research with different purposes.

	Acquisition Information of the in vivo Spectra		
	Chapter 2 Section 2.2.1	Chapter 3 Section 3.2.1	Chapter 4 Section 4.2.2
Magnetic field strength	3.0 T	9.4 T	3.0 T
Magnet model	Siemens Trio (VB17A)	Agilent 160AS	Siemens Trio (VB17A)
Coil	32-channel head coil	Single channel surface coil	32-channel head coil
Target	Human Brain	Rat Brain	Human Brain
Number of subjects	5	15	5
	Pulse sequence parameters of the in vivo Spectra		
Pulse sequence	PRESS	SPECIAL	PRESS
Repetition time / Echo time	2 sec / 30 ms	5.5 sec / 3.45 ms	2 sec / 30 ms
Spectral width	2 kHz	5 kHz	2 kHz
Data points	2048	2048	2048
Phase cycling	8 step	32 step	8 step
Voxel position	Left frontal lobe	Cerebellum / Thalamus	Left frontal lobe
Voxel size	8 cm^3 (2 x 2 x 2 cm^3)	23.5 mm^3 (3.0 x 2.8 x 2.8 mm^3)	8 cm^3 (2 x 2 x 2 cm^3)
Number of signal averages	64 (8 averages x 8 times)	384 (32 averages x 12 times)	64 (8 averages x 8 times)

Table 1-2. Preparation and usage of in vivo brain spectra.

	Acquisition Information of the Phantom Spectra		
	Chapter 2 Section 2.2.2	Chapter 3 Section 3.2.3	Chapter 4 Section 4.2.2
Purpose of use	Metabolite basis set	Metabolite basis set	Metabolite basis set
Magnetic field strength	3.0 T	9.4 T	3.0 T
Magnet model	Siemens Trio (VB17A)	Agilent 160AS	Siemens Trio (VB17A)
Coil	32-channel head coil	Single channel surface coil	32-channel head coil
Target	300 mL glass bottle	50 mL cylindrical tubes	300 mL glass bottle
Number of subjects	17 (each metabolite)	3 (different combination of the metabolite concentration)	17 (each metabolite)
Considered metabolites	Ala, Asp, Cr, GABA, Glc, Gln, Glu, GSH, GPC, Lac, mI, NAA, NAAG, PCr, PCh, PE, Tau		
Concentration	50mM; NAAG at 10mM, PCh at 36.05mM	50mM	50mM; NAAG at 10mM, PCh at 36.05mM
pH	7.0 – 7.5	7.0 – 7.5	7.0 – 7.5
	Pulse sequence parameters of the Phantom Spectra		
Pulse sequence	PRESS	SPECIAL	PRESS
Repetition time / Echo time	2 sec / 30 ms	10 sec / 3.45 ms	2 sec / 30 ms
Spectral width	2 kHz	5 kHz	2 kHz
Data points	2048	2048	2048
Phase cycling	8 step	32 step	8 step
Voxel position	Isocenter	Isocenter	Isocenter
Voxel size	8 cm^3 (2 x 2 x 2 cm)	23.5 mm^3 (3.0 x 2.8 x 2.8 mm)	8 cm^3 (2 x 2 x 2 cm)
Number of signal averages	256	128	256

Table 1-3. Preparation and usage of brain phantom spectra.

	Information of the Simulated Brain Spectra		
	Chapter 2 Section 2.2.3	Chapter 3 Section 3.2.4	Chapter 4 Section 4.2.2
Language	Python 3.5.2	Python 3.6	Python 3.5.2
Total number of simulated spectra	50,000	110,000	100,000
Training / Validation / Test set	40,000 / 5,000 / 5,000	80,000 / 20,000 / 10,000	80,000 / 10,000 / 10,000
	Information of the Metabolites in the Simulated Brain Spectra		
Total number of metabolites	17	27	17
Type of metabolite basis set	Phantom	Simulated (GAMMA)	Phantom
Concentration range	Table 2-1	Figure 3-1	Table 2-1
	Information of the MMs baseline in the Simulated Brain Spectra		
Total number of MMs group	9	10	10
Total number of MMs linefunction	17	25	10
Relative MMs group amplitudes	Table 2-2	Table 2 in reference 156	Table 2-2
	Information of the Spectral Quality of the Simulated Brain Spectra		
Linewidth range	10 – 20 Hz	10 – 30 Hz	10 – 20 Hz
SNR range	5 – 15	5 – 30	5 – 15
Frequency shift range	± 10 Hz	± 21 Hz	± 10 Hz
Zeroth order phase shift range	± 5°	± 1°	± 5°

Table 1-4. Preparation and usage of simulated brain spectra.

	Information of the Simulated, in vitro or Modified in vivo Spectra		
	Chapter 3 Section 3.3.2	Chapter 4 Section 4.2.2 & 4.2.4	Chapter 4 Section 4.2.2 & 4.2.4
Purpose of use	Performance compared to LCModel	Responses for OOD data	Performance compared to LCModel
Data type	in vitro phantom	Simulation	modified in vivo
Total number of data set	54	3,000	50
Linwidth range	10 – 20 Hz	4.9 – 40.3 Hz	Table 4-2
SNR range	5 – 30 Hz	3.5 – 40	Table 4-2

Table 1-5. Preparation and usage of simulated, in vitro or modified in vivo spectra.

Chapter 2. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain

2.1. Introduction

^{1}H-MRS provides a noninvasive means of extracting biochemical information from the brain. Typically, spectra are acquired at a shortest TE for maximal signal yield against signal decay due to transverse relaxation (T_2) and for coupled spins J-modulation as well. Then, spectral fitting is performed either in the time-domain [116] or in the frequency-domain [55] for the quantification of individual metabolites. At clinical magnetic field strength, however, in vivo brain spectra are subject to low SNR and spectral overlap between metabolite signals, which are exacerbated by limited field homogeneity across the spectroscopic voxel. They are known to strongly influence the quantitative outcome. [55] Although data post-processing is routinely performed for improved SNR and/or linewidth by applying filters on the FID signal in the time domain, their performance is limited due to different FID signal decay rates for different spin systems. [58, 125]

In addition, metabolite signal at short TE is always superimposed on the spectral baseline across the entire spectral range, which is mainly contributed by MMs. [35] Unlike SNR and spectral overlap between metabolite signals that can be improved by higher main field strength and a better shimming technology, spectral baseline is unavoidably present in the spectra at short TE, and is the major hindrance for accurate metabolite quantification by fitting. [55, 126] To this end, development of a means of robust metabolite quantification from the brain spectra that are severely degraded by low SNR, broadened linewidth and unknown spectral baseline has long been a major technical issue in ^{1}H-MRS.

Given the recent accomplishment of deep learning in a variety of different

The work presented in this chapter has been published; Lee, H. H., & Kim, H. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. Magn Reson Med 2019, 82(1), 33-48. doi:10.1002/mrm.27727.

A part of the work presented in this chapter in association with the spectra simulation has been published; Lee, H. H., & Kim, H. Parameterization of spectral baseline directly from short echo time full spectra in 1H-MRS. Magn Reson Med 2017, 78(3), 836-847. doi:10.1002/mrm.26502.

tasks [74, 127] and its potential application in ^{1}H-MRS [107, 110, 111, 128, 129], this chapter aimed to develop and evaluate a CNN that maps degraded generic in vivo brain spectra at short TE into noise-free, line-narrowed, baseline-removed metabolite-only spectra (intact metabolite spectra). The subsequent metabolite quantification from such intact metabolite spectra is achieved by solving a simple inverse problem using a metabolite basis set. The CNN was trained on simulated brain spectra. The robust performance of the proposed method against poor SNR and linewidth and unknown spectral baseline was tested first on simulated spectra and further on in vivo spectra from healthy volunteers.

2.2. Methods and Materials

All MR data were collected on a 3.0 T MR system (Magnetom Trio; Siemens Medical Solutions, Erlangen, Germany) using a cylindrically polarized body coil for RF transmission and a 32-channel head coil for signal reception.

2.2.1. Acquisition of in vivo Spectra

The study was approved by the Institutional Review Board (IRB). Scout images were acquired in all three orthogonal directions using a T_2-weighted fast spin echo sequence (TR/TE = 2000/25 ms, echo train length = 8, field of view = 220 × 220 mm^2, matrix size = 256 × 256, number of slices = 15 (no gap), slice thickness = 8 mm). Based on the scout images, a voxel (2 × 2 × 2 cm^3) was placed on the left frontal lobe containing mainly white matter. Then, auto-shimming was performed over the voxel. MRS data were collected using a PRESS sequence [49, 50] (TR/TE = 2000 / 30 ms, spectral bandwidth = 2 kHz, number of data points = 2048). A water-unsuppressed spectrum was collected for the screening of the quality of the spectrum such as the measurement of the linewidth of water signal (number of signal averages (NSA) = 8). Then, a total of 8 water-suppressed spectra each with NSA = 8 were collected consecutively for each subject (a total of 40 spectra from 5 subjects). The 8 water-suppressed spectra were combined to generate another 8 spectra with NSA equivalent to 8, 16, 24, 32, 40, 48, 56, and 64 for each subject, which were used for the evaluation of the performance of the CNN against different SNR of the spectra. The acquired FID data were zero-filled to 4096, Fourier transformed, and zeroth order phase corrected by using jMRUI (v.5.0). [116] No first order phase correction

and line-broadening was applied. Then, the residual water signal at ~ 4.7 ppm was removed by using the HLSVD filter. [119] Finally, the processed spectra were cropped to 1024 data points corresponding to ~ 0.5-4.5 ppm range as with the simulated spectra.

2.2.2. Acquisition of Metabolite Phantom Spectra

For the simulation of brain spectra for training and testing of the CNN, data were collected from the aqueous solutions of metabolites in glass bottles (300 mL) for 17 metabolites individually (Ala, Asp, Cr, GABA, Glc, Gln, Glu, GSH, GPC, Lac, mI, NAA, NAAG, PCr, PCh, PE, Tau; NAAG at 10 mM, PC at 36.05 mM, and the rest at 50 mM in a mixture with TSP (1 mM) and sodium azide (NaN_3; 0.1%); pH = 7.0 − 7.5; Sigma-Aldrich). Each phantom was placed at the isocenter of the magnet. After defining a voxel also at the isocenter, auto-shimming was performed followed by manual refinement. Spectra were acquired using the same sequence and sequence parameters that were used for the human subjects except for a larger NSA of 256.

2.2.3. Simulation of Brain Spectra

In the simulation of brain spectra the metabolite phantom spectra were used as a spectral basis set and the true in vivo brain spectra were used for the extraction of a priori knowledge such as SNR and linewidth of in vivo spectra. Simulation was performed by using in-house scripts written in Python (v.3.5.2; Python Software Foundation).

First, the upper and lower bounds of the metabolite concentrations in normal human brain were determined according to the literature [34, 58, 130, 131] (Table 2-1). These concentration ranges were evenly divided according to the total number of simulated spectra (N = 50,000) for all metabolites. Then, metabolite spectra were generated by combining all metabolite phantom spectra according to randomly selected relative metabolite concentration ratios within the concentration bounds. The GT target spectra (intact metabolite spectra) were prepared by using these metabolite spectra without spectral baseline and artificial addition of noise, line-broadening and frequency/phase shift (Figure 2-1).

Second, the spectral baseline in the range of ~ 0.5 – 4.5 ppm was simulated

using 17 Gaussian model functions based on the previous studies. [132, 133] These model functions were combined into 9 groups (M1-M9) and their relative amplitudes were normalized according to the literature (Table 2-2). [132, 133]Then, spectral baselines were generated by randomly varying the linewidth of the 17 model functions and the relative amplitude of the M1-M9 groups within ± 20% and ± 10% from their initial values, respectively. Third, the relative proportions of the metabolite signal and baseline signal in typical in vivo brain spectra were investigated from the true in vivo spectra prior to combining these two signal groups. Based on a previous study [134] 43 initial data points (empirically determined) were truncated from the FID of the in vivo data (NSA = 64, N = 5) and their Fourier transforms were considered as a surrogate of baseline-nulled metabolite-only spectra. Then, by comparing the spectra with and without the FID truncation a metabolite-to-baseline signal ratio of 1 : 0.876 was determined. The metabolite signal and baseline signal were combined by randomly varying the ratio within ± 25% (Figure 2-1). Forth, Lorentzian line broadening was applied such that the linewidth of the total NAA (tNAA = NAA + NAAG; ~ 2.0 ppm) was broadened from ~ 5 Hz as measured in phantom to 10 – 20 Hz. A zeroth order phase shift was also applied in the range of ± 5°.

Fifth, after Fourier transformation all simulated spectra were cropped to 1024 data points (~ 0.5-4.5 ppm) for efficient CNN training. Then, the SNR of the spectra were adjusted by adding random noise such that the mean SNR of the tNAA (~ 2.0 ppm), total Cr (tCr = Cr + PCr; ~ 3.0 ppm), and total Cho (tCho = PCh + GPC; ~ 3.2 ppm) is in the range of ~ 5-15. The spectral noise was estimated by measuring the standard deviation in the 8.0-10.0 ppm range prior to the cropping of the spectra to 1024 data points. A frequency shift was also applied to the spectra at this stage in the range of ± 10 Hz.

Finally, a total of 50,000 brain spectra were simulated and randomly assigned into a training (N = 40,000), a validation (N = 5,000) and a test (N = 5,000) sets (these data sets are referred to as the original data sets). The range of the spectral amplitudes was -248.5 – 14125.0 (2150.9 ± 1045.7) for the whole simulated brain spectra and 6.2 – 10875.0 (746.2 ± 1049.0) for the corresponding whole GT target spectra.

Metabolite	Concentration	
	Lower limit (mmol L^{-1})	Upper limit (mmol L^{-1})
Ala	0.1	1.5
Asp	1.0	2.0
Cr	4.5	10.5
GABA	1.0	2.0
Glc	1.0	2.0
Gln	3.0	6.0
Glu	6.0	12.5
GPC	0.5	2.0
GSH	1.5	3.0
Lac	0.2	1.0
mI	4.0	9.0
NAA	7.5	17
NAAG	0.5	2.5
PC	0.5	2.0
PCr	3.0	5.5
PE	1.0	2.0
Tau	2.0	6.0
Glx	9.0	18.5
tCho	1.0	4.0
tCr	7.5	16.0
tNAA	8.0	19.5

Table 2-1. Ranges of the relative metabolite concentrations used in the simulated brain spectra.

MM group	Line function[a]	Initial amplitude[b]	Chemical shift (ppm)[c]	Initial linewidth (Hz)[d]
M1	M1-09	0.72	0.90	21
M2	M2-12	0.28	1.20	19
M3	M3-14	0.38	1.36	16
M4	M4-16	0.05	1.63	8
	M4-17		1.68	13
	M4-18		1.81	13
M5	M5-20	0.78	2.02	29
	M5-21		2.08	21
	M5-23		2.25	18
	M5-26		2.61	5
M6	M6-30	0.3	2.97	14
M7	M7-31	0.11	3.11	18
	M7-32		3.22	10
	M7-33		3.27	10
M8	M8-37	0.71	3.67	34
	M8-38		3.80	12
M9	M9-40	1	3.96	37

Table 2-2. Spectral characteristics of the Gaussian linefunctions and macromolecule (MM) groups used in the simulated spectral baseline.

[a] Separate line functions were used only when the chemical shift difference between the line functions reported in the literature is larger than 0.05 ppm.

[b] The initial relative amplitudes of the MM groups were normalized to the amplitude of M9. These initial values were randomly varied in the range of ± 10%.

[c] For the line functions reported with different chemical shifts in the literature, their mean value was used.

[d] In the simulation of spectral baseline, these initial linewidths were randomly varied in the range of ± 20%.

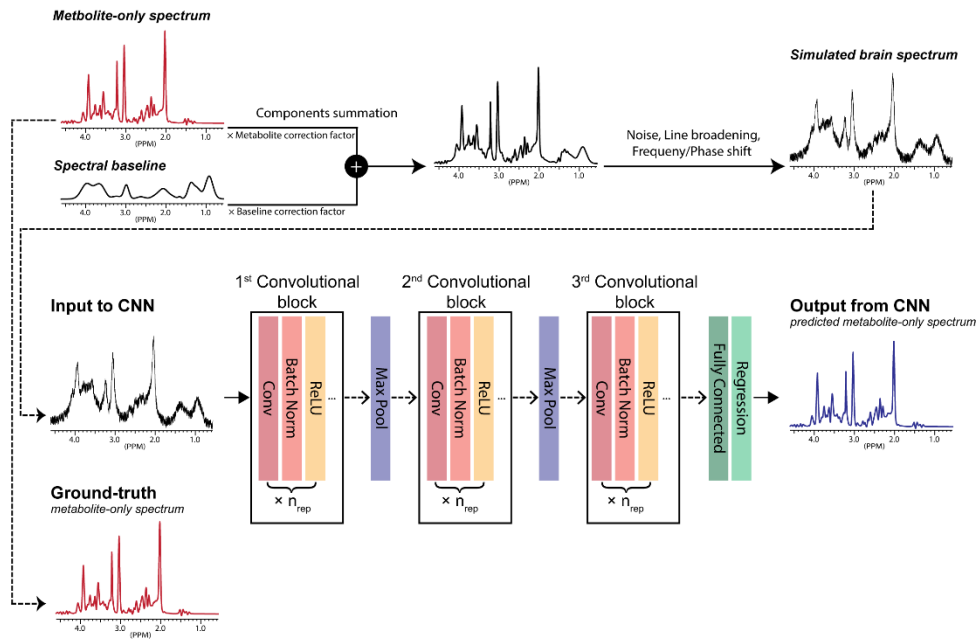


Figure 2-1. A schematic of the simulation of brain spectra and the training of the convolutional neural network (CNN). The metabolite-only spectra simulated by linearly combining the metabolite phantom spectra according to randomly selected relative concentration ratios are used as the ground truth in the training of the CNN. To mimic in vivo brain spectra the metabolite-only spectra are combined with noise, line broadening, frequency/phase shift, and spectral baseline. These simulated brain spectra are used as the input to the CNN in the training. The CNN is trained to map the brain spectra degraded by low SNR, broadened linewidth, frequency/phase shift, and unknown spectral baseline into the noise-free, line-narrowed, frequency/phase shift-corrected, baseline-removed, intact metabolite spectra. (Conv: convolution, Batch Norm: batch normalization, ReLU: rectified linear unit, Max Pool: max pooling)

2.2.4. Design and Optimization of CNN

CNNs were designed and optimized by using MATLAB neural network toolbox (v.9.4 (R2018a); Mathworks Inc., Natick, MA, USA) on 4 GPUs (NVIDIA Titan Xp). Initially, a CNN was designed that consisted of 1 input layer, 3 convolution blocks, 2 max pooling layers, 1 fully connected layer, and 1 regression layer. Each convolution block was composed of n-repetitions (n_{rep}) of one-dimensional convolution-batch normalization-activation layers (Figure 2-1). The simulated spectra in the training set (height × width × channels × array = 1024 × 1 × 1 × 40000) were zero-center normalized.

Using the original training and validation data sets, the n_{rep}, initial learning rate, momentum of the stochastic gradient descent with momentum algorithm (SGDM), and L_2 regularization parameter were Bayesian-optimized [135] for each of the convolution filter sizes of 3, 5, 7, 9, 11, 15, and 25. The range of n_{rep} in the optimization was 1 – 4. The ranges of the initial learning rate, momentum of SGDM, and L_2 regularization were 1×10^{-10} – 1×10^{-4}, 0.80 – 0.95, and 1×10^{-10} – 1×10^{-2}, respectively. The number of convolution filters was set to the integer closest to $64/\sqrt{n_{rep}}$ at the first convolution block and doubled at each of the following blocks. The convolutions were performed with stride = 1 and pad size = (filter size - 1)/2 to match the input and output sizes of the convolution layers. A ReLU activation function was used in the activation layer. For the training of the network an SGDM was used. The loss function was MSE. Other parameters used in the Bayesian optimization were: pool size and stride of max pooling = 2 × 1 and 1, batch size = 32, 3 validations per epoch, and maximum epoch = 100. The learning rate was scheduled to drop at the 91st epoch by a factor of 0.1. The maximum number of objective function evaluations and maximum total optimization time for each filter size were 30 and 12 hours, respectively.

2.2.5. Evaluation of the Reproducibility of the Optimized CNN

To evaluate the reproducibility of the optimized CNN parameters, a 10-fold cross validation was performed. Specifically, the 45,000 simulated spectra in the original training and validation sets were randomly split into 9 equally sized groups (thus, a total of 10 folds including the original test set). Each group was assigned to a test set

(N = 5,000), while those spectra in the remaining 9 groups were randomly divided into a training set (N = 40,000) and a validation set (N = 5,000). The Bayesian-optimized CNN was trained on each of these training sets (the corresponding validation sets being also used during training), and the resulting mean MSE between the GT and the CNN-predicted intact metabolite spectra for the test sets were examined.

2.2.6. Metabolite Quantification from the Predicted Spectra

Given that the simulated target spectra of the CNN were generated by a linear combination of the metabolite phantom spectra only, quantification of individual metabolites from the CNN-predicted intact metabolite spectra amounts to solving a simple inverse problem. That is,

$$\mathbf{C} = \mathbf{S} \times \text{pinv}(\mathbf{b}). \quad (2.1)$$

where, for N-predicted spectra to be analyzed for metabolite quantification, **C** is an N-by-17 matrix containing the relative concentrations of the 17 metabolites, **S** is an N-by-1024 matrix containing the N-predicted spectra with 1024 data points, pinv denotes the Moore-Penrose pseudoinverse of a matrix, and **b** is a 17-by-1024 matrix containing the basis spectra of the 17 metabolites with 1024 data points.

2.2.7. Evaluation of CNN in Metabolite Quantification

For the simulated spectra in the test set, the relative metabolite concentrations estimated by the proposed method were directly compared with the GT values, and the resulting mean absolute percent errors (MAPE) were calculated.

For in vivo spectra, metabolite concentrations were estimated by using the proposed method from the spectra with different NSA of 8, 16, 24, 32, 40, 48, 56, and 64. The results with NSA = 64 were compared with those obtained by using the LCModel analysis (v.6.3-1J; [55]) that is most widely used for metabolite quantification in ^{1}H-MRS. The vendor-provided spectral basis set was used in the LCModel analysis. The results were also compared with those reported in the literature. [64, 136-141] Specifically, the references used for obtaining the expected concentration ranges of the metabolites were: Glx (= Gln + Glu), mI, tCho, tCr [141], GABA, GSH [137], Gln/Glu [139], PC/GPC [140], Cr/PCr [64, 138] and NAAG/NAA. [136] The expected concentration ranges of the metabolites with

respect to tNAA were calculated by mean ± (2 × SD) where SD denotes standard deviation and was estimated by assuming that the covariance between the concentrations of tNAA and the rest of the metabolites was zero. To evaluate the robustness of the proposed method against low SNR, the metabolite concentrations obtained with NSA = 8, 16, 24, 32, 40, 48 and 56 were compared with those obtained with NSA = 64. The same analysis was also performed with the LCModel analysis. In the case of the LCModel analysis the resulting CRLB values were also shown.

For in vivo spectra, metabolite concentrations were estimated using the proposed method from spectra with NSAs of 8, 16, 24, 32, 40, 48, 56 and 64. The results with NSA = 64 were compared with those obtained by using the LCModel analysis (v.6.3-1J; [55]). The results were also compared with those reported in the literature [64, 136-141]. To evaluate the robustness of the proposed method to low SNR, the metabolite concentrations obtained with NSA = 8, 16, 24, 32, 40, 48 and 56 were compared with those obtained with NSA = 64. For LCModel analysis, the resulting CRLB values were also shown.

2.2.8. Statistical Analysis

All statistical analyses were performed by using PRISM (v.6.01; GraphPad Software, Inc., La Jolla, CA, USA). For pair-wise group comparison a Student t-test was performed. A P-value of less than 0.05 was considered statistically significant.

2.3. Results

2.3.1. SNR Distribution of the Simulated Spectra

The SNR distribution of the Glx (~ 2.1-2.4 ppm), mI (~ 3.5-3.6 ppm), tCho (~ 3.2 ppm), tCr (~ 3.0 ppm) and tNAA (~ 2.0 ppm) for the simulated spectra in the test set are shown in Figure 2-2, for which the SNR can be measured directly from the spectra. The wide range of the relative metabolite concentrations in the simulated spectra is clearly demonstrated.

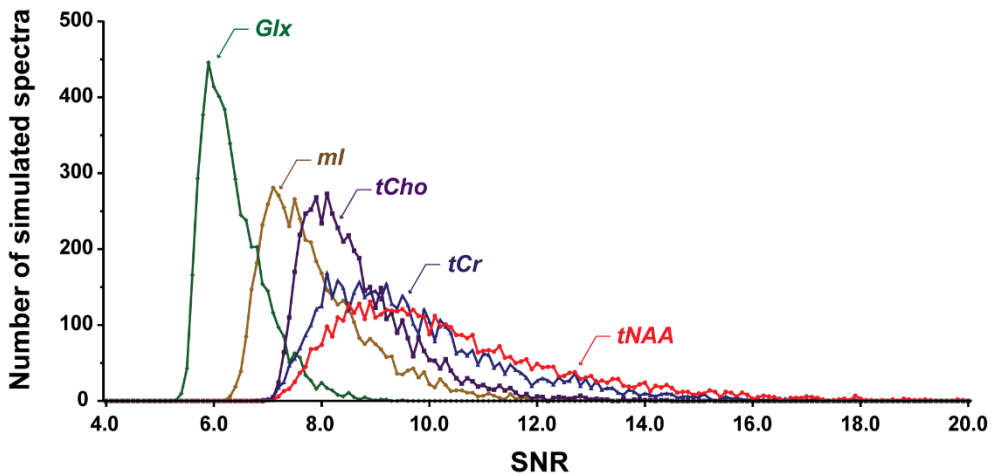


Figure 2-2. The SNR distribution of the Glx, mI, tCho, tCr, and tNAA for the simulated spectra in the test set (N = 5,000). For these metabolites the SNR can be measured directly from the spectra. The SNR ranges were 5.28 – 9.48 (6.37 ± 0.59) for Glx, 6.23-13.32 (7.89 ± 1.00) for mI, 7.05 – 15.19 (8.67 ± 1.04) for tCho, 6.84 – 18.26 (9.71 ± 1.72) for tCr, and 6.90 – 20.74 (10.41 ± 2.14) for tNAA. The signal was estimated by measuring the maximum amplitude in ~ 2.1 – 2.4 ppm, ~ 3.5 – 3.6 ppm, ~ 3.2 ppm, ~ 3.0 ppm, and ~ 2.0 ppm for Glx, mI, tCho, tCr, and tNAA, respectively, and the noise was estimated by measuring the standard deviation in 8.0 – 10.0 ppm before the cropping of the spectra down to 1024 data points.

2.3.2. Optimized CNN

The Bayesian-optimized CNN parameters were: convolution filter size = 15×1, number of convolution filters = 32, 64, and 128 for the 1st, 2nd, and 3rd convolution blocks, n_{rep} = 4 (thus, a total of 41 layers with 12 convolution layers), initial learning rate = 3.0709×10^{-8}, SGDM momentum = 0.8024, and L_2 regularization = 1.0291×10^{-4}.

In the 10-fold cross validation using the optimized CNN parameters, the mean MSE between the GT and the CNN-predicted spectra ranged 4904.5 – 5050.0 (4970.2 ± 45.2). Defining the maximum MSE as the mean squared value of the GT target spectrum (i.e., in the case where the predicted spectrum has zero amplitude across the whole spectral range (null spectrum)), and %MSE as the MSE normalized to the maximum MSE, the %MSE corresponding to the mean MSE of 4970.2 ± 45.2 is estimated to be 0.3 ± 0.0 %. Using the original train/validation/test sets that were used for the Bayesian optimization of the CNN, the mean MSE was 4945.3 (%MSE = 0.3 ± 0.0 %) (the 4th lowest among the 10 CNNs in the cross validation). All results reported from here were obtained with the CNN that was trained by the original data set.

2.3.3. Representative Simulated and CNN-predicted Spectra

The representative simulated spectra in the test set with different SNR, linewidth, and relative signal intensity of the peaks are shown in Figure 2-3(A)-(D) along with the GT (Figure 2-3(E)-(H)) and the CNN-predicted intact metabolite spectra (Figure 2-3(I)-(L)). A high tCho (~ 3.2 ppm) and a low tNAA (~ 2.0 ppm) signals are observed in Figure 2-3(A) and Figur 2-3(D) with respect to their well-known relative signal intensity relationship (tNAA > tCr (~ 3.0 ppm) > tCho) in normal human brain. Despite the severe degradation of the simulated input spectra due mainly to poor SNR and linewidth, the CNN effectively recovered the spectral details in 3.2 – 3.8 and 2.1 – 2.9 ppm ranges (Figure 2-3(I)-(L)). The line-narrowing capability of the CNN is also clearly shown. That is, the signal amplitude of the major peaks such as tCho (~ 3.2 ppm), tCr (~ 3.0 ppm), and tNAA (~ 2.0 ppm) appear larger in the CNN-predicted spectra (Figure 2-3(I)-(L)) than in the simulated brain spectra (Figure 2-3(A)-(D)) resulting from the effective line-narrowing by the CNN. The linewidth (LW) of the tNAA (~ 2.0 ppm) ranging 10.4 – 16.2 Hz in the simulated brain spectra

(Figure 2-3(A)-(D)) are all narrowed into ~ 5 Hz in the CNN-predicted spectra (Figure 2-3(I)-(L)), which corresponds to the linewidth of NAA in phantom. The spectra reconstructed by linearly combining the metabolite phantom spectra according to the estimated relative metabolite concentrations are also shown (Figure 2-3(M)-(P)). The individual spectral components of the reconstructed spectra are shown more clearly in Figure 2-4. The difference spectra between the GT and the CNN-predicted spectra show only a small amount of residual signal mainly in the major peak areas such as tCho, tCr, and tNAA (Figure 2-3(Q)-(T)). The difference spectra between the reconstructed spectra and the predicted spectra show no visible residual signal (Figure 2-3(U)-(X)), which demonstrates that the CNN fulfills its tasks of removing noise, line broadening, frequency/phase shift, and spectral baseline as trained such that the predicted spectra can be completely accounted for by a linear combination of the metabolite phantom spectra only.

2.3.4. Metabolite Quantification in Simulated Spectra

The results of the metabolite quantification from all simulated spectra in the test set (N = 5,000) by using the proposed method are shown in Figure 2-5. For all 17 metabolites MAPE was 20.67 ± 16.71%. Excluding Ala, GPC, Lac, NAAG, and PC (MAPE > 20%), MAPE was 12.49 ± 4.35%. The MAPE were less than 10% for Cr (8.65 ± 7.16%), Glu (8.16 ± 6.85%), mI (7.61 ± 6.77%), NAA (4.64 ± 4.02%), Glx (5.36 ± 4.48%), tCho (7.87 ± 7.35%), tCr (3.66 ± 3.05%), and tNAA (4.00 ± 3.38%). GABA and GSH were quantified with MAPE of 17.5 ± 11.76% and 12.72 ± 10.04%, respectively.

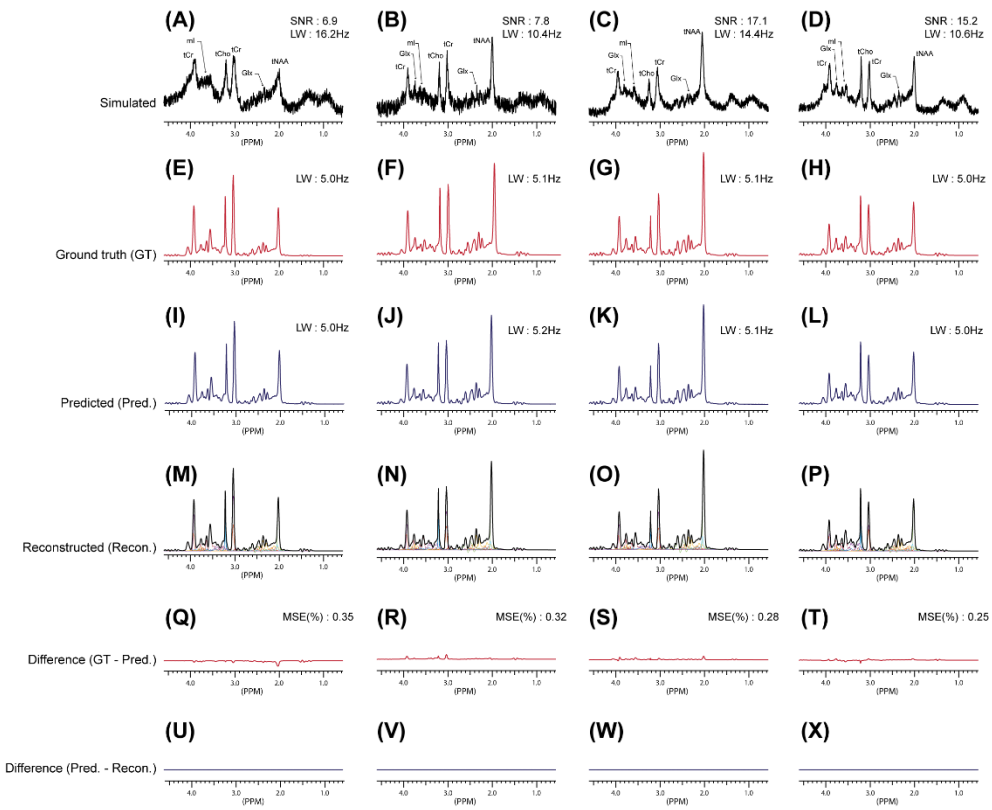


Figure 2-3. Representative simulated brain spectra in the test set and corresponding CNN-predicted intact metabolite spectra. (A)-(D): simulated brain spectra with noise, line broadening, phase and frequency shift, and spectra baseline, which were used as the input to the CNN. (E)-(H): ground truth (GT) spectra that were composed solely of the metabolite phantom spectra. (I)-(L): CNN-predicted intact metabolite spectra (Pred.). (M)-(P): reconstructed spectra (Recon.) by linearly combining the metabolite phantom spectra according to the relative metabolite concentrations estimated by solving the inverse problem using the CNN-predicted spectra. The metabolite phantom spectra used for the reconstruction are also shown. (Q)-(T): difference spectra obtained by subtracting the CNN-predicted spectra from the ground truth spectra (GT – Pred.). (U)-(X): difference spectra obtained by subtracting the reconstructed spectra from the CNN-predicted spectra (Pred. – Recon.). %MSE = {mean square of (GT – Pred.)}/{mean square of (GT)}×100.

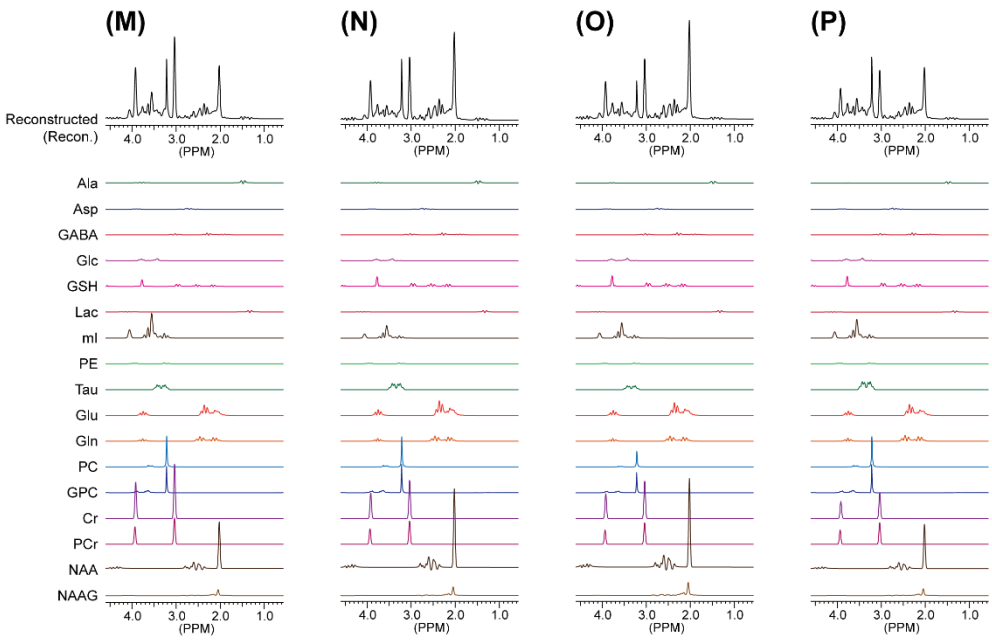


Figure 2-4. The individual metabolite spectra consisting of the reconstructed spectra shown in Figure 2-3(M)-(P). They were obtained by solving the inverse problem for the CNN-predicted spectra.

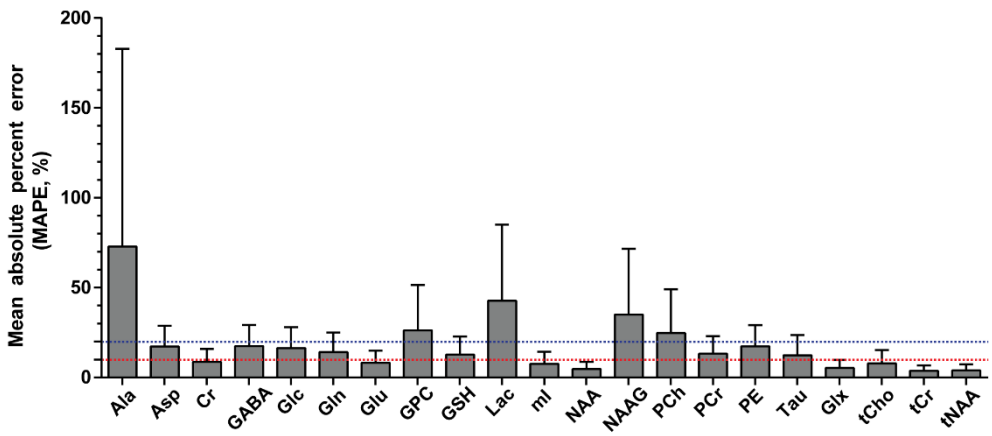


Figure 2-5. Mean absolute percent error (MAPE) of the concentrations of the metabolites estimated over all simulated brain spectra in the test set (N = 5,000).

2.3.5. Representative in vivo and CNN-predicted Spectra

The in vivo spectra with NSA equivalent to 64 and 8 are shown for all five human subjects in Figure 2-6(A)-(E) and Figure 2-7(A)-(E), respectively, along with the corresponding CNN-predicted intact metabolite spectra (Figure 2-6(F)-(J) and Figure 2-7(F)-(J)). The SNR and linewidth of the spectra with NSA = 64 measured from tNAA ranged 14.9-17.8 and 10.4-12.4 Hz, respectively. As in Figure 2-3, the spectral details that were obscured by the noise and line broadening in the in vivo spectra were effectively recovered in the CNN-predicted spectra. The line-narrowing capability of the CNN is also clearly shown. The difference spectra between the CNN-predicted spectra and the reconstructed spectra by linear combination of the metabolite phantom spectra ('Difference (Pred – Recon)' in Figure 2-6(P)-(T) and Figure 2-7(P)-(T)) show no visible residual signal and thus demonstrate complete removal of spectral baseline. The difference spectra between the CNN-predicted spectra in Figure 2-7(U)-(Y) demonstrate the robust performance of the CNN against low SNR.

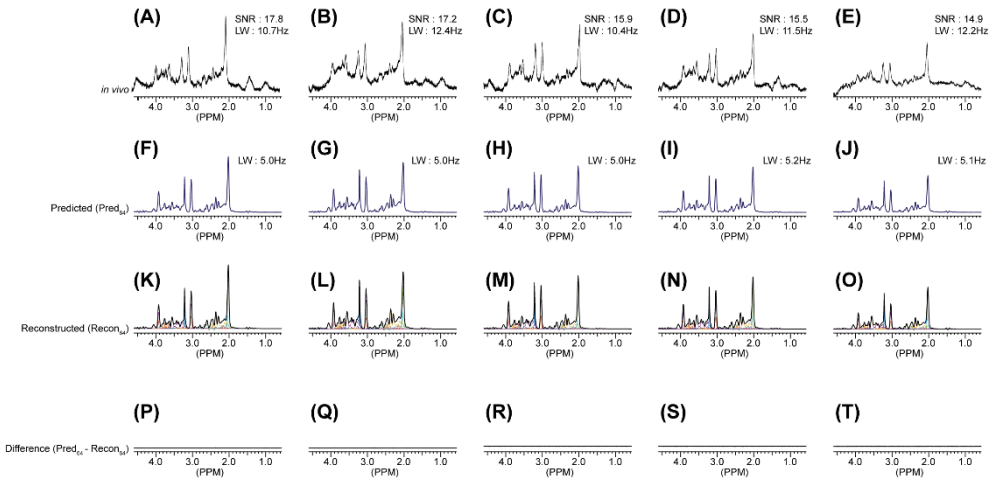


Figure 2-6. In vivo brain spectra with NSA = 64 ((A)-(E)) from the five healthy subjects and corresponding CNN-predicted intact metabolite spectra ((F)-(J)). The difference spectra ($Pred_{64}$ – $Recon_{64}$) ((P)-(T)) between the CNN-predicted intact metabolite spectra (Pred) and the reconstructed spectra (Recon) ((K)-(O)) by linearly combining the metabolite phantom spectra demonstrate that the CNN successfully mapped the degraded in vivo spectra into the noise-free, line-narrowed, baseline-removed, intact metabolite spectra that can be completely accounted for by a linear combination of the metabolite phantom spectra only. %MSE = {mean square of ($Pred_{64}$ – $Pred_8$)}/{mean square of ($Pred_{64}$)}×100.

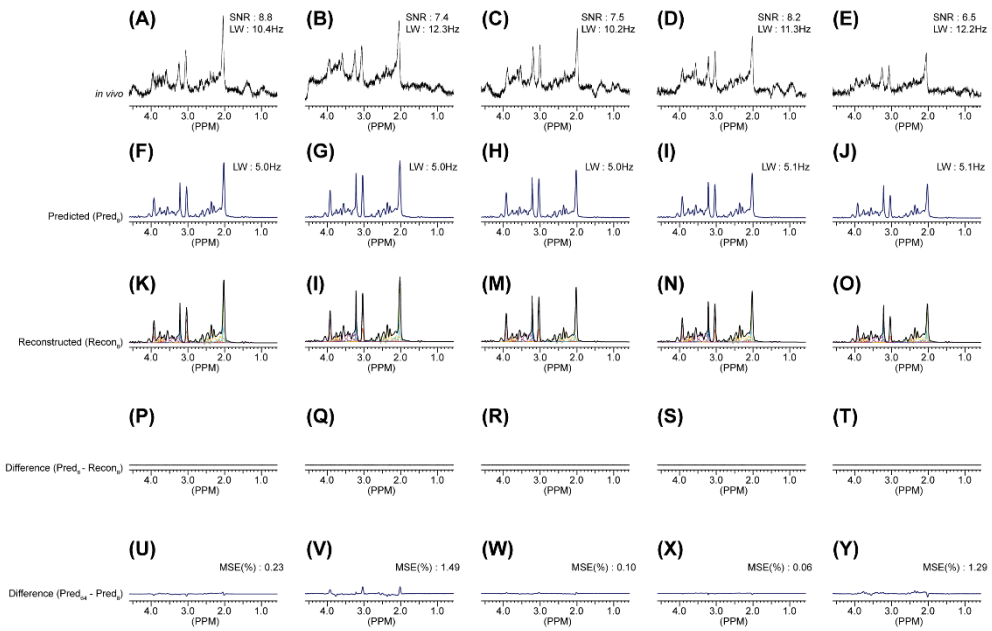


Figure 2-7. In vivo brain spectra with NSA = 8 ((A)-(E)) from the five healthy subjects and corresponding CNN-predicted intact metabolite spectra ((F)-(J)). The difference spectra ($Pred_8$ – $Recon_8$) ((P)-(T)) between the CNN-predicted intact metabolite spectra (Pred) and the reconstructed spectra (Recon) ((K)-(O)) by linearly combining the metabolite phantom spectra demonstrate that the CNN successfully mapped the degraded in vivo spectra into the noise-free, line-narrowed, baseline-removed, intact metabolite spectra that can be completely accounted for by a linear combination of the metabolite phantom spectra only. The difference spectra ($Pred_{64}$ – $Pred_8$) ((U)-(Y)) between the CNN-predicted metabolite spectra for the in vivo spectra with NSA = 64 and 8 (i.e., the difference spectra between (F)-(J) in Figure 2-6 and (F)-(J) in this Figured 2-7) demonstrate the robust performance of the CNN against low SNR. %MSE = {mean square of (Pred64 – Pred8)}/{mean square of (Pred64)}×100.

2.3.6. Metabolite Quantification in in vivo Spectra

The results of metabolite quantification from the in vivo spectra with NSA = 64 by using the proposed method and the LCModel analysis are compared in Figure 2-8 for those metabolites that are most commonly reported in the literature (Figure 2-8 (A)-(D)), challenging to detect (Figure 2-8(E)-(F)), and difficult to separate from each other (Figure 2-8(G)-(J)). Overall, the metabolite concentrations estimated from the five human subjects are within or close to the expected ranges (mean ± 2 × SD) from the literature except mI, GSH and especially Cr/PCr for the LCModel analysis and PCh/GPC for the proposed method. In the LCModel analysis the PCh/GPC values were not available (Figure 2-8(H)) because PCh were not detectable (amplitude = 0 and CRLB = 999%) for all five subjects.

For the two approaches the variations in metabolite concentration estimated across the in vivo spectra with NSA = 8, 16, 24, 32, 40, 48, and 56 with respect to the concentrations estimated with NSA = 64 are compared in Figure 2-9 for those metabolites included in Figure 2-8. The variations in metabolite concentration and the CRLB values are the mean values over the five human subjects except for Cr with NSA = 8 in the LCModel analysis where amplitude = 0 and CRLB = 999% ('not detectable') were returned for one subject. Therefore, this case was excluded in the calculation of the mean values. For the LCModel analysis, such a case of 'not detectable' occurred with PCh for all subjects across all NSA, and therefore, the results of PCh were excluded in Figure 2-9(J). For the proposed method the variations were ~ 10% or less for the majority of the metabolites except for GSH. For the LCModel analysis the estimated concentrations appear more variable, particularly for Cr and GABA.

The whole run time for a pre-processed single in vivo spectrum was estimated to be 4.13 seconds on a PC using CPU only (Intel® Core™ i5-3570 processor (3.40 GHz); 16 GB RAM), which included loading of the files (input spectrum, metabolite basis set, and the trained CNN), prediction of the intact metabolite spectrum from the input spectrum by the CNN, and quantification of the individual metabolites from the intact metabolite spectrum by solving the inverse problem.

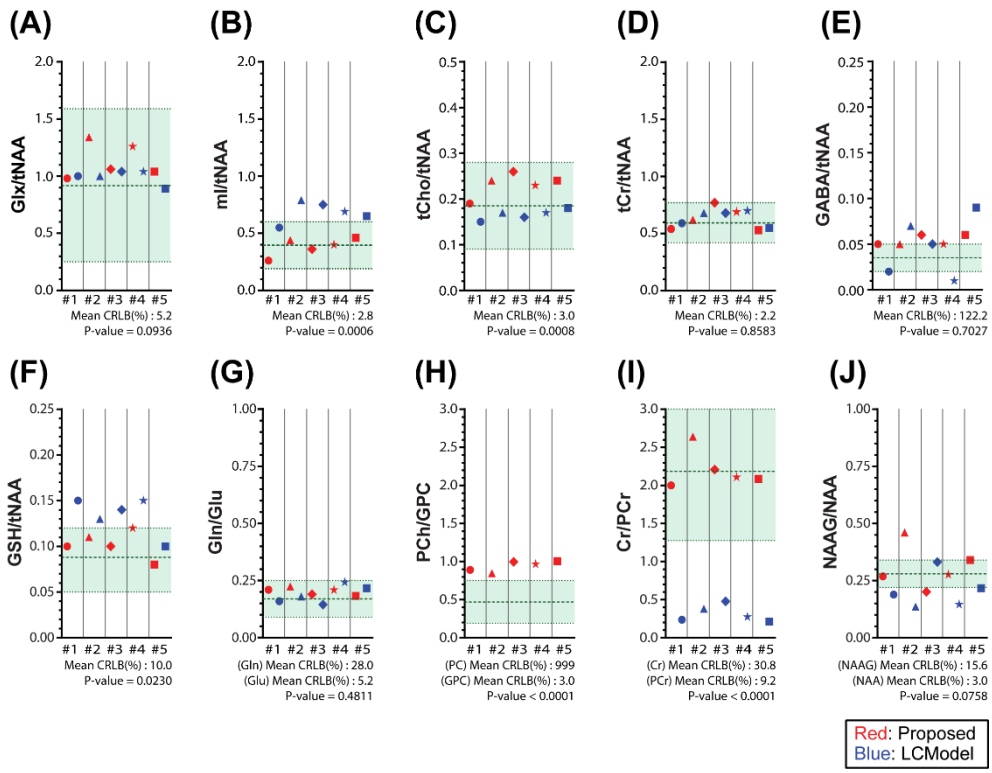


Figure 2-8. Comparison of the relative metabolite concentrations estimated by the CNN (red symbols; left) and the LCModel analysis (blue symbols; right) from the in vivo brain spectra (NSA = 64) of the five healthy subjects (#1 - #5). Different symbols were used for different subjects. The expected concentration ranges of the metabolites from the literature were marked with dotted lines (the mean values marked with thick-dotted lines).

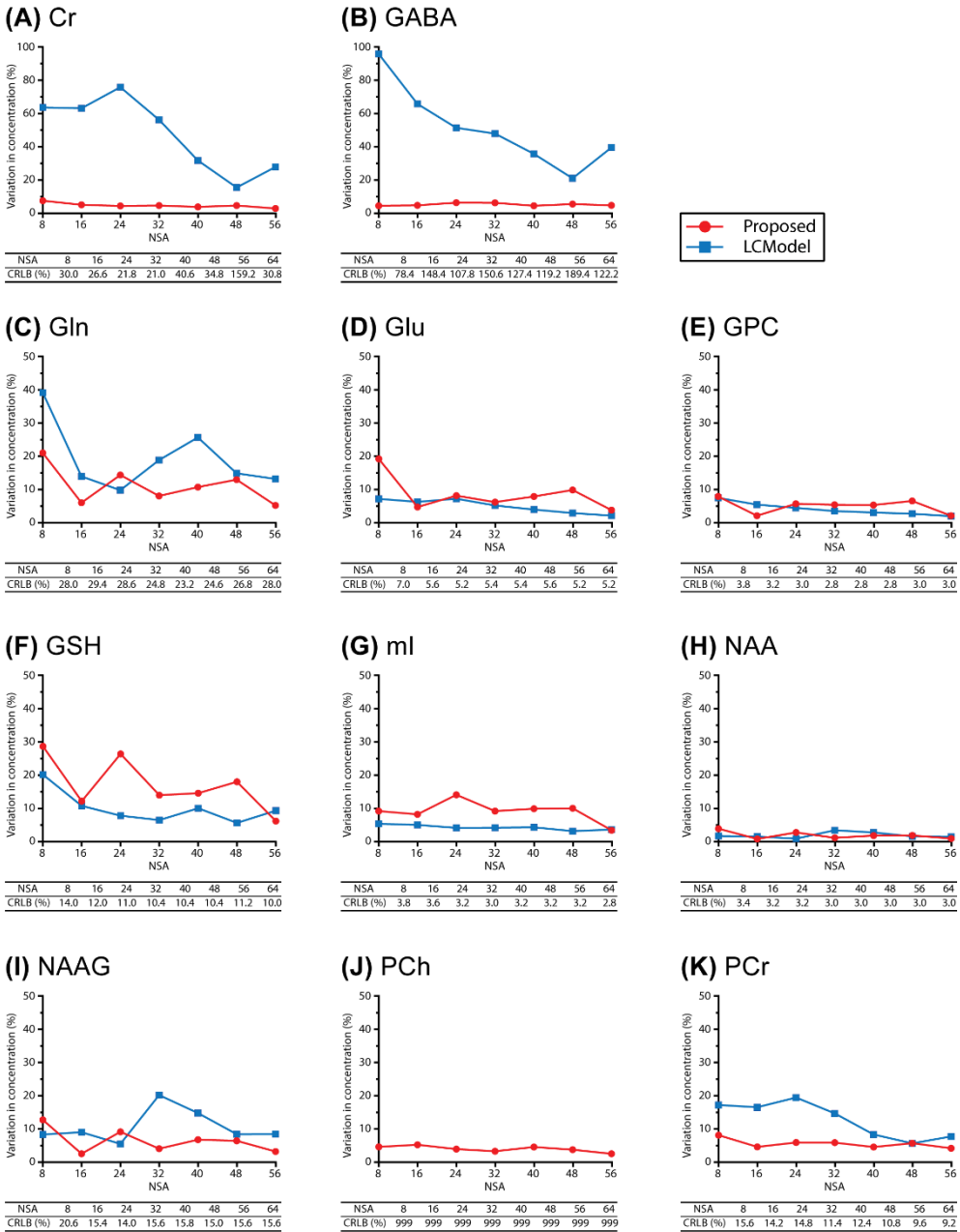


Figure 2-9. The variations (%) in the metabolite concentrations estimated by the proposed method (red lines and circles) and the LCModel analysis (blue lines and squares) across the in vivo spectra with NSA = 8, 16, 24, 32, 40, 48, and 56 with respect to the concentrations estimated from the spectra with the maximum NSA of 64.

2.4. Discussions

2.4.1. Motivation of Study

In a typical clinical setting, brain metabolite spectra are indispensably degraded due to many factors including, but not limited to, low SNR, overlap between metabolite signals, line broadening, and unknown spectral baseline. For the quantification of individual metabolites from the degraded brain spectra, a nonlinear least squares fitting has been the most widely used approach. [55, 116] However, even with the current state-of-the-art software package, it can be challenging. [126] Spectral editing methods [31] such as a multiple quantum filter and a J-difference editing can effectively isolate a target metabolite signal from its neighboring metabolite signals and spectral baseline, but is not suitable for profiling of all MR visible metabolites. Therefore, development of a more robust method for brain metabolite quantification is a remaining issue in ^{1}H-MRS.

To this end, several pioneering studies reported potential applicability of deep learning in ^{1}H-MRS in line with its rapidly expanding applications in a variety of different tasks. For instance, Kyathanahally et al [129] reported the application of CNNs in the detection and removal of spectroscopic artifacts following their earlier work using machine learning. [142] Gurbani et al [128] reported a fully automated management of spectral quality using a CNN. The potential applications of machine learning for metabolite quantification have also been explored by a few research groups. Using a random forest regression Das et al [107] reported the utility of machine learning for the quantification of major metabolite peaks such as Cho, Cr, Glx, mI and NAA in the simulated and in vivo ^{1}H-MRS spectra. Using deep learning Hatami et al [110] reported a CNN that was trained on the time-domain data set and capable of quantifying individual metabolites in simulated spectra. Iqbal et al [111] also reported the potential application of a CNN for metabolite quantification using two dimensional simulated spectra. To the best of our knowledge, however, none of these previous studies evaluated the performance of their machines on in vivo spectra of human brain for all representative MR visible brain metabolites including Glu separately from Gln and those metabolites suffering from low concentrations and severe spectral overlap such as GABA [143] and GSH. [118]

Inspired by these previous studies and motivated by the remaining issue, we

developed a CNN that was trained to map in vivo brain spectra degraded by poor SNR and linewidth, frequency and phase shift, and unknown baseline into noise-free, line-narrowed, frequency and phase shift-corrected, baseline-removed, intact metabolite spectra through a sequence to sequence regression. As the target spectra used in the training of the CNN were generated solely by a linear combination of the metabolite phantom spectra with different relative concentration ratios, the quantification of individual metabolites from the CNN-predicted spectra is nothing but a simple inverse problem involving only one matrix inversion and one matrix multiplication. The difference spectra (Figure 2-3 (Q)-(T) and (U)-(X); Figure 2-6(P)-(T); Figure 2-7(P)-(T)) clearly demonstrate that the CNN fulfills its tasks of removing noise, line broadening, phase and frequency shift, and spectral baseline as trained such that the CNN-predicted spectra can be completely accounted for by a linear combination of the metabolite phantom spectra only.

2.4.2. Metabolite Quantification on Simulated and in vivo Brain Spectra

In this study, Ala, GPC, Lac, NAAG, and PC were quantified with MAPE > 20% for the simulated brain spectra in the test set. Ala and Lac are known to be present in mammalian brain at a concentration of < 1 mM [58], and could be detected only upon increased concentrations under pathological conditions. [144] In this study, they were the two lowest in concentration among the metabolites included in the simulated spectra (Table 2-1). Thus, the low concentrations in combination with the overlapping MMs signal (M3 in Table 2-2) [132, 133] may explain their relatively high MAPE. The quantification of NAAG separately from NAA is also known to be difficult due to low concentrations and spectral overlap. [144, 145] In this study the difficulty should be exacerbated by the wide range of line broadening applied to the simulated spectra, resulting in the relatively high MAPE. As shown in Figure 2-6, whereas the linewidth of the in vivo spectra with NSA = 64 ranged 10.4 – 12.4 Hz, the linewidth of the simulated spectra ranged 10 – 20 Hz. Likewise, the large line broadening in our simulated spectra most likely influenced the separate quantification of GPC and PCh, given the spectral overlap from mI, Glc, and Tau [58] in addition to between the two metabolites. It should be noted, however, that these results were obtained across all simulated spectra in the test set (N = 5,000), a portion of which had metabolite SNR substantially lower than that found in typical

human brain spectra in addition to the wide range of the line broadening. For instance, whereas the typical SNR range of tNAA in vivo is reported to be 23.5 ± 11.3 [19], it was 10.41 ± 2.14 (6.90 – 20.74) in this study as shown in Figure 2-2. Excluding these five metabolites, an overall MAPE of 12.49±4.35% was achieved for the rest 12 metabolites.

Despite the poorer SNR and linewidth ranges in the simulated spectra than in typical in vivo spectra our CNN achieved MAPE close to 10% or less for mI, NAA, and Tau, and even for Cr, Gln, Glu, GSH, and PCr that are already difficult enough to quantify in the short TE spectra with typical SNR and linewidth at clinical field strength. Specifically, due to the spectral proximity the separate quantification of Cr and PCr is known to be possible only at 7.0 T or higher. [58] The differentiation between Gln and Glu has long been an issue for MR spectroscopists. [146] GSH as well as GABA can be detected only upon the use of spectral editing methods. [58, 144] Nonetheless, our CNN achieved MAPE close to 10% for GSH and less than 20% for GABA in the simulated spectra. For the true in vivo spectra the concentrations of the metabolites of interest estimated by using the proposed method and the LCModel analysis were comparable to the expected concentration ranges except PCh/GPC for the proposed method and mI, GSH and Cr/PCr for the LCModel analysis. For the proposed method the deviation of PC/GPC from the expected range is due to the difficulty of quantifying both metabolites (MAPE > 20% in the simulated spectra). The difficulty of quantifying PC was also encountered with the LCModel analysis where it was not detectable for all subjects across all NSA. Given the potential dependence of metabolite concentrations on the brain regions and relative proportion of the gray/white matters [147], the deviation of mI as well as GSH from the expected concentration ranges with the LCModel analysis may need to be carefully interpreted. For Cr/PCr, however, the estimated values with the LCModel analysis were substantially deviated from the expected range of ~ 1.3-3.0 [64, 138], which may be in support of the efficacy of the line-narrowing provided by the proposed method.

2.4.3. Metabolite Quantification Robustness against Low SNR

One of the potential advantages of our CNN is the robustness against low SNR. Despite the substantial difference in SNR between the in vivo spectra with

NSA equivalent to 8 and 64 (SNR 7.7 ± 0.9 vs. 16.2 ± 1.2), the resulting CNN-predicted intact metabolite spectra are in close agreement with each other as demonstrated in the difference spectra (Figure 2-7 (U)-(Y)). Further, despite the wide range of the SNR across the in vivo spectra with different NSA (8 through 64), the majority of the metabolites of interest showed only ~ 10% variations in concentration with respect to the concentrations estimated with the maximum NSA of 64 using the proposed method. On the other hand, larger variations were observed with the LCModel analysis, particularly for Cr and GABA. Given that the experimental settings associated with the SNR of ^{1}H-MRS brain spectra (such as number of receiver channels, coil design, voxel size, and NSA) can be quite diverse even for different studies in the same laboratory, the robust performance of the proposed method against different SNR can be advantageous. In addition, while far more improvement is required prior to the routine use of deep learning-based metabolite quantification, results in this study all together are in support of the feasibility of implementing a sub-minute ^{1}H-MRS of human brain (e.g., with an NSA of only 16) by exploiting the robustness of our CNN against low SNR, which then can further extend the applicability of ^{1}H-MRS in clinical studies.

2.4.4. Study Limitation

In the simulation of spectra that mimic in vivo spectra this study did not include several factors that may influence the accuracy of metabolite quantification such as residual water signal, other spectroscopic artifacts [129, 142], lipid, and first order phase distortion. For instance, incomplete water suppression can result in substantial baseline distortion. Therefore, the goal of a follow-up study should be to further improve the capability of the CNN upon more realistically simulated in vivo spectra such that the CNN in itself can either filter out contaminated spectra prior to further analysis or still extract uncontaminated intact metabolite spectra therefrom. Or at least, a separate processing unit for spectral quality control [128, 129, 142, 148] should be integrated into the working pipeline prior to the CNN. Nonetheless, the observation that our CNN successfully mapped the degraded in vivo spectra into the intact metabolite spectra with negligible residual signal in the difference spectra (Figure 2-6 (P)-(T) and Figure 2-7 (P)-(T)) without being ever trained on the actual in vivo spectra is in support of the utility of our simulated brain spectra.

This study was evaluated CNN only on the in vivo spectra that were acquired with the same voxel size ($2 \times 2 \times 2$ cm^3) and from the same brain location (left frontal lobe). However, simulated spectra have a wide range of relative metabolite concentration ratios and substantially low SNR and large line broadening with respect to those for typical human brain spectra (Figure 2-2 and 2-3).

The spectral baseline was also modeled with varying relative amplitude and linewidth for all MMs groups in consideration of its potential regional and/or disease state-dependent variability [37, 38, 149, 150]. Therefore, the proposed CNN may also be tolerant of changes in the size and location of the voxel to a certain extent. However, in order to apply CNN to the spectra with drastically deviated relative metabolite concentration ratios from those found in normal human brain such as brain tumor spectra, the CNN should be retrained accordingly with additional simulated spectra.

Figure 2-10 and Figure 2-11 show the responses of CNN in its current form to previously unseen simulated spectra that mimic brain tumor spectra (Figure 2-10) and that are severely contaminated by strong water signal (Figure 2-11 (B)), ghosting artifact (Figure 2-11 (C)), or both. (Figure 2-11(D)), respectively.

The spectrum in Figure 2-10 (A) is one of the spectra in the test set. Although tCho appears higher and tNAA appears lower than those in the normal human brain spectra, they are still within the range of the relative metabolite concentration ratios used in the simulation of our original 50,000 spectra for the training/validation/test of the CNN. Despite the poor SNR, the difference spectra between the ground truth spectrum and the predicted spectrum by the CNN show only a small amount of residual signal mainly in the major peak areas such as tCho, tCr, and tNAA. In Figure 2-10 (B), the spectrum now contains strong tCho and Lac signal, of which the relative concentration ratios substantially deviate from those in the training set. As a result, a relatively large residual tCho signal is observed at ~ 3.2 ppm in the difference spectrum (GT – Pred.). For Lac, the CNN appears to have treated it as a non-metabolite signal. The Lac signal is almost completely suppressed in the predicted spectra together with the spectral baseline. Consequently, the difference spectrum (GT – Pred.) shows as much Lac signal as in the ground truth spectrum (the residual Lac signal is also observed at ~ 4.1 ppm). n Figure 2-10 (C), the spectrum contains high tCho signal and low tNAA signal, of which the relative

concentration ratios substantially deviate from those in the training set (Lac concentration ratio being within the range used for the original simulation). In addition, strong lipid signals were also included in the spectrum. For the modeling of the lipid signals, 10 line functions were used, and combined into 5 groups (~ 0.9, ~ 1.3, ~ 2.0, ~ 2.3, and ~ 2.8 ppm with the relative amplitude ratio of ~ 0.5 : 1 : 0.1 : 0.1 : 0.1 [39, 133, 151]). Similar to Figure 2-10 (B), relatively large residual signals of tCho and tNAA are observed in the difference spectrum (GT – Pred.) at ~ 3.2 and ~ 2.0 ppm, respectively. As with the Lac signal in Figure 2-10 (B), the CNN appears to have treated the strong lipid signal (esp., the methylene group at ~ 1.3 ppm) as a non-metabolite signal (this time correctly) and almost completely suppressed it in the predicted spectra together with the spectral baseline.

The spectrum in Figure 2-11 (A) is one of the spectra in the test set, from which the CNN resulted in the most accurate quantitative outcome. Almost negligible residual signals are observed in the corresponding difference spectrum (GT – Pred.). A strong residual water signal (Figure 2-11 (B)), a strong ghosting artifact (Figure 2-11 (C)), or both residual water signal and ghosting artifact (Figure 2-11 (D)) were artificially added to the spectrum in Figure 2-11 (A) to simulate those spectra shown in Figure 2-11 (B) through (D), respectively. In Figure 2-10 (B), the strong residual water signal that the CNN has never seen before was completely suppressed in the predicted spectrum (GT – Pred.) together with the spectral baseline. However, higher residual signals are observed in that predicted spectrum. In Figure 2-11 (C), a strong ghosting artifact that the CNN has never seen before completely ruined the spectrum in the ~ 3.5 – 4.3 ppm range. Nonetheless, the CNN successfully predicted the metabolite only spectrum in that spectral region according to the minor residual signals in the resulting difference spectrum (GT – Pred.). However, the strong ghosting artifact appears to have influenced the prediction of other spectral regions such as ~ 3.2 and ~ 2.0 ppm regions according to the slightly higher residual tCho signal and more noticeable residual tNAA signal in the corresponding difference spectrum in Figure 2-11 (C) with respect to those in the difference spectrum in Figure 2-11 (A). In Figure 2-11 (D), the strong residual water signal and ghosting artifact are combined to completely ruin the spectrum in the ~ 3.5 – 4.3 ppm range. Despite the fact that the CNN has never seen before such a severely degraded spectrum, it almost completely suppressed the contaminators in the

predicted spectrum. The residual signals in the resulting difference spectrum, however, are not negligible across the ~ 2.0 – 4.0 ppm range.

In summary, our CNN has a great potential for robust metabolite quantification from normal human brain spectra. However, the capability of our CNN should be further improved in order to apply it to those spectra that substantially deviate from typical, normal human brain spectra due to severe contamination resulting from unwanted signals such as strong residual water signal and ghosting artifact and/or due to aberrant relative metabolite concentration ratios as in brain tumor spectra. This would require more extensive training of the CNN on more realistically simulated in vivo spectra such that the CNN in itself can either filter out those atypical spectra prior to further analysis or still extract uncontaminated intact metabolite spectra therefrom.

To the best of our knowledge, the vendor-provided basis set is most widely used in the LCModel analysis. However, for more direct comparison between the two approaches, the metabolite phantom spectra that were used for the simulation of the brain spectra need to be incorporated as a basis set in the LCModel analysis. Combined with voxel-specific spectral baseline information such as metabolite-nulled spectra [41, 152], they would have improved the performance of the LCModel analysis in our study. Although our CNN requires data pre-processing such as Fourier transformation of the FID signal followed by the zeroth and first order phase correction, they are only minor works. In terms of fully automated metabolite quantification, a neural network may be preferred that directly uses the time domain signal as an input [107] and is capable of managing data quality. [128, 129, 142, 148] Such an advanced CNN might also be developed for MRSI, in which case the processing time for MRSI data would substantially be reduced. The feasibility of automatic quality management for MRSI data sets has been demonstrated previously. [128, 148]

One additional limitation of this study is the difficulty of clearly understanding what our CNN has learned for the mapping of the degraded brain spectra into the intact metabolite spectra. To address such a black box problem of artificial intelligence development of techniques for visualizing neural networks is one of the active research areas in deep learning. [153] For those who may be interested in, the layer activations of the CNN are shown in Figure 2-12, Figure 2-

13, Figure 2-14 and Figure 2-15 for several representative network layers. Finally, given that we employed a quite generic neural network architecture in the design of our CNN and that the function mapping the degraded brain spectra into intact metabolite spectra can be approximated by using different deep learning models and architectures [154], development of a CNN with better performance than ours is likely.

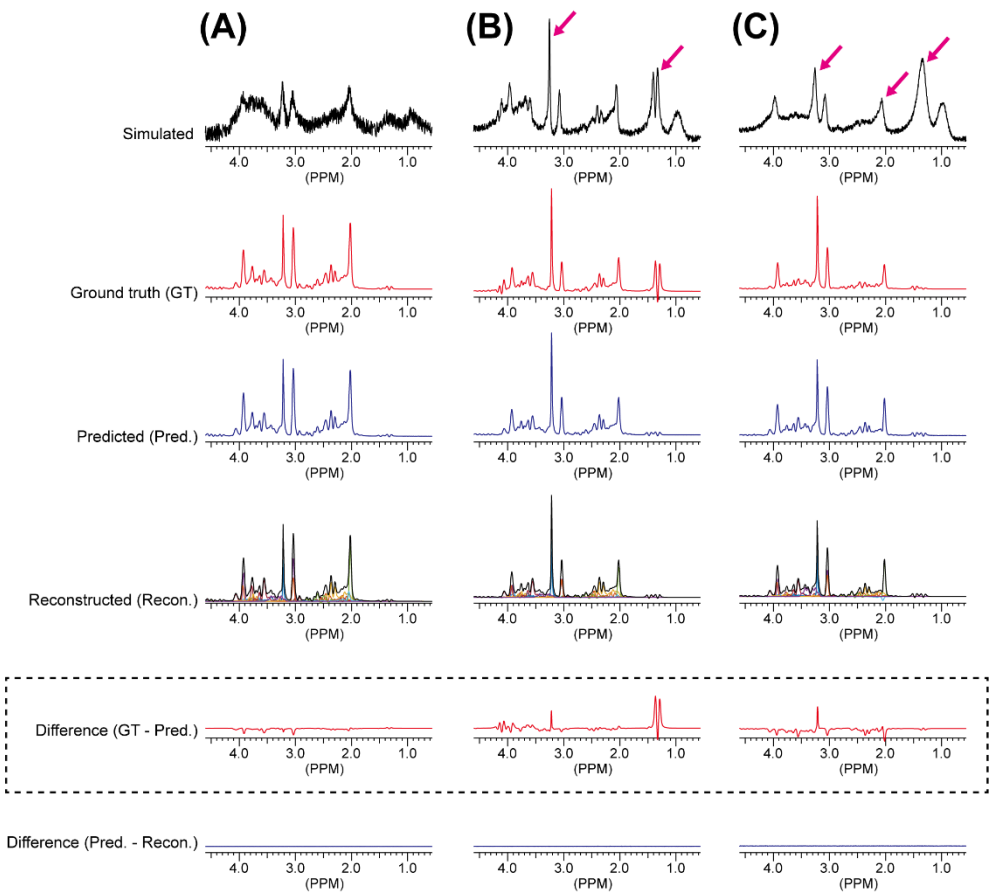


Figure 2-10. The responses of the CNN in its current form to the simulated spectra (A) that mimic brain tumor spectra (B) – (C).

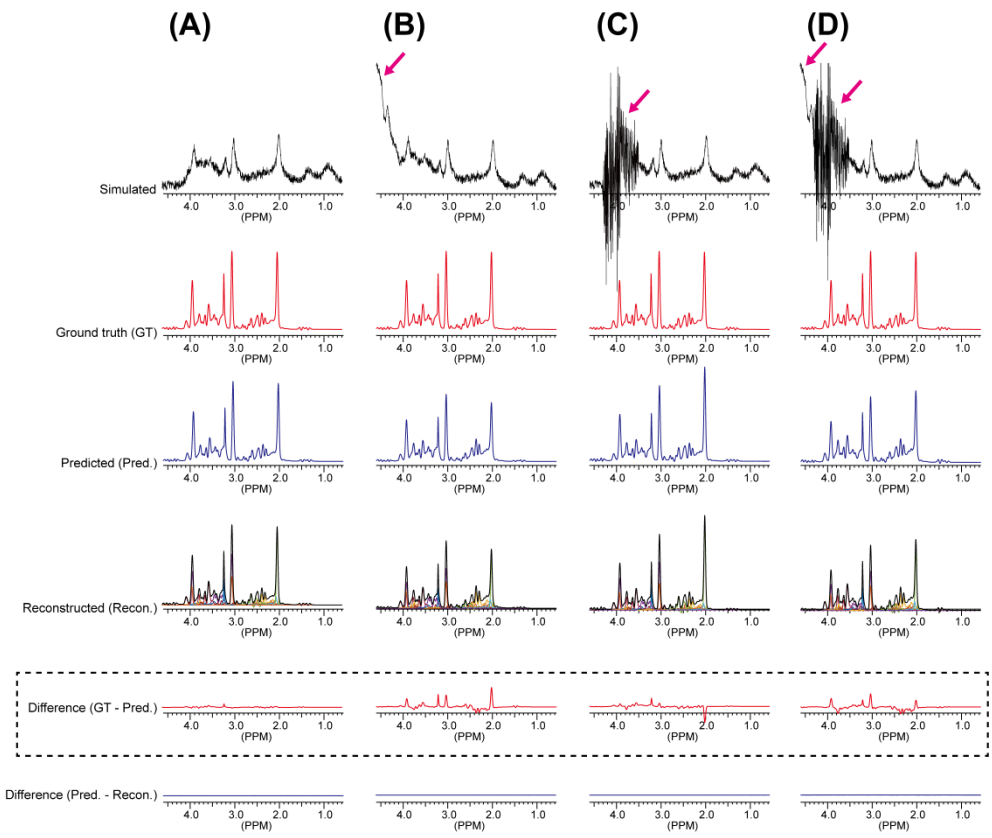


Figure 2-11. The responses of the CNN in its current form to the simulated spectra (A) that are severely contaminated by strong residual water signal (B), ghosting artifact (C), or both (D).

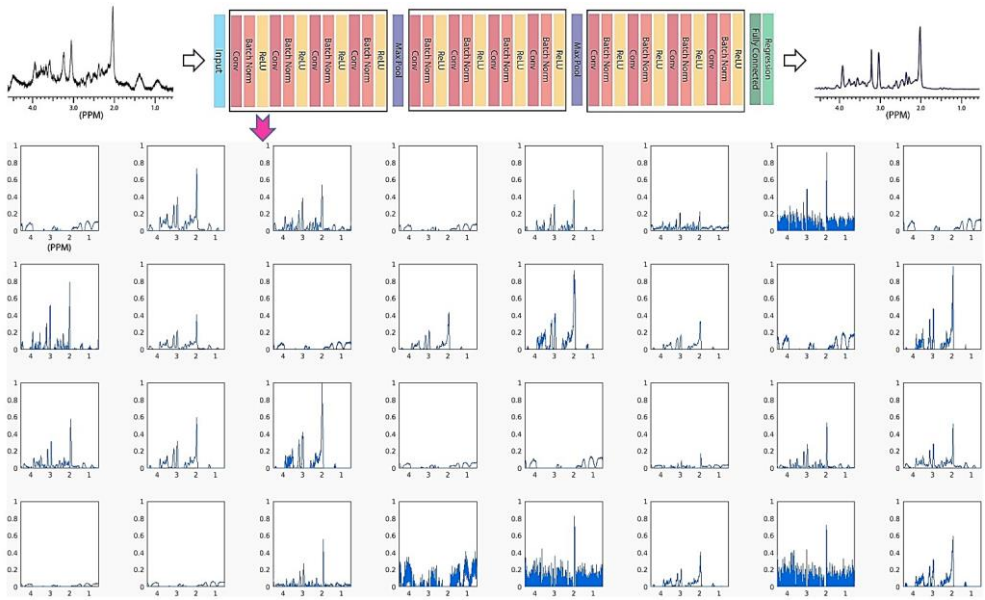


Figure 2-12. The activations of the first ReLU layer for all of the 32 channels.

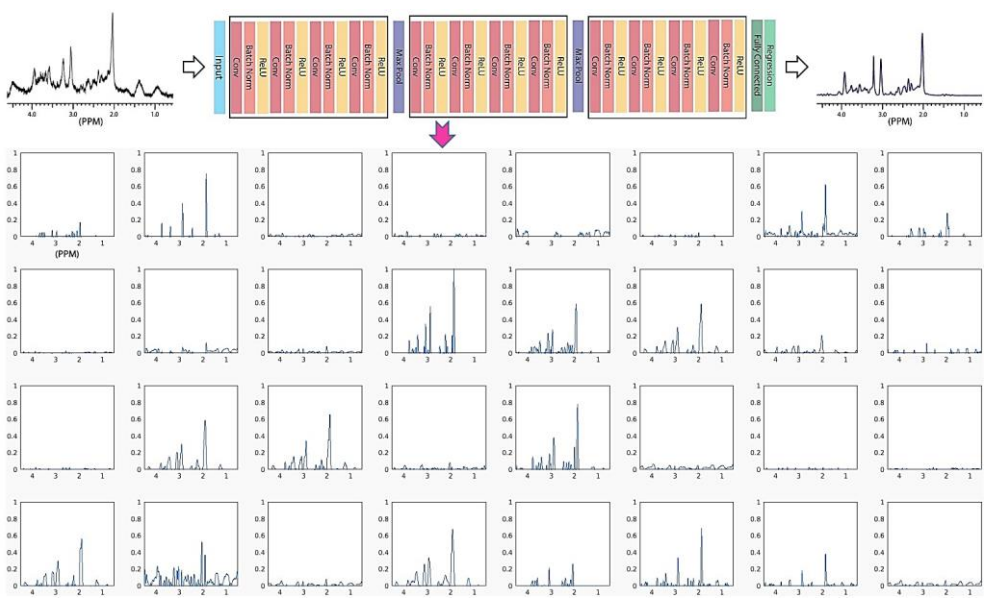


Figure 2-13. The activations of the fifth ReLU layer for all of the 32 channels.

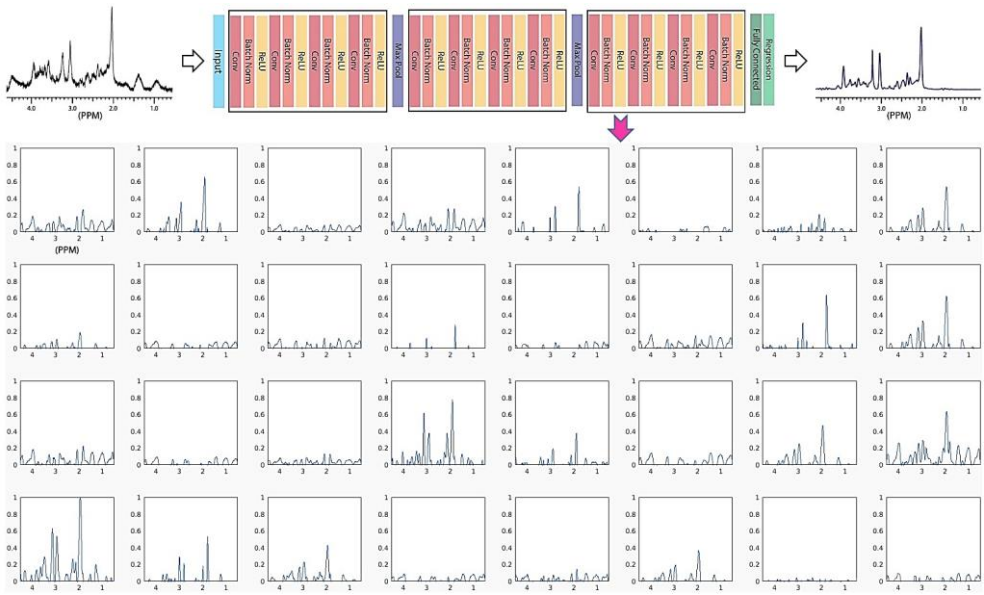


Figure 2-14. The activations of the ninth ReLU layer for all of the 32 channels.

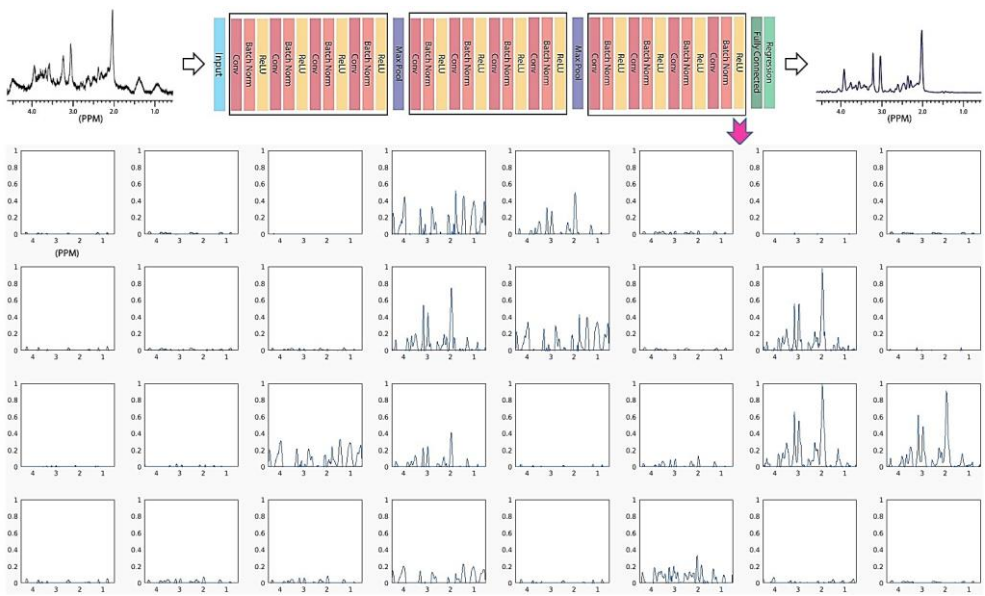


Figure 2-15. The activations of the twelfth ReLU layer for all of the 32 channels.

Chapter 3. Deep learning-based target metabolite isolation and big data-driven measurement uncertainty estimation in proton magnetic resonance spectroscopy of the brain

3.1. Introduction

For quantification of individual metabolites from brain spectra acquired at a short echo time (TE) by using by using ^{1}H-MRS, a NLSF has been most commonly used. [55, 116] However, it can be challenging due to limited SNR, spectral overlap between metabolite signals and the presence of unknown baseline. [35, 55, 126, 155] Therefore, development of a more robust method is still on demand in ^{1}H-MRS.

In accordance with the recent revolutionary advances in deep learning [99], studies supporting the applicability of machine learning/deep learning in ^{1}H-MRS have been rapidly accumulating. [107, 108, 110, 111, 128, 129, 142] In these studies the potential of the machine learning-/deep learning-based spectral quality management [128, 129, 142] and metabolite quantification [107, 108, 110, 111] has been clearly demonstrated. For machine learning-/deep learning-based metabolite quantification, however, estimation of the measurement uncertainty in the machine-predicted metabolite concentrations such as the CRLB in the LCModel analysis [55] should also be important [108], and yet has not been reported to date.

In NLSF the accuracy of metabolite quantification is known to depend mainly on SNR, linewidth and degree of spectral overlap [55], and this appears to hold also in the deep learning-based metabolite quantification to a certain extent. [108, 110, 111] Therefore, knowledge about those three factors may be utilized for the prediction of the quantitative errors in the machine-predicted metabolite concentrations. While SNR and linewidth are typically measured from a representative singlet such as tNAA, the degree of spectral overlap or signal-to-background ratio (SBR) as its quantitative metric needs to be measured from each

The work presented in this chapter has been published; Lee, H. H., & 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(4),1689-1706. doi:10.1002/mrm.28234.

individual metabolites. Thus far, the outputs of the CNNs used in the deep learning-based metabolite quantification are in the form of direct metabolite concentrations [108, 110], individual metabolite spectra [111], or whole (metabolite-only) spectrum as described in Chapter 2. For the second approach the estimation of SBR of individual metabolites from the CNN-output is more straightforward, and the function to be approximated by a CNN may be computationally less costly than the first approach where the CNN is required to learn the mapping from input spectra directly to metabolite concentrations.

To this end, this chapter aimed to develop and evaluate CNNs specific to each individual metabolites, which isolate signals only from one designated metabolite and one fixed reference metabolite (e.g., tCr) and suppress the rest of the signals including baseline signal and spectral noise. The estimation of SBR, as well as quantification, of the metabolites are achieved from the spectrally isolated and denoised CNN-output spectra. Then, the resulting SBR that is subject to error due to limited performance of the CNNs is adjusted according to the relationships between the CNN-predicted SBR and the corresponding GT SBR that were pre-defined for each metabolite by exploiting the big training data set. The adjusted SBR together with the SNR and linewidth of the input spectrum are used for the prediction of the quantitative errors. The accuracy of the metabolite quantification by the deep learning-based spectral isolation and the feasibility of the big data-driven quantitative error estimation were tested first on two different groups of simulated rat brain spectra and further on in vivo spectra at 9.4T.

3.2. Methods and Materials

3.2.1. Acquisition and Analysis of in vivo Rat Brain Spectra

The animal study was approved by the Institutional Animal Care and Use Committee (IACUC). In vivo data were acquired from heathy Sprague-Dawley rats (n= 15, 7±1 weeks old, 220±30 g) to extract a prior knowledges such as SNR and linewidth for the simulation of rat brain spectra.

MR data were collected on a 9.4T animal MR scanner using a single channel surface coil (20-mm in diameter) for RF transmission and signal reception (Agilent Technologies, Santa Clara, CA, USA) as previously described. **[156]** After

acquisition of scout images, voxels were defined in the cerebellum and the thalamus ($3.0\times2.8\times2.8$ mm^3) for each animal. Then, auto-shimming was performed followed by manual adjustment.

^{1}H-MRS data were collected using a SPECIAL sequence [51] (TR/TE= 5500 / 3.45 ms, 2048 data points, spectral BW= 5 kHz, and 32-step phase cycling). For minimal voxel displacement [157], the carrier frequency was shifted by -2.3 ppm from the water resonance. A variable pulse power and optimized relaxation delays (VAPOR) [158] module in combination with OVS modules [45] preceded the main sequence. To investigate the performance of the CNNs against different SNR, 12 spectra with a NSA of 32 were acquired consecutively from each voxel. Then, by combining these data 12 spectra with different NSA ranging 32 – 384 (step size=32) were generated.

Metabolite-nulled spectra were collected as a surrogate of the spectral baseline from the thalamus of three rats by using double inversion [156], and used for one of the simulated spectra groups.

The acquired data were zero-filled to 4096, Fourier transformed, and 0^{th}- and 1^{st}-order phase corrected by using jMRUI (v.5.2) [116] with no line-broadening. The residual water signal was removed by the HLSVD filter. [119] In the case where the in vivo data are used as CNN-inputs, the processed spectra were cropped to 1308 data points corresponding to 0.5 – 4.5 ppm range as with the simulated spectra.

3.2.2. Simulation of Metabolite Basis set

The spectral basis set was simulated for 27 metabolites in response to a single spin-echo by using GAMMA [59] according to the reported chemical shifts and J-coupling constants [34]. They were alanine Ala, Asp, GABA, Glc, Gln, Glu, GPC, GSH, Lac, mI, NAA, NAAG, PCh, PCr, PE and Tau. Those metabolites in the first group and the second group are referred to as major and minor metabolites, respectively.

3.2.3. Acquisition of Metabolite Basis set in Phantom

Metabolite phantoms were made in cylindrical tubes (50 mL) for each one of the 17 major metabolites (TSP (0.5 mM) and sodium azide (NaN_3; 0.1%) also included in the phantoms; pH= 7.0 – 7.5; Sigma-Aldrich). Spectra were acquired by using the

same sequence that was used for the in vivo data collection except for TR (=10 sec) and NSA (=128).

3.2.4. Simulation of Rat Brain Spectra using Simulated Metabolite and Baseline Basis Sets

Based on the simulated metabolite basis set and a priori knowledge from the in vivo spectra, rat brain spectra were simulated by using an in-house script written in Python (v.3.6; Python Software Foundation, Wilmington, DE) as previously described in Section 2.2.3.

First, the upper and lower bounds of the metabolite concentrations for normal rat brain were determined according to the literature (Figure 3-1). [132, 159] These concentration ranges were evenly divided by the number of spectra to be simulated for the optimization of the CNNs (N=100,000) for all metabolites. Then, metabolite spectra were generated by combining all simulated metabolite basis spectra according to randomly selected relative metabolite concentration ratios within the concentration ranges.

Second, the spectral baseline in the range of 0.5 – 4.5 ppm was simulated using 25 Gaussian model functions based on the previous studies. [132, 159] These model functions were grouped into 10 resonance groups (M1-M10) and their relative amplitudes were normalized according to the literature. [132, 159] Then, spectral baselines were generated by randomly varying the linewidth of the 25 model functions and the relative amplitude of the M1-M10 groups within ± 20 % and ± 10 % from their initial values, respectively.

Third, the metabolite signal and baseline signal were combined by randomly varying the initial ratio of 1 : 0.9 within ± 25 % followed by the application of Lorentzian line broadening (10 – 30 Hz as measured from tNAA (~2.0 ppm)) and phase-/frequency-shift (± 1° and ± 21 Hz, respectively) (Figure 3-2).

Forth, all simulated spectra were cropped to 1308 data points (0.5 – 4.5 ppm) for efficient CNN training. Then, the SNR of the spectra were adjusted by adding random noise such that the SNR of tNAA was in the range of ~5 – 30. The spectral noise was estimated by measuring the two standard deviations [160] in the ~8.0 – 10.0 ppm range prior to the cropping of the spectra.

Fifth, for each of the simulated brain spectra the corresponding GT target spectra (metabolite-specifically isolated spectra) were prepared for individual metabolites without adding the spectral baseline and noise. The tCr was used as the reference metabolite for quantification and thus the two singlets of tCr were also included in all of the GT target spectra.

Finally, 100,000 spectra were simulated for the optimization of the CNNs. In addition, 10,000 spectra were simulated for the testing of the optimized CNNs. This set of the simulated spectra using the simulated metabolite and baseline basis sets is referred to as simulated spectra set I.

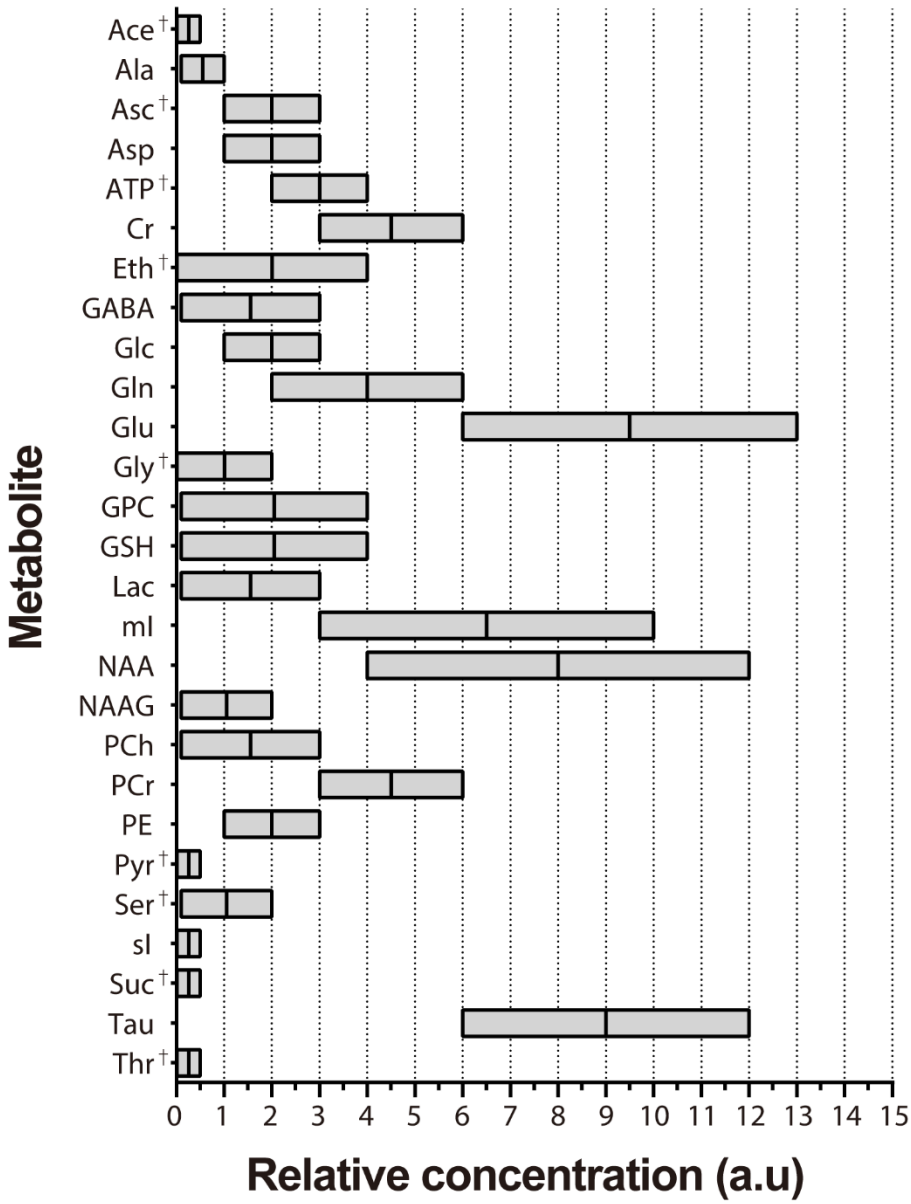


Figure 3-1. The relative metabolite concentration ranges used for the simulated spectra set I. Those 10 metabolites referred to as minor metabolites are indicated with †. The rest of the 17 metabolites are referred to as major metabolites. The solid line in the middle of each bar indicates the mean value.

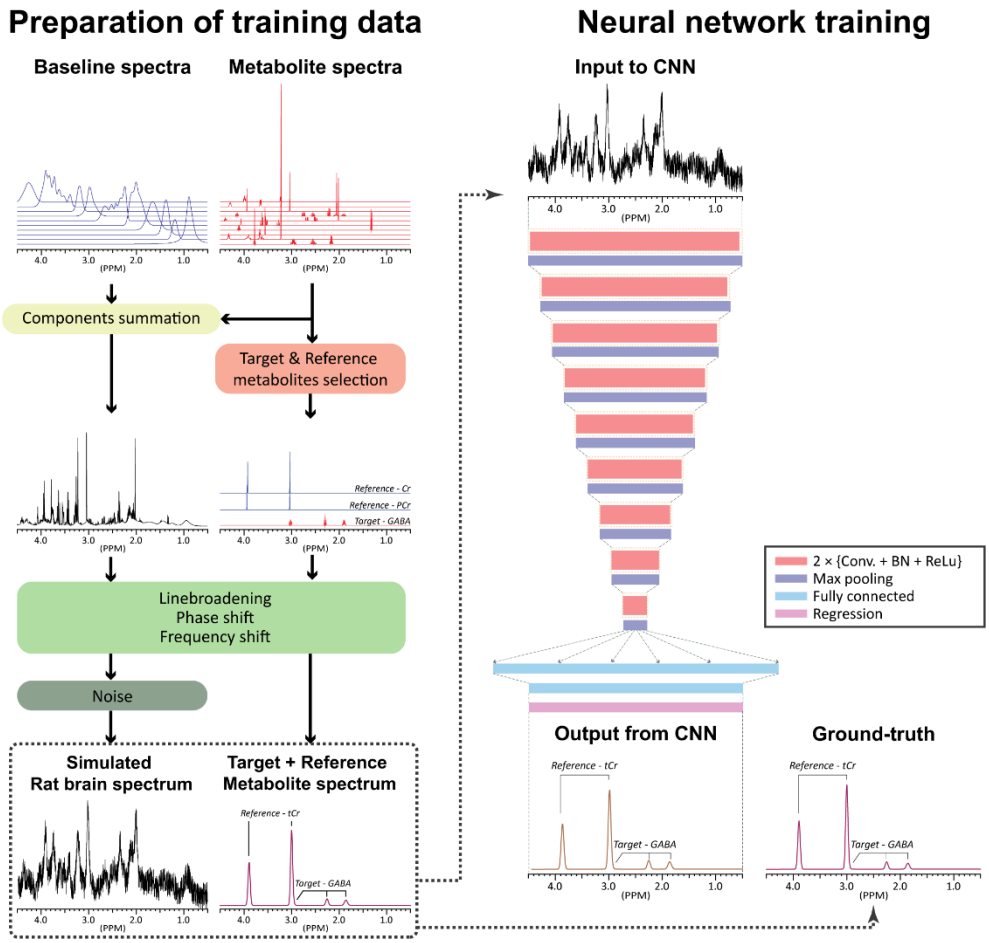


Figure 3-2. A schematic of the simulation of rat brain spectra at 9.4T. The simplified structure of the CNN is also shown. The simulated spectra with different relative metabolite concentrations, baseline, SNR, linewidth, frequency shift, and phase shift are used as CNN-input spectra. For each CNN-input spectrum the constituent metabolite spectra without artificial spectral noise are individually saved and used as ground truth spectra. As a reference signal for quantification tCr signal is combined with all individual metabolite spectra. Using these CNN-input and ground truth spectra CNNs are trained one by one for each metabolite such that each CNN filters out signals only from a designated target metabolite (GABA in this example) and tCr. After completion of training, an input rat brain spectrum is fed into the 27 CNNs and tCr-normalized quantitative results are obtained for the 27 metabolites.

Metabolites	Absolute concentration (mM)			Ground truth for LCModel and jMRUI (/tCr concentration)*			Ground truth for CNN (/tCr area)†			LCM GT / CNN GT‡
	Phantom 1	Phantom 2	Phantom 3	Phantom 1	Phantom 2	Phantom 3	Phantom 1	Phantom 2	Phantom 3	
Ala	0.5	1	0.75	0.07	0.11	0.08	0.08	0.12	0.09	0.89
Asp	3	1	2	0.43	0.11	0.21	0.29	0.07	0.14	1.49
Cr	4	5	4.5	0.57	0.56	0.47	-			
GABA	2	1	0.5	0.29	0.11	0.05	0.69	0.27	0.13	0.42
Glc	2	3	1	0.29	0.33	0.11	0.32	0.37	0.12	0.9
Gln	5	3	4	0.71	0.33	0.42	0.99	0.46	0.58	0.74
Glu	8	10	12	1.14	1.11	1.26	1.54	1.5	1.71	0.72
GPC	1	3	2	0.14	0.33	0.21	0.18	0.42	0.27	0.34
GSH	0.5	0.75	1	0.07	0.08	0.11	0.21	0.25	0.31	0.78
Lac	1	2	3	0.14	0.22	0.32	0.14	0.21	0.3	1.04
mI	6	9	8	0.86	1	0.84	1.42	1.66	1.4	0.6
NAA	9	6	12	1.29	0.67	1.26	1.36	0.71	1.34	0.94
NAAG	0.5	0.25	0.1	0.07	0.03	0.01	0.09	0.04	0.01	0.6
PCh	1	2	0.5	0.14	0.22	0.05	0.4	0.63	0.15	0.34
PCr	3	4	5	0.43	0.44	0.53	-			
PE	1	1.5	2	0.14	0.17	0.21	0.09	0.1	0.13	1.67
Tau	12	10	8	1.71	1.11	0.84	2.24	1.45	1.1	0.76

Table 3-1. Metabolite concentrations in the three phantoms and corresponding ground truth metabolite content.

* Ground truth metabolite content upon measurement by LCModel and jMRUI where the concentrations of the metabolites were normalized by the concentration of tCr.

† Ground truth metabolite content upon measurement by CNNs where the area of the signals of the metabolites were normalized by the area of the methyl signal of tCr at ~3.0 ppm in the CNN-output spectra.

‡ The ratio of LCModel to CNN ground truth.

3.2.5. Simulation of Rat Brain Spectra using Metabolite Phantom Spectra and in vivo Baseline

To evaluate the trained CNNs further on more realistically simulated in vivo spectra, spectra were simulated by using metabolite phantom spectra in combination with in vivo baseline.

Three metabolite phantoms were made in cylindrical tubes (50 mL) including those 17 major metabolites with different relative metabolite concentration ratios (Table 3-1) (TSP (0.5 mM) and NaN_3 (0.1 %) also included; pH=7.0-7.5). Spectra were acquired by using the same sequence that was used for the in vivo data collection except for TR (=10 sec) and NSA (=128).

The phantom spectra were line-broadened (10 Hz) and then combined with the in vivo baselines. For each of the three combined spectra the linewidth and SNR were adjusted in the ranges of ~10 – 20 Hz (step size = 5 Hz) and ~5 – 30 (step size = 5), respectively. Thus, a total of 54 spectra were generated from the original three combined spectra (3 linewidth × 6 SNR × 3 spectra = 54 spectra). This set of the simulated spectra using the metabolite phantom spectra and in vivo baseline is referred to as simulated spectra set II.

3.2.6. Design and Optimization of CNN

Only the real (absorption) part of the spectrum was used as the input and output of the CNNs. A base CNN was designed (Figure 3-2) and optimized for individual metabolites by using Tensorflow framework (Tensorflow v1.4.1; Google LLC, Mountain View, CA) on 4 graphics processing units (GPUs; NVIDIA Titan Xp). The base CNN consisted of 1 input layer, 9 convolution blocks, 9 max pooling layers, 2 fully connected layer, and 1 regression layer. Each convolution block was composed of 2 repetitions of one-dimensional convolution – batch normalization – rectified linear unit (ReLU) activation layers. The following hyperparamters were commonly used in the CNN optimization: max pooling pool size=2×1, max pooling stride=1, batch size=128, 1 validation per epoch. For the training of the CNNs an adaptive moment (ADAM) estimation algorithm [84] was used. The loss function was MSE.

The 100,000 spectra simulated for the optimization of the CNNs in the simulated spectra set I were randomly split into a training (N=80,000) and a

validation (N=20,000) sets. The spectra in the training set were min-max normalized (height × width × channels × array = 1308 × 1 × 1 × 80,000). Then, the CNNs were Bayesian-optimized [135] with a coarse-to-fine search for individual metabolites. First, the following coarsely selected hyperparameter ranges were used with max epoch=3 and number of objective evaluations = 300: filter size = 3 × 1 to {int(input height/2n)} × 1 where int stands for a closest integer and n is convolution block index (1–9) for each convolution layer; number of filters = 1 to 100 for each convolution layer; the number of neurons of the first fully connected layer = {1 to 5} × 1308; initial learning rate = 1 × 10-5 to 1 × 10-2. From this first optimization step, 10 sets of the optimized hyperparameters were selected for each metabolite that resulted in the 10 smallest validation losses. Then, the hyperparameter ranges were fine-tuned according to the lower and upper limits of the above optimized hyperparameters, and an additional Bayesian optimization was performed with max epoch=5 and number of objective evaluations = 200. After this second optimization step, 10 sets of the optimized hyperparameters were selected again for each metabolite.

Then, the 100,000 spectra simulated for the optimization of the CNNs in the simulated spectra set I were randomly divided into 5 groups and a 5-fold cross validation was performed [161] for each of the 10 hyperparameter sets for individual metabolites. From this step the best hyperparameter set was determined for each metabolite based on the mean validation loss over the 5-folds.

Finally, a training set (N=80,000) and a validation set (N=10,000) were randomly assigned from the 100,000 spectra for the optimization of the CNNs in the simulated spectra set I, and CNNs were trained with the optimized hyperparameter (Table 3-2) for individual metabolites (maximum epoch=2000). An early stopping rule was applied with 100 epochs patience. The learning rate was scheduled to drop every 10 epochs by a factor of 0.9. The total optimization time for each metabolite was 120–168 hours.

Metabolites	Filter size in convolution layers									Number of filters in convolution layers									Number of neurons*	Initial learning rate
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	1st	2nd	3rd	4th	5th	6th	7th	8th	9th		
Ace	178	153	133	65	27	6	9	3	3	18	56	49	13	20	59	33	82	62	5693	0.007404
ATP	168	237	31	43	8	12	8	3	3	24	63	92	44	17	8	76	62	51	4668	0.006661
Ala	96	245	78	37	9	14	7	4	3	11	39	51	15	21	57	53	24	29	5855	0.006354
Asc	17	156	73	63	30	14	4	4	3	32	73	30	79	27	39	77	71	47	5763	0.005434
Asp	22	146	154	11	32	3	4	4	3	7	89	44	26	60	91	18	56	75	6316	0.006156
Eth	260	248	140	21	27	4	6	5	3	26	58	12	58	10	5	75	56	47	6110	0.006247
GABA	33	24	17	57	19	6	6	3	3	49	33	66	22	16	50	51	97	86	6476	0.007486
Glc	85	114	126	43	17	7	10	4	3	86	25	55	59	88	72	67	43	60	4599	0.006746
Glu	168	69	4	40	14	10	9	5	3	52	24	29	38	22	73	21	58	45	6353	0.006154
Gln	65	185	134	9	34	16	4	3	3	14	72	7	73	4	42	29	28	48	5886	0.007347
GPC	118	179	8	57	7	8	5	5	3	27	20	14	33	75	34	77	36	90	6345	0.006534
GSH	36	110	38	16	5	18	8	4	3	56	37	38	76	31	50	13	54	58	6360	0.005054
Gly	81	204	111	58	4	10	7	5	3	64	8	26	41	69	57	77	52	40	5506	0.006368
Lac	11	28	19	46	23	9	4	4	3	51	24	61	20	22	23	36	86	83	6109	0.00644
mI	75	16	135	18	13	15	4	4	3	28	33	66	67	61	29	62	87	48	6245	0.006057
NAA	199	78	8	65	8	5	4	5	3	9	36	46	77	25	60	42	30	85	5910	0.005944
NAAG	168	257	8	13	12	14	7	3	3	19	73	34	71	12	69	20	75	82	5257	0.006486
PCh	27	153	133	23	12	4	4	4	3	31	34	19	73	50	10	39	30	90	6404	0.006901
PE	65	197	66	13	12	11	5	5	3	56	34	18	50	19	55	67	95	31	4961	0.006891
Pyr	22	300	101	16	23	5	7	4	3	84	22	55	11	9	62	44	59	83	5283	0.006836
sI	30	297	116	39	13	15	7	4	3	19	50	12	78	57	25	14	20	80	6000	0.006479
Ser	257	10	99	6	37	6	7	5	3	26	67	28	67	5	38	36	57	73	5716	0.006272
Suc	141	204	89	46	11	14	5	4	3	75	40	7	87	25	15	38	71	65	4530	0.006261
Tau	22	253	109	9	19	10	5	4	3	37	16	22	30	32	76	79	29	89	6225	0.006225
Thr	29	105	104	68	21	18	5	4	3	8	17	54	37	22	72	66	47	88	5846	0.006129
tCr	12	106	27	50	4	4	5	5	3	37	7	71	57	54	72	72	61	61	4814	0.006912

Table 3-2. Optimized hyperparameters of the CNNs for individual metabolites.

* The number of neurons of the first fully connected layer

3.2.7. Metabolite Quantification from the CNN-predicted Spectra

Prior to the use of the trained CNNs data needs to be pre-processed manually such as zero-filling, Fourier transform, phase and frequency correction, removal of residual water signal, and cropping the 0.5 – 4.5 ppm range of the real part of the spectrum. Metabolite quantification is achieved by feeding the pre-processed input spectrum into the 27 metabolite-specific CNNs and measuring the area of the metabolite signal (summing signal over the known spectral region) relative to that of the methyl signal of tCr (~3.0 ppm) in the output spectra. This process was included in the working pipeline for fully automated quantification. In the case where metabolite signal overlaps with the reference signal (such as GABA), the area of the rest of the metabolite signal is measured with respect to the methylene signal of tCr (~3.9 ppm) and then converted to that with respect to the methyl signal according to the area ratio between the two signals of tCr.

3.2.8. Prediction of Quantitative Error

In the preliminary observation the losses for both training and validation sets rapidly decrease until ~100th epoch. After that, the training loss decreases slowly but continuously towards its minimum. The corresponding validation loss also decays but in a far slower manner. Given this initial observation, CNNs were saved in the middle of training at the 100th epoch where the training and validation loss decay curves start deviating from each other. Then, CNN-predicted spectra were obtained for all metabolites from the training set by using the saved CNNs.

Figure 3-3(A) illustrates the metabolite-specific measurement of SBR for GABA as an example. The signal in SBR refers to the area of the peaks of a spectrally isolated metabolite. The background refers to the area of the rest of the signal in the spectral regions where the area of the spectrally isolated metabolite signal was estimated. The SBR is measured from both the CNN-predicted spectrum and corresponding GT spectrum. The SNR and linewidth are measured from tNAA in the input spectrum commonly for all metabolites.

Once the SNR, linewidth, and SBR are measured for all metabolites from the input spectra and each one of the CNN-predicted spectra, a 3-D SNR-linewidth-SBR space (SLS space) is formed (Figure 3-3(B)) for each metabolite where the SBR

is from the CNN-predicted spectra (SBR_{CNN}). Then, the SLS space is segmented heuristically using 3-D cuboidal bins. Given that the SNR and linewidth were evenly varied in the simulation of the spectra, their ranges are set to 5-35 and 10-30 Hz and divided into 30 and 20 equally spaced pieces, respectively. On the other hand, the SBR was not controlled during simulation of the spectra. Therefore, it is not necessarily evenly varying and its variation and range are metabolite-dependent. For this reason the range of the SBR (e.g., 0-0.35 for GABA) is divided with a variable bin width along the SBR-axis such that all bins contain a fixed, identical number of data points. In practice, the SBR range was heuristically divided into 20 pieces and each bin contained 6-7 data points ($80000 / (30 \times 20 \times 20) \sim 6.7$) such that the predicted error is large enough to include the GT error while maintaining the high correlation between the two errors. Next, the relationship between the SBRs estimated from the CNN-predicted spectra and the GT spectra is stored for all data points in the SLS space (Figure 3-3(C)). At the same time a 3-D error space (absolute percent errors of metabolite concentrations) is also stored as a function of SNR, linewidth, and the SBR measured from the GT spectra (SBR_{GT}) for each metabolite (Figure 3-3(D)).

In summary, for an input spectrum, the CNNs predict metabolite-specific isolated spectra. The SBR of the metabolites are measured from the CNN-predicted spectra (Figure 3-3(A)). The SNR and linewidth are measured from the input spectrum. Using the co-isolated tCr in the CNN-predicted spectra, the relative concentrations of the metabolites are estimated (Figure 3-3(A)). Then, a bin corresponding to the SLS coordinate is identified from the 3-D SLS space for each metabolite (Figure 3-3(B)). For all data points in that bin the SBR are adjusted to the GT SBR (Figure 3-3(C)). Finally, a set of measurement errors corresponding to the data points with the adjusted SBR are obtained from the 3-D error space (Figure 3-3(D)), among which the maximum error is chosen as the predicted error.

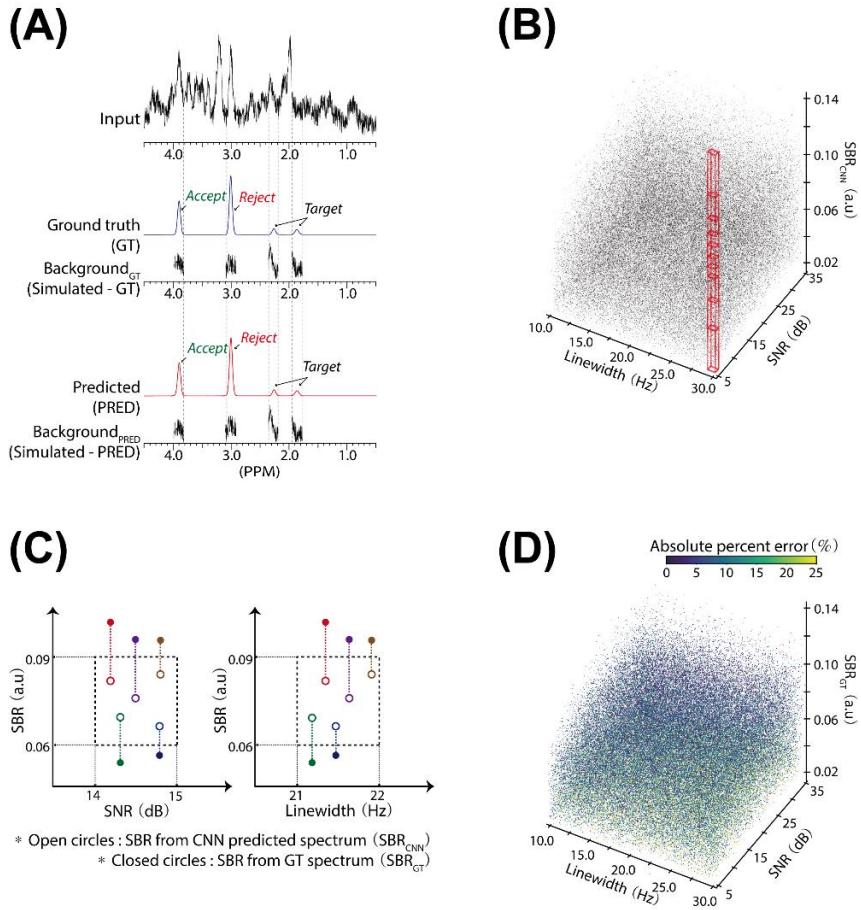


Figure 3-3. The construction of the database from the training data set for the prediction of the quantitative errors in the CNN-predicted metabolite concentrations. (A): The signal-to-background-ratio (SBR) of a target metabolite (GABA in this example) is estimated by measuring the signal area of the target metabolite and its background (Background) from both the CNN-predicted spectrum (PRED) and corresponding ground truth spectrum (GT). At the same time, the SNR and linewidth are measured from the CNN-input spectrum (Input). In this case, the GABA signal at ~3.0 ppm overlaps with the reference signal (the methyl signal of tCr at ~3.0 ppm). Therefore, the area of the methyl signal of tCr is not used (Reject). Instead, the methylene signal of tCr at ~3.9 ppm is estimated (Accept), and the area of the rest of the GABA signal (at ~2.3 and ~1.9 ppm) is measured with respect to the methylene signal of tCr. Then, it is converted to that with respect to the methyl signal of tCr according to the area ratio between the two signals of tCr. (B): Using the relationships among the SNR, linewidth, and SBR estimated from the CNN-predicted spectra (SBR_{CNN}) from the 80,000 spectra in the training data set, the SNR-linewidth-SBR space (SLS space) is formed for each target metabolite. Then, the space is segmented with 3-D bins, each of which contains ~6-7 data points. (C): the relationships between the SBR of a target metabolite estimated from the CNN-predicted spectra (SBR_{CNN}) and that from the GT spectra (SBR_{GT}) are stored for all data points in the SLS space. (D): a 3-D space of the quantitative errors obtained from the training data set is also stored for each target metabolite as a function of the SNR, linewidth, and SBR measured from the GT spectra (SBR_{GT}).

3.2.9. Evaluation of Proposed Method

For the simulated spectra sets I and II, the metabolite concentrations and corresponding errors predicted by the proposed method were directly compared with the GT values. For the predicted concentrations the resulting MAPE was calculated for each metabolite. For the predicted errors whether or not they are inclusive of (no less than) the GT errors was examined along with the correlations between the predicted and the GT errors.

For the simulated spectra set II, the metabolite concentrations and errors predicted by the proposed method were also compared with the concentrations and CRLB obtained by the LCModel (v.6.3-1J; [55]) and jMRUI (QUEST [54]) where the metabolite basis set acquired in phantom was used and the baselines used for the simulation were incorporated into the spectral basis set for better performance of the software. It should be noted in this analysis that CRLB are estimations of the minimum possible error that can be achieved given the amount of information available (i.e., not predictions of the real errors but of the level of uncertainty). Therefore, the results need to be interpreted carefully.

For the in vivo spectra the CNN-predicted metabolite concentrations were compared with those reported in the literature. [162]

3.2.10. Statistical Analysis

All statistical analyses were performed by using PRISM (v.6.01; GraphPad Software, Inc., La Jolla, CA, USA). For pair-wise group comparison a Student t-test was performed. A *P*-value of less than 0.05 was considered statistically significant. For multiple pair-wise group comparison the Bonferroni correction was applied by adjusting the significance level. For correlations between two variables the Pearson's correlation coefficient, r, was calculated.

For the analysis using LCModel and jMRUI only those results with CRLB≤50% [160] were included in the statistical analysis.

3.3. Results

3.3.1. Performance of Proposed Method on Simulated Spectra Set I

First, the performance of the proposed method was evaluated by using the 10,000 spectra simulated for the testing of the CNNs in the simulated spectra set I.

The SNR and linewidth of the input spectra ranged 4.88-30.23 (17.51±7.13) and 10.11-30.02 Hz (19.97±5.80 Hz), respectively.

Using the CNN-based spectral isolation method the MAPE of the reference metabolite tCr were 1.92% and 2.56% for the methyl (~3.0 ppm) and methylene (~3.9 ppm) peaks, respectively, and the ratio of their relative areas was 2:3.01, over the 10,000 test spectra.

Figure 3-4 (A) shows the MAPE of the 25 metabolites obtained by the CNN-based method. For the minor metabolites MAPE were mainly over 10%, but were less than 10% for Asc and ATP. For the major metabolites MAPE were close to 10% or less except for Ala, GPC, NAAG and PCh (14.79±11.12%). The MAPE of Gln, Glu, mI, NAA, and Tau were close to 5% or less. The correlations between the GT metabolite concentrations and the CNN-predicted metabolite concentrations for the simulated spectra set I are given in Figure 3-5.

Figure 3-4 (B) shows the percentages of the predicted quantitative errors that were inclusive of (no less than) the GT errors (absolute percent errors) for individual metabolites over the whole spectra in the test set. They are all larger than 85% except for Glc, GPC, and PCh. For the majority of the metabolites the predicted errors by the proposed method were highly correlated with the GT errors ($r=0.81 \pm 0.13$; 0.88 ± 0.09 for the major metabolites only). The correlation coefficients are shown for all metabolites in Table 3-3. The relationships between SBR_{GT} and SBR_{CNN} as well as between the GT errors and SBR_{GT}, SBR_{CNN}, SNR, and linewidth are shown for the 15 major metabolites in Figure 3-6.

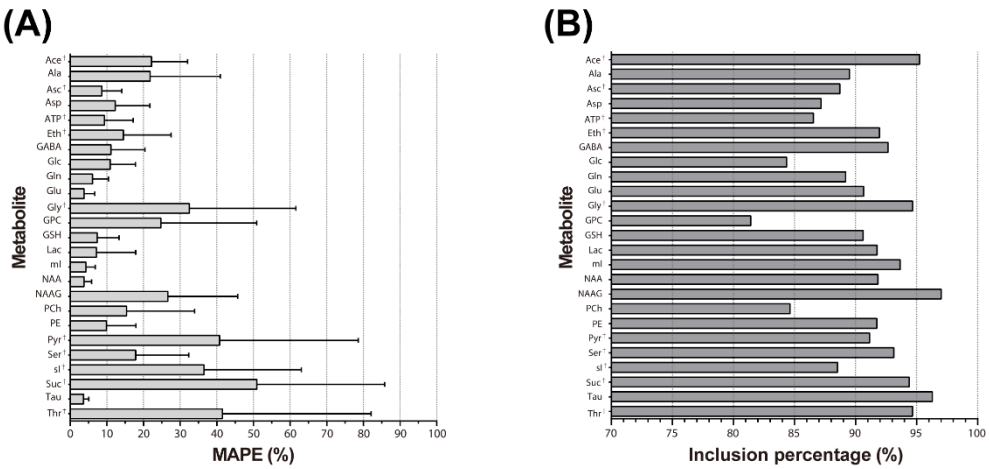


Figure 3-4. The performance of the proposed method on the simulated spectra by using the simulated metabolite and baseline basis sets (the 10,000 spectra simulated for the testing of the CNNs in the simulated spectra set I;). (A): Mean absolute percent errors (MAPE) of the 27 metabolites obtained by using the CNNs. The error bars indicate the standard deviations. (B): The percentages of the predicted quantitative errors that are inclusive of (no less than) the ground truth errors. For both A and B, those metabolites referred to as minor metabolites are marked with †.

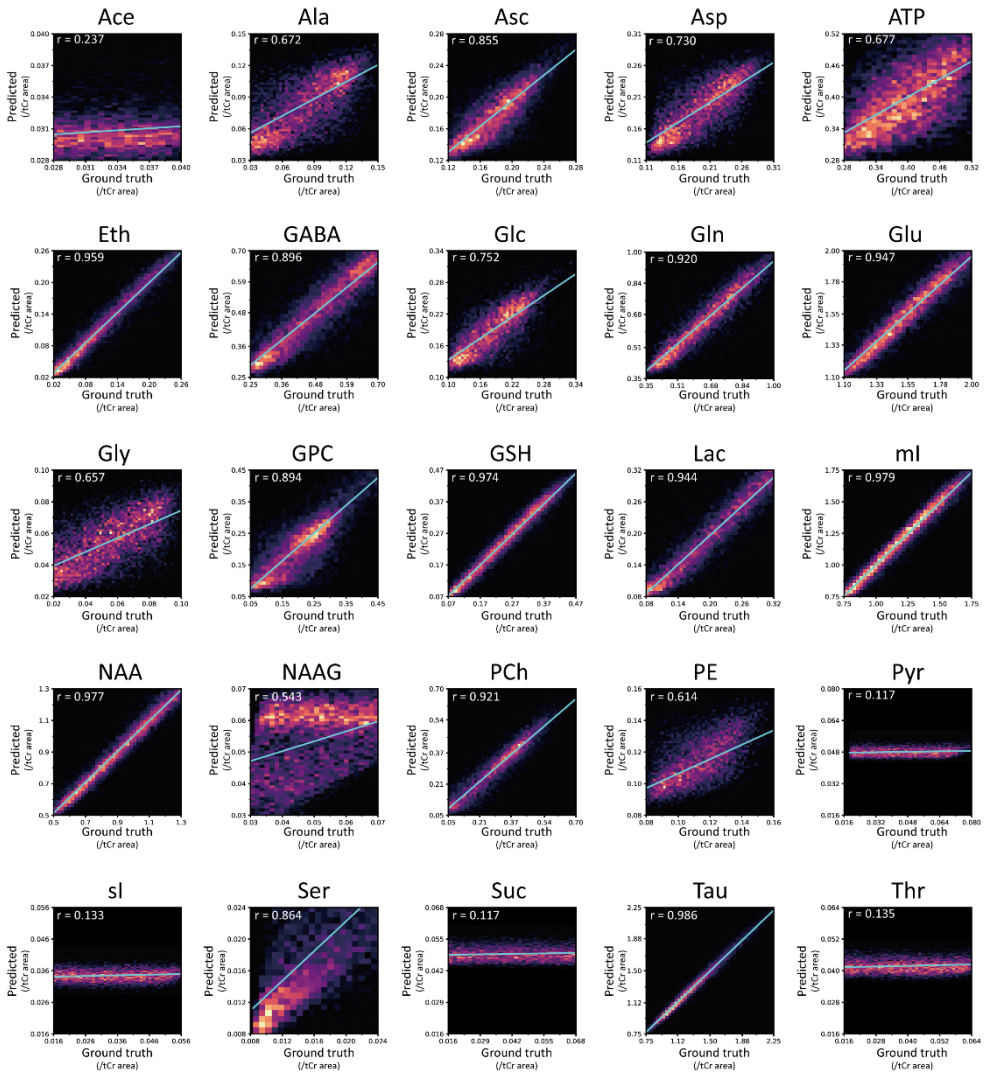


Figure 3-5. The correlations between the GT metabolite concentrations and the CNN-predicted metabolite concentrations for the simulated spectra set I. For the 25 metabolites (out of the 27 metabolites, Cr and PCr were used as the reference (tCr)), correlation (r) ranged from 0.117 (Pyr) to 0.986 (Tau). ($p < 0.001$ for all metabolites) The mean r value was 0.700±0.309. For the 15 major metabolites (Ala, Asp, GABA, Glc, Gln, Glu, GPC, GSH, Lac, mI, NAA, NAAG, PCh, PE, and Tau), correlation ranged from 0.543 (NAAG) to 0.986 (Tau). The mean r value was 0.850±0.147. For Asc, ATP, Gln, Glu, GSH, Lac, mI, NAA, PE, and Tau, which were quantified with the mean absolute percent error (MAPE) less than 10% by the CNNs, the mean r value was 0.887±0.1

Metabolite	Correlation (r)[†]
Ace[*]	0.701
Ala	0.706
Asc[*]	0.738
Asp	0.919
ATP[*]	0.825
Eth[*]	0.62
GABA	0.921
Glc	0.717
Gln	0.942
Glu	0.981
Gly[*]	0.663
GPC	0.75
GSH	0.746
Lac	0.915
mI	0.982
NAA	0.988
NAAG	0.832
PCh	0.962
PE	0.949
Pyr[*]	0.733
Ser[*]	0.829
sI[*]	0.559
Suc[*]	0.596
Tau	0.983

Table 3-3. Correlations between the ground truth error (absolute percent error) and the predicted error by the proposed method for all metabolites.

[*]Those referred to as minor metabolites in the text.

[†] For all metabolites the corresponding P-values were <0.001.

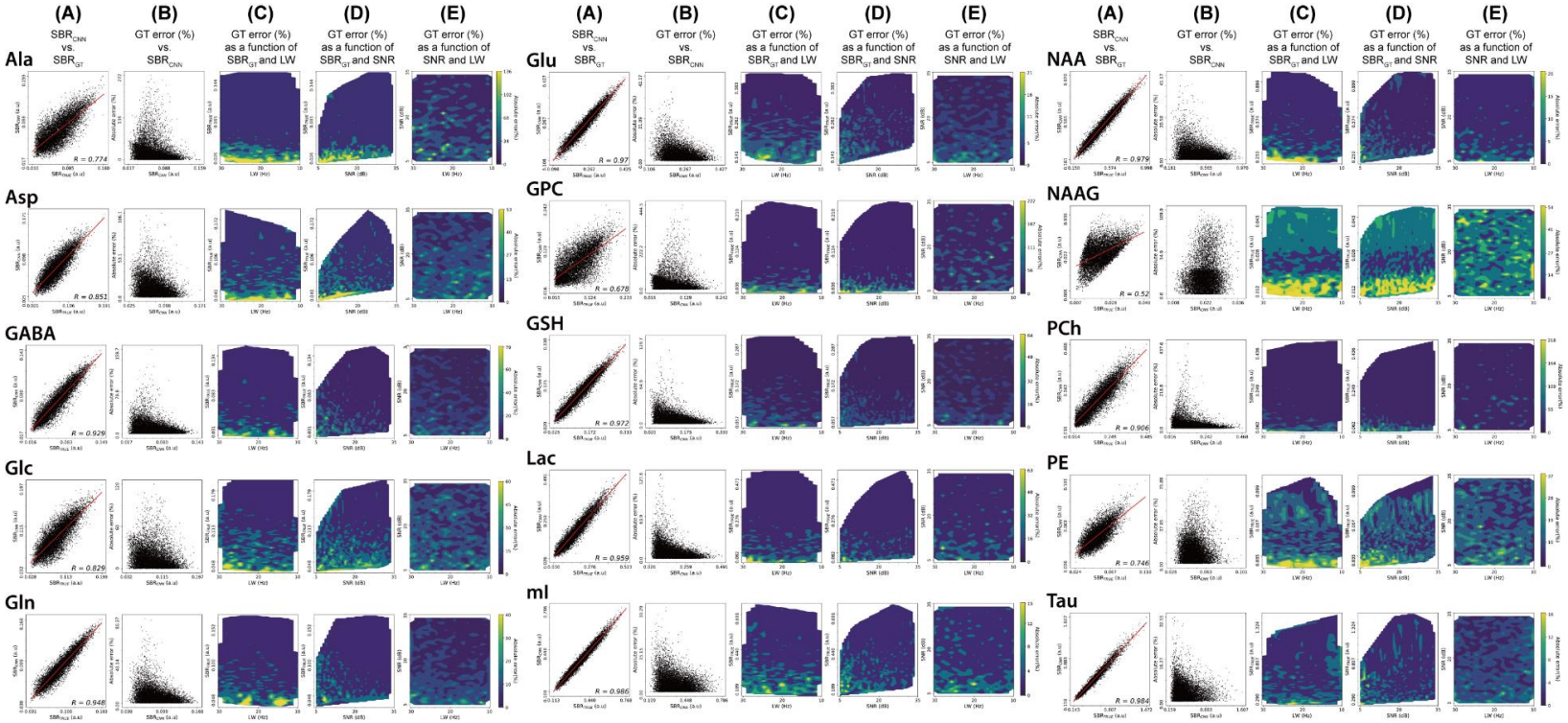


Figure 3-6. The relationships between SBR_{GT} and SBR_{CNN} as well as between the GT errors and SBR_{GT}, SBR_{CNN}, SNR, and linewidth (LW). To minimize the confounding effect among the three variables of SBR_{GT}, SNR, and LW, we provided the contour plots of the GT error (%) as a function of the two of the three variables in (C)-(E) above, instead of scatter plots between the error and each one of the three variables. Note the different scales of the GT (absolute) errors (%) for the contour plots. For all of the correlation coefficients shown in (a), $p < 0.001$.

3.3.2. Performance of Proposed Method, LCModel, and jMRUI on Simulated Spectra Set II

The SNR and linewidth of the simulated spectra set II ranged 5.01-30.09 (17.51 ± 5.79) and 9.95-19.98 Hz (15.02 ± 2.35 Hz), respectively.

The representative spectra are shown in Figure 3-7 where the input spectra came from the same source spectrum (metabolite phantom 2 in Table 3-1) but with different combinations of SNR and linewidth. The corresponding CNN-predicted spectra are also shown. For the majority of the metabolites the CNN-predicted spectra pairs in Figure 3-7 (A)-(B) and in Figure 3-7 (C)-(D) (pairs in terms of the linewidth) are comparable despite the substantial difference in SNR (5 vs. 30).

The MAPE from the proposed method, LCModel, and jMRUI are compared in Figure 3-8 (A). Overall, the proposed method outperforms LCModel and jMRUI. For the proposed method the mean MAPE over the major metabolites were increased from 14.79 ± 11.12 % to 23.07 ± 16.36 %, but Gln, Glu, mI, NAA, and Tau remained as the five metabolites with the lowest MAPE (~10% or less) as compared to the MAPE obtained from the simulated spectra set I shown in Figure 3-4. The correlations between the GT metabolite concentrations and the CNN-predicted metabolite concentrations for the simulated spectra set II are given in Figure 3-9. The correlations between the metabolite concentrations estimated by the proposed method, LCModel, and jMRUI are compared for the individual metabolites in Table 3-4.

The predicted errors by the proposed method and the CRLB from the LCModel and jMRUI are shown with respect to the GT errors (MAPE) in Figure 3-8 (B)-(D), respectively. The predicted errors are inclusive of the GT errors for the majority of the metabolites. The CRLB values tend to be smaller than the GT errors as they are not the predictions of the real errors but the estimations of the minimum possible errors and only those data with CRLB ≤ 50 % were included in the analysis. The correlations of the GT errors with the predicted errors and CRLB from LCModel and jMRUI are summarized for each metabolite in Table 3-5. For the proposed method the correlations were ~ 0.7 or higher (0.78 ± 0.05) and statistically significant for all 15 major metabolites.

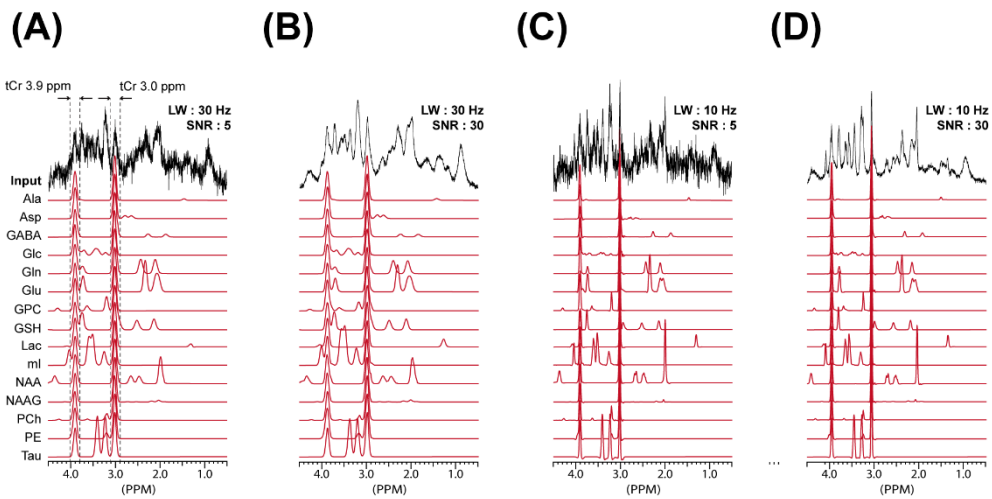


Figure 3-7. The representative rat brain spectra simulated by using the metabolite phantom spectra and in vivo baselines (simulated spectra set II) and the corresponding CNN-predicted spectra for the major metabolites. All CNN-input spectra (Input) have the same relative metabolite concentrations but with different combinations of SNR and linewidth (LW).

	CNN vs. LCM		CNN vs. jMRUI		LCM vs. jMRUI	
	r	P-value	r	P-value	r	P-value
Ala	0.256	0.031	0.531	< 0.001	0.064	0.324
Asp	0.683	< 0.001	0.336	0.007	0.046	0.369
GABA	**0.579**	< 0.001	**0.697**	< 0.001	**0.423**	0.001
Glc	-0.365	0.997	0.278	0.021	0.130	0.175
Gln	0.490	-0.006	-0.006	0.517	-0.197	0.923
Glu	**0.737**	< 0.001	**0.659**	< 0.001	**0.797**	< 0.001
GPC	-0.422	0.999	0.145	0.148	-0.145	0.851
GSH	0.033	0.406	0.505	< 0.001	0.574	< 0.001
Lac	**0.732**	< 0.001	**0.747**	< 0.001	**0.928**	< 0.001
mI	**0.507**	< 0.001	**0.448**	< 0.001	**0.575**	< 0.001
NAA	**0.922**	< 0.001	**0.975**	< 0.001	**0.950**	< 0.001
NAAG	0.504	< 0.001	-0.161	0.878	-0.562	< 0.001
PCh	0.885	< 0.001	-0.239	0.959	0.295	0.015
PE	0.579	< 0.001	0.197	0.077	-0.983	< 0.001
Tau	**0.904**	< 0.001	**0.828**	< 0.001	**0.849**	< 0.001

Table 3-4. The correlations between the metabolite concentrations estimated by the proposed method (CNN), LCModel (LCM), and jMRUI for the simulated spectra set II.

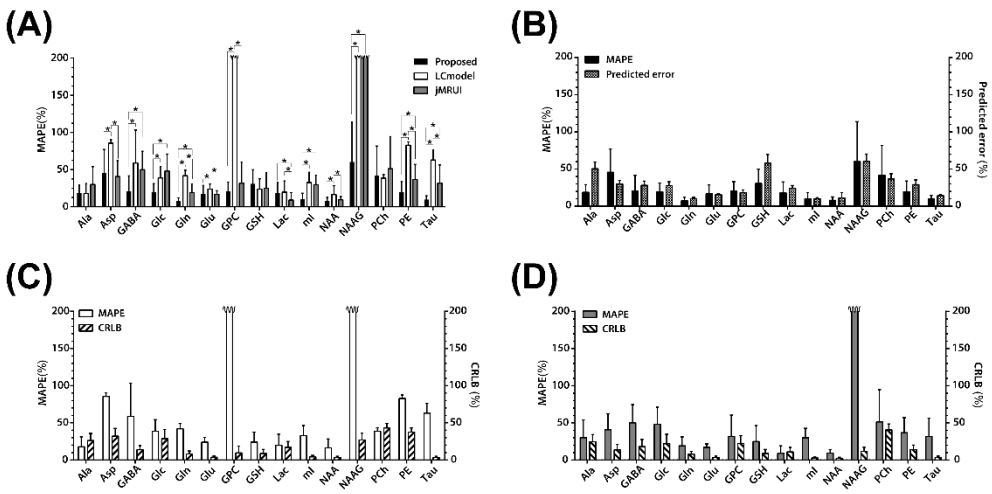


Figure 3-8. The performance of the proposed method, LCModel, and jMRUI on the simulated spectra by using the metabolite phantoms and in vivo baselines (simulated spectra set II; N=54). (A): Comparison of the mean absolute percent errors (MAPE) obtained from the proposed method, LCModel, and jMRUI. * statistically significant difference (Bonferroni corrected). (B): Comparison between the ground truth errors (MAPE) and the predicted errors by the proposed method. (C): Comparison between the ground truth errors (MAPE) and the CRLB from the LCModel. (D): Comparison between the ground truth errors (MAPE) and the CRLB from the jMRUI. In the interpretation of C and D, it should be recalled that CRLB are not predictions of the real errors but estimations of the minimum possible errors and that only those data with CRLB ≤ 50% were included in this analysis.

Metabolites	Proposed			LCM + phantom basis set			jMRUI + phantom basis set		
	Correlation	*P-value*	N	Correlation	*P-value*	N	Correlation	*P-value*	N
Ala	0.702	0.011	54	0.439	< 0.001	49	0.599	< 0.001	50
Asp	0.758	0.004	54	0.404	< 0.001	14	0.541	< 0.001	50
GABA	0.777	0.003	54	0.590	< 0.001	54	0.258	0.070	34
Glc	0.744	0.006	54	0.984	< 0.001	16	0.633	< 0.001	30
Glu	0.831	< 0.001	54	-0.096	0.49	54	0.215	0.065	51
Gln	0.727	0.007	54	-0.229	0.095	54	0.246	0.041	51
GSH	0.785	0.003	54	-0.155	0.512	52	0.039	0.404	42
GPC	0.692	0.013	54	0.584	< 0.001	54	0.255	0.036	51
Lac	0.842	< 0.001	54	0.140	0.188	53	0.607	< 0.001	51
mI	0.802	0.002	54	-0.603	< 0.001	54	-0.220	0.060	51
NAA	0.857	< 0.001	54	-0.203	0.141	54	0.299	0.017	51
NAAG	0.704	0.01	54	0.335	< 0.001	13	-0.236	0.057	46
PCh	0.792	0.002	54	-0.983	< 0.001	3	0.137	0.313	15
PE	0.814	0.001	54	0.715	< 0.001	3	0.090	0.265	51
Tau	0.863	< 0.001	54	0.328	0.016	54	0.540	< 0.001	51

Table 3-5. Correlations of the absolute percent error with the predicted errors by the proposed method (Proposed) and with CRLB obtained from the LCModel (LCM) and jMRUI for the major metabolites.

* Only those cases with CRLB ≤ 50% were included.

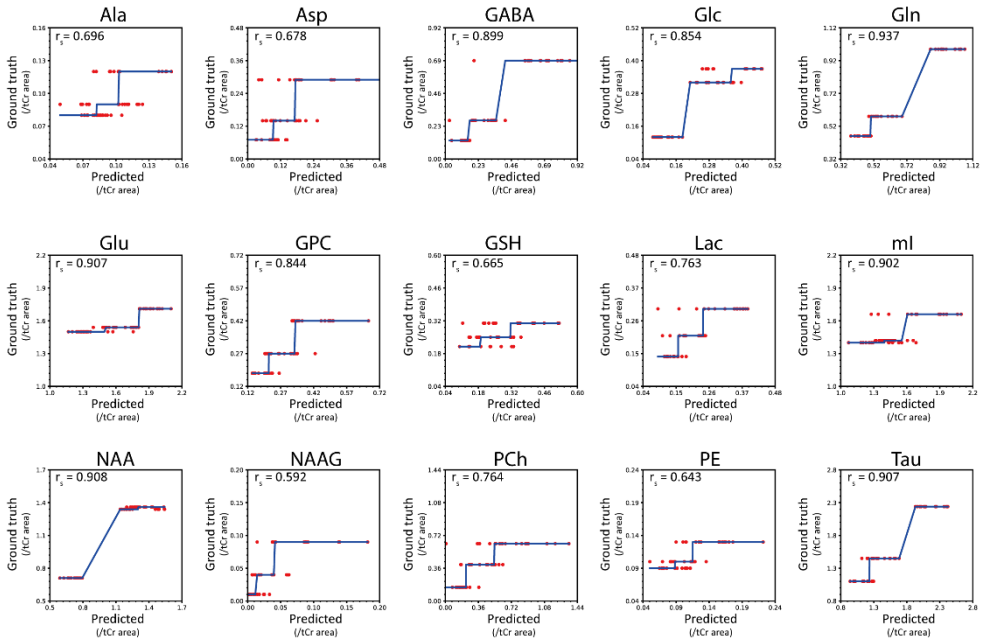


Figure 3-9. The correlations between the GT metabolite concentrations and the CNN-predicted metabolite concentrations for the simulated spectra set II. Given that the ground truth metabolite concentrations of the simulated spectra set II are discrete (3 different concentrations for each metabolite), Spearman's rank correlation coefficient was estimated, r_s. ($p < 0.001$ for all metabolites) The higher r_s indicates higher monotonic association between the ground truth and the CNN-predicted metabolite concentrations. The results are shown below. For the 15 major metabolites (out of the 17 major metabolites, Cr and PCr were used as the reference (tCr)), r_s ranged from 0.592 (NAAG) to 0.937 (Tau). The mean r value was 0.797±0.117. Note that NAAG and Tau showed the lowest and the highest correlations, respectively, also for the simulated spectra set I. For Gln, Glu, mI, NAA, and Tau, which are the five metabolites with the lowest MAPE (~10% or less), the mean r_s value was 0.912±0.014. In summary, for those metabolites with low MAPE values (~10% or less), there are strong correlations between the ground truth metabolite concentrations and the CNN-predicted metabolite concentrations (r being close to 0.9 or higher). These results (Figures 3-5 and 3-6) clearly demonstrate that the CNNs are not just generating one of the values within the range they are trained, but have actually learned the operation of quantification.

3.3.3. Proposed Method Applied to in vivo Spectra

For the spectra with NSA=384 the SNR and linewidth ranged 5.33 – 28.45 (17.81 ± 4.99) and 11.47 – 18.63 Hz (15.32 ± 2.05 Hz), respectively.

The representative in vivo spectra from the cerebellum (A, B) and the thalamus (C, D) of a rat are shown in Figure 3-10 for NSA of 32 (A, C) and 384 (B, D). As in Figure 3-7, the CNN-predicted spectra pairs in Figure 3-10 (A)-(B) and in Figure 3-10 (C)-(D) are comparable for the majority of the metabolites despite the substantial difference in NSA. The quantitative analysis supporting the relatively robust performance of the CNNs against different SNR (NSA of 32 through 384 with a step size of 32) is given in Figure 3-11 in comparison with the LCModel for the major metabolites.

The metabolite concentrations and corresponding predicted errors estimated from the in vivo data (NSA=384) by using the proposed method are shown for the cerebellum and the thalamus in Figure 3-12 (A)-(B), respectively, where Gln, Glu, mI, NAA, and Tau were the five metabolites with the lowest predicted errors (less than 10%) for both brain regions as in Figure 3-8 (B) that was obtained from the simulated spectra set II. Overall, the metabolite concentrations estimated by the proposed method were close to or within the reported ranges (Figure 3-12 (C)). [162]

3.3.4. Processing Time

It took 6.9 sec to quantify the 15 major metabolites (normalized to tCr) from a pre-processed in vivo spectrum on a PC using CPU only (Intel® Core™ i7-7700k processor (4.50 GHz); 48 GB RAM). The runtime includes loading of the files (input spectrum and metabolite-specific SLS spaces and CNNs), prediction of the metabolite spectra by the CNNs, normalization of the peak areas to tCr, and prediction of the quantitative errors.

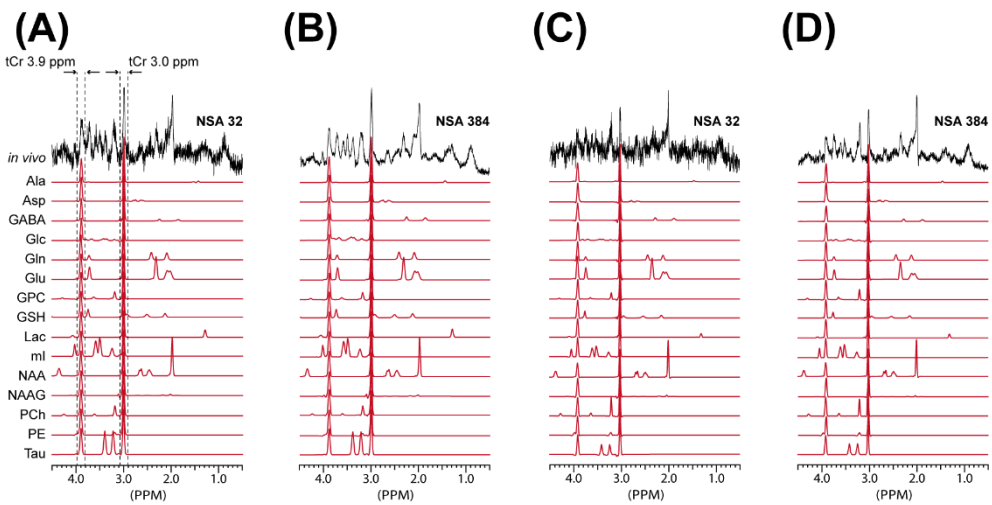


Figure 3-10. The representative in vivo rat brain spectra from the cerebellum ((A), (B)) and the thalamus ((C), (D)) of a rat and the corresponding CNN-predicted spectra for the major metabolites. The spectra A and C with NSA=32 were included in the reconstruction of the spectra B and D with NSA=384, respectively.

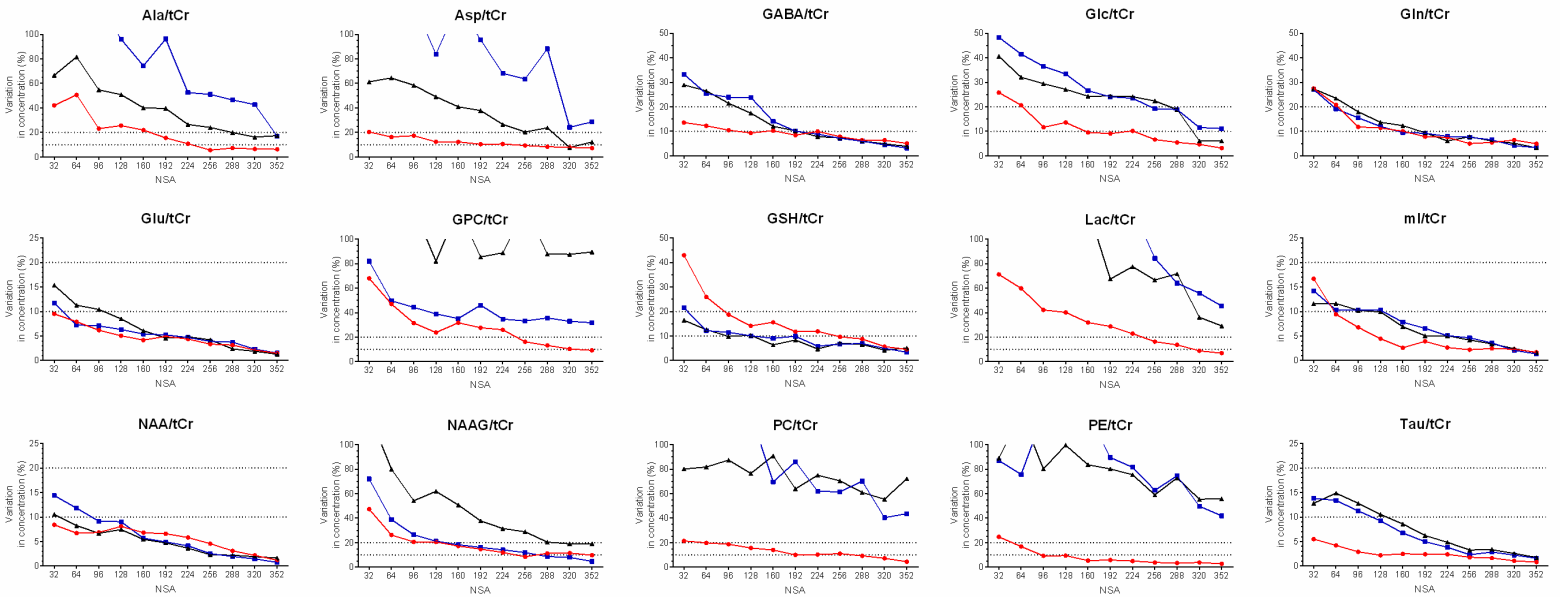


Figure 3-11. The performance of the proposed deep learning-based method against different SNR of the in vivo rat brain spectra. The variations in metabolite concentrations estimated by the CNNs from the in vivo spectra are shown for the major metabolites over the NSA ranging 32 through 352 with a step size of 32 (red line on circles). The variation in metabolite concentrations at each NSA was calculated with respect to the metabolite concentrations estimated at the maximum NSA of 384. The data from the cerebellum and the thalamus were combined in this analysis. The results obtained by the LCModel are also shown. The LCModel analysis was performed with (black line on triangles) and without (blue line on squares) the baseline line functions. Overall, the performance of the proposed method appears more robust than that of the LCModel against different SNR of the spectra. Note the different y-axis (variation in concentration (%)) scales.

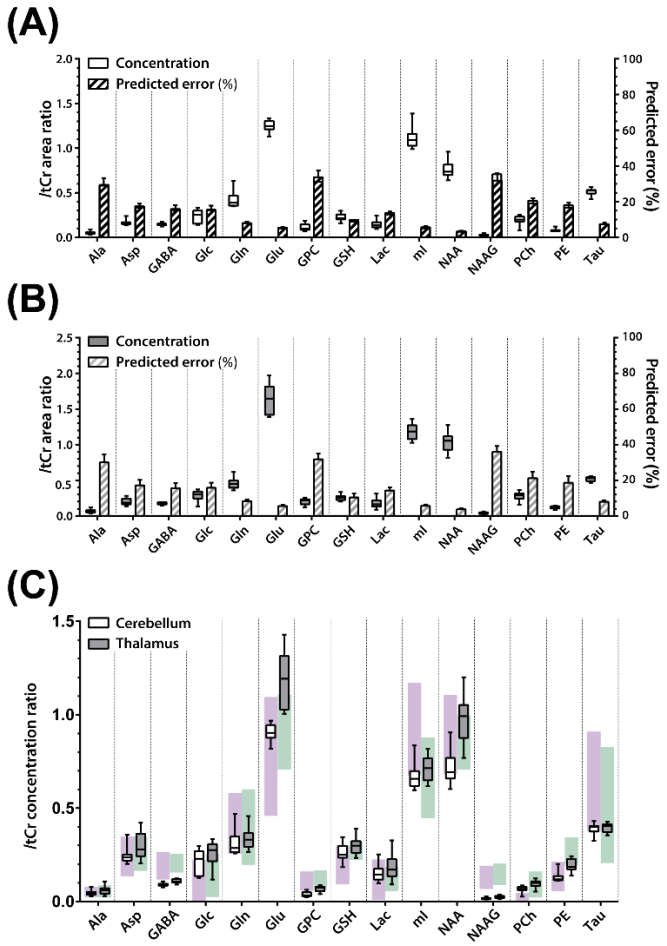


Figure 3-12. The concentrations and predicted errors of the major metabolites estimated from the in vivo data (NSA=384) by using the proposed method. A and B: The metabolite concentrations (box-and-whisker plot; left y-axis) and the corresponding predicted errors in % (shaded bar graph; right y-axis) for the cerebellum (A) and the thalamus (B) of the rats. The concentrations were estimated by measuring the signal area of the metabolite normalized to the signal area of the methyl group of tCr in the CNN-predicted spectra. The vertical bounds of the box (for the concentrations) indicate the lower and upper quartiles, and the line inside the box indicate the median value. The whiskers indicate the minimum and maximum values. The height of the shaded bar and the vertical line extending therefrom (for the predicted errors) indicate the mean and standard deviation, respectively. C: The comparison of the CNN-predicted metabolite concentrations to the reported concentration ranges. The reported concentration ranges of the metabolites are marked as bands (the left pink band for the cerebellum and the right green band for the thalamus for each metabolite). Note that those reported metabolite concentrations were estimated by using the LCModel and thus normalized to the concentrations of tCr in C unlike those concentrations estimated by the CNNs and thus normalized to the signal area of the methyl group of tCr in A and B. For this reason, those concentrations shown in A and B were converted by using the LCModel GT to CNN GT ratio in Table 3-1 such that they can be directly comparable to the reported concentrations. This explains the use of "concentrations (/tCr concentrations)" as the label of the y-axis in C instead of "concentrations (/tCr area)" as in A and B.

3.4. Discussions

3.4.1. Summary of the Study

The recent development of deep learning is revolutionary and its application to ^{1}H-MRS is rapidly growing. To translate the current development in deep learning-based ^{1}H-MRS ultimately into clinical applications, the assessment of measurement uncertainty is an important prerequisite. [108]

The SNR, linewidth, and degree of spectral overlap are known to be the major factors that influence the NLSF-based metabolite quantification, and this appears also to be the case with the deep learning-based ^{1}H-MRS. [108, 110, 111] For instance, the SNR-dependent performance of the CNNs was reported from all of the previous deep learning-based metabolite quantification studies. [108, 110, 111] In addition, as described in Chapter 2, those metabolites that were difficult to quantify using the LCModel also resulted in relatively large errors in the CNN-based metabolite quantification.

To this end, we first defined SBR as a quantitative metric of the degree of spectral overlap, which requires metabolite-specific measurement unlike SNR and linewidth. Next, inspired by the previous studies [111], we developed CNNs that are capable of isolating signals only from a specific target metabolite and a reference metabolite while suppressing the rest of the signals and spectral noise. Using this approach, first of all, the quantification of metabolites is achieved by simply measuring the area of target signals relative to that of the reference signals in the CNN-output spectrum, which is automated in our study. It also greatly simplifies the metabolite-specific SBR measurement from the CNN-output. Thus, the current study is clearly distinguished from the Chapter 2 in which the CNN-output contains signal from all metabolites and the quantification of the individual metabolites is achieved by multiple regression using a metabolite basis set. The narrowed linewidth in the output spectrum therein also makes it difficult to directly measure the metabolite SBR.

The SBR thus measured, however, is subject to errors due to the limited performance of the CNNs. Therefore, the next step was to adjust the SBR empirically by exploiting the big spectral data. In this process the relationships between the CNN-predicted SBR and GT SBR were investigated for each metabolite from the

80,000 simulated rat brain spectra, and stored in a 3-D database where the SNR, linewidth and SBR ranges were segmented heuristically such that the predicted error is large enough to include the GT error while maintaining the high correlation between the two errors. Finally, a group of candidate quantitative errors are obtained from a pre-defined 3-D error space for each metabolite, among which the maximum error was chosen as the predicted error.

3.4.2. Performance of Proposed Method on Simulated Spectra

For the simulated spectra set I, the MAPE was 14.79 ± 11.12% and the correlation between the GT errors and predicted errors was 0.88±0.09 over the all major metabolites. For the simulated spectra set II, the overall MAPE was elevated to 23.07 ± 16.36% and the correlation between the GT errors and predicted errors was lowered to 0.78 ± 0.05 despite the comparable SNR (17.51 ± 5.79) and narrow linewidth (15.02 ± 2.35 Hz) of the spectra relative to those of the simulated spectra set I (SNR=17.51 ± 7.13 and linewidth = 19.97 ± 5.80 Hz).

The use of the simulated basis set and resulting potential difference in lineshape between the simulated basis spectra and phantom spectra of the metabolites could at least in part be responsible for the different performance of the CNNs. The possibility of the presence of unwanted signals such as residual water signal and/or artifact in the metabolite phantom spectra and/or in vivo baseline may not completely be excluded, given the difficulty of perfectly identifying those unwanted signals using visual inspection. This result may reemphasize the importance of the spectral basis set and thorough screening and modeling of unwanted signals in the preparation of a training data set in machine learning-/deep learning-based ^{1}H-MRS. [128, 129, 142]

Nonetheless, for the simulated spectra set II, the overall quantitative accuracy of the proposed method was higher than those of the LCModel and jMRUI, and the predicted errors were inclusive of, and highly correlated with, the GT errors for the majority of the major metabolites. These observations clearly support the potential applicability of the proposed method in noninvasive metabolomics of the rat brain at 9.4T.

3.4.3. Proposed Method Applied to in vivo Spectra

Using the proposed method Gln, Glu, mI, NAA, and Tau were quantified with the predicted errors less than 10%. They were also the five metabolites with the lowest MAPE (~10% or less) using the simulated spectra set II as well. Given that the SNR and linewidth of the in vivo data were comparable to those of the simulated spectra set II, that the predicted errors were inclusive of the GT errors for most of the metabolites using the simulated spectra set II, and that the CNNs were trained on the wide ranges of the relative metabolite concentration ratios, the actual quantitative errors with the in vivo data might likely be smaller than the predicted errors.

While the use of our trained CNNs requires data pre-processing such as Fourier transform, phase and frequency correction, and removal of residual water signal, these are routinely practiced minor tasks. Once a pre-processed spectrum is prepared and fed into the CNNs 15 major metabolites can automatically be quantified along with the corresponding quantitative errors in ~7 sec on CPU only. Therefore, upon the development of an automatic pre-processing unit, fully automated metabolite quantification with error estimation could be implemented online without the need of high-performance GPUs.

3.4.4. Robustness of CNNs against Different SNR

One of the great advantages with deep learning-based ^{1}H-MRS is that the CNN-output can flexibly be controlled by a suitable choice of training labels. Combined with an effectively designed and optimized CNN along with a large enough training sample size, therefore, a CNN can be made capable of spectral denoising as previously demonstrated in Chapter 2. In this study the training input contained those spectra with SNR as low as 5 – 6, while the training labels were almost free from noise. As a consequence of this training input/label setup, the CNN-predicted metabolite concentrations were comparable despite the substantial difference in SNR (Figure 3-7) and across different NSA (Figure 3-11). Such denoising capability of CNNs would allow for exploiting far smaller voxels for minimized partial volume effect or substantially reduced scan time, and thus facilitate the clinical applicability of ^{1}H-MRS.

3.4.5. CRLB and Predicted Error

CRLB has been widely used in the NLSF-based metabolite quantification as a measure of the reliability of fitting. CRLB is of great help in the screening of the fitted data prior to statistical analysis upon a suitable choice of its cutoff value. [123] In the comparison between the predicted errors from the proposed method and the CRLB from LCModel and jMRUI, CRLB tended to be smaller than the GT errors. However, in general, CRLB values are not necessarily larger than measurement errors (%) as they are estimations of the minimum possible errors (i.e., not predictions of the real errors but of the level of uncertainty). Furthermore, for the large errors the CRLB values can naturally be smaller than the measurement errors (%) in our study as we included only those data with CRLB≤50% in the analysis (i.e., max CRLB = 50%). The relatively low correlations of the CRLB values with the GT errors could also be understood in the similar context (i.e., the definition and restricted dynamic range of CRLB). Therefore, while the relatively small CRLB values (GPC and NAAG with LCModel and NAAG with jMRUI; Figure 3-8) and the statistically significant negative correlations (mI and PCh with LCModel; Table 3-5) for some of the metabolites need further speculation, the relationships between the measurement errors and the CRLB values in our study should be interpreted carefully.

It is known that CRLB values are indicative of the precision of the fitted outcome but not the accuracy. [163] On the other hand, the proposed error prediction, although heuristic and empirical, is based on the GT errors and thus indicative of the accuracy of the CNN-predicted metabolite concentrations as demonstrated by the high correlations between the predicted errors and the GT errors, and by the ranges of the predicted errors that are inclusive of the GT errors. In our heuristic approach the predicted error ranges can be adjusted wider to make sure their complete inclusion of the GT error ranges for all metabolites. But then, it may reduce the correlations between the predicted and the GT errors. For a well-trained CNN the measurement errors and thus the candidate errors would be smaller, and therefore the predicted error ranges would become narrower accordingly. Of course, a larger amount of data would further improve the accuracy of the predicted errors in this heuristic and empirical approach. It may also simplify the optimization procedure for the prediction of the errors by taking advantage of more evenly distributed data

points along the SBR-axis in the SLS space. Further, it may allow reduction of the volume of the bins followed by SBR correction for the whole data points inside a bin using a calibration function rather than using the current one-to-one correction (Figure 3-3(C)).

3.4.6. Study Limitation

Approach of the in this study to the assessment of measurement uncertainty is heuristic and developed specific to the case where the SBR of metabolites are used in the assessment. For general applications, therefore, a theory-oriented, formal approach is still required.

In our study we generated the database for the big data-driven error correction by using the training data set for its far larger sample size than the validation data set. To avoid overfitting of the CNNs to the training data set, therefore, they were saved in the middle of training at the epoch where the training and validation losses were comparable. However, ideally, more spectra can be simulated and assigned as an additional validation data set, and the database can be generated therefrom.

In the prediction of the measurement errors the SNR and linewidth of the input spectrum were estimated from tNAA. Therefore, if the concentrations of tNAA vary, the SNR of the spectrum is also measured to be varied even in the case where the effective SNR of all other metabolites remain the same. To minimize potential degradation of the performance of the proposed method in error prediction in this case, those spectra that may well approximate such input spectra should already be stored in the database in our data-driven approach. Thus, our approach is subject to high computational cost for optimization and to demanding large data storage. One of the potential solutions to such an issue may be to improve further the performance of the CNNs particularly against different SNR and linewidth. This study showed the feasibility of developing CNNs that are relatively robust against different SNR (e.g., Figures 3-7 and 3-10), and also demonstrated in Chapter 2, the feasibility of developing CNNs that are relatively robust against different linewidth as well. As the performance of the CNNs against different SNR and linewidth improves, the dependence of the proposed method on the internal reference peak such as tNAA would be reduced. Note that the SBR is naturally dependent on linewidth. Note also

that the SBR defined in our study is also dependent on SNR, given that the signal of a metabolite is directly obtained from the denoised, CNN-predicted spectrum and the background is estimated by subtracting the denoised spectrum from the input spectrum that still contains noise. Taken these altogether, the prediction of the measurement error might also be achieved by solely relying on the SBR without referring to an internal reference upon the improved performance of the CNNs.

This study was trained the CNNs to generate spectrally isolated target signal as an output for each individual metabolites. However, upon the availability of hardware with higher performance, a single CNN may be developed that is capable of generating the output spectra for all metabolites at a time or directly providing their concentrations altogether, while maintaining the quantitative accuracy achieved with the current approach. The preprocessing required by our CNNs could also be incorporated into the development of such a single CNN including a far more tolerable phase shift range in particular. It is interesting to note that the individual CNNs appear to utilize almost whole spectral region for the generation of the output spectrum despite the fact that the output spectrum has multiple spectral regions that are completely devoid of signal (Figure 3-13, Figure 3-14 and Figure 3-15).

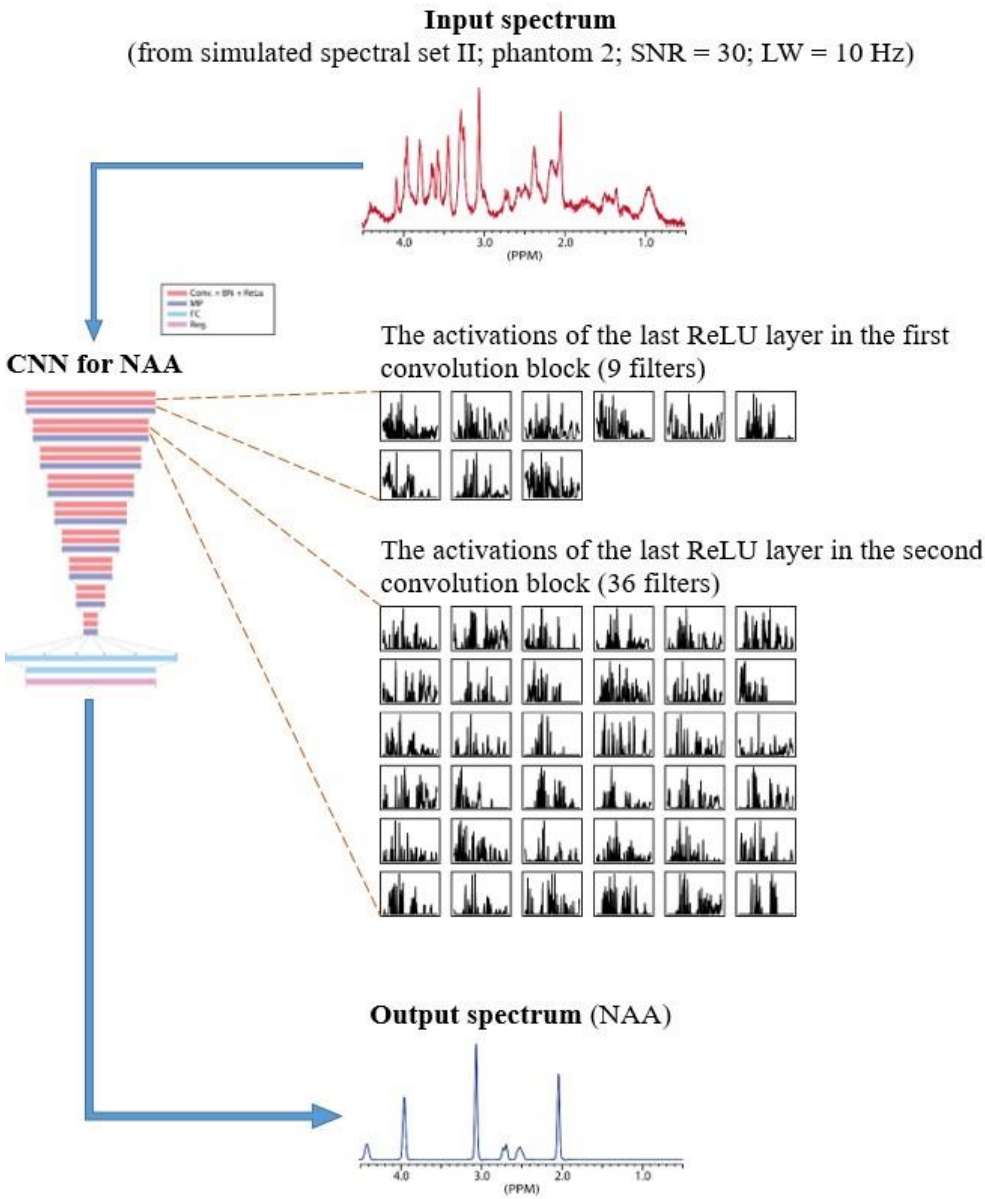


Figure 3-13. The activations of the CNN trained for NAA in response to a simulated spectrum. One spectrum from the simulated spectra set II was used as an input spectrum. Figure shows the activations of the last ReLU layers in the first and the second convolution blocks of the CNN for NAA. The output spectrum contains signals from NAA at ~ 2.0, 2.6, and 4.4 ppm as well as from the tCr as a reference at ~ 3.0 and 3.9 ppm. In particular, there is no signal at all in the 0.5-2.0 ppm range. Nonetheless, for the majority of the filters, the ReLU layers activate across the whole spectral regions in addition to those regions where the signals from NAA and tCr are present.

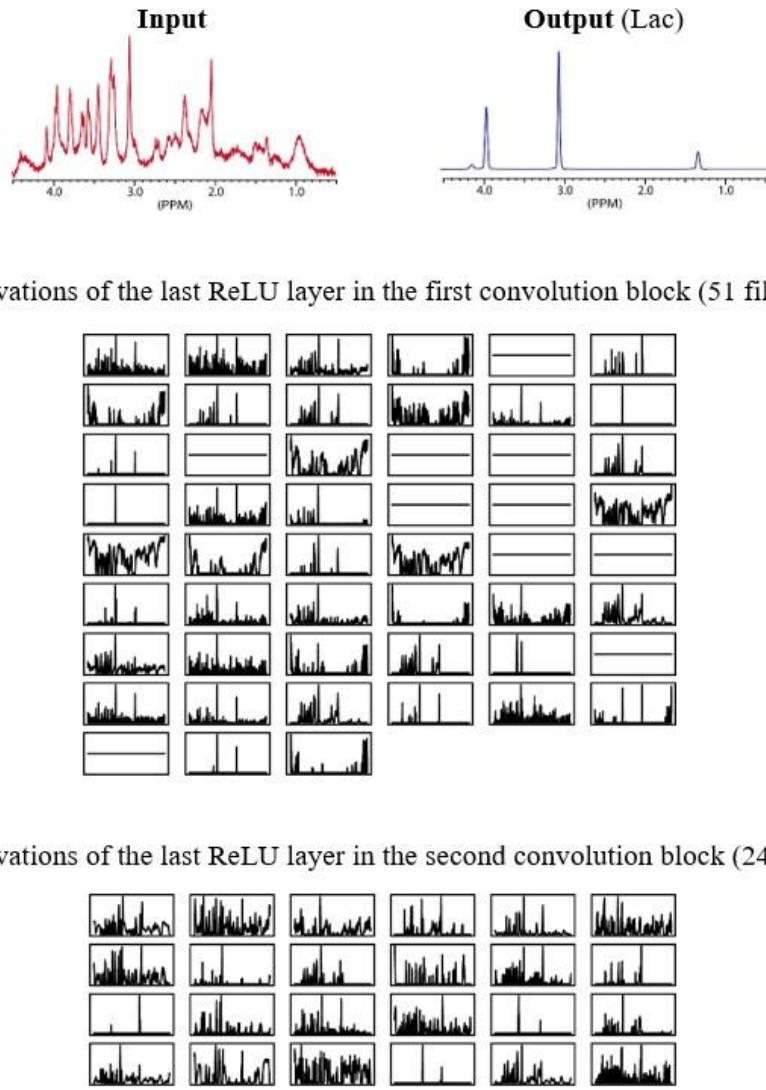
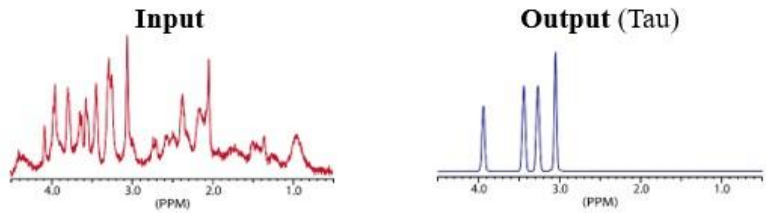
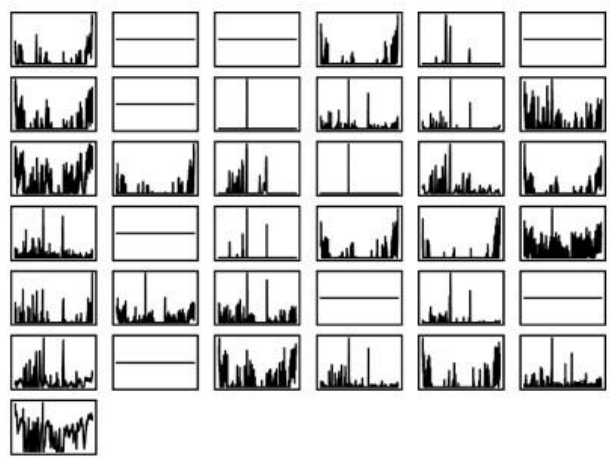


Figure 3-14. The activations of the CNN trained for Lac in response to a simulated spectrum. Figure shows the activations of the last ReLU layers in the first and the second convolution blocks of the CNN for Lac which has signals with the largest separation (~ 1.3 and 4.1 ppm) among the metabolites with a coupled spin system. The same input spectrum that was used for NAA above was also used. In the output spectrum, there is no signal in the 1.5-3.0 ppm range. Nonetheless, for many of the filters, the ReLU layers activate even in the 1.5-3.0 ppm range as well.



The activations of the last ReLU layer in the first convolution block (37 filters)



The activations of the last ReLU layer in the second convolution block (16 filters)

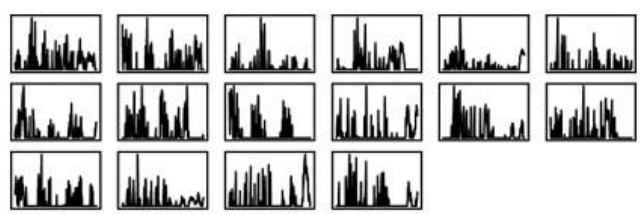


Figure 3-15. The activations of the CNN trained for Tau in response to a simulated spectrum. The figure shows the activations of the ReLU layers in the first and the second convolution blocks of the CNN for Tau which has signals with the smallest separation (~ 3.3 and 3.4 ppm) among the metabolites with a coupled spin system. The same input spectrum that was used for NAA and Lac above was also used. In the output spectrum, there is no signal in the 0.5-3.0 ppm range. Nonetheless, for many of the filters, the ReLU layers activate even in the 0.5-3.0 ppm.

Chapter 4. Bayesian deep learning-based proton magnetic resonance spectroscopy of the brain: metabolite quantification with uncertainty estimation using Monte Carlo dropout

4.1. Introduction

As explained in the previous chapters, quantification of the individual brain metabolites by using ^{1}H-MRS at a clinical field strength can be challenging due to limited SNR, spectral overlap not only between metabolites but also between metabolites and MMs, and the presence of unknown spectral baseline. [36, 55] To address this issue, nonlinear least squares fitting (NLSF) is widely used [55, 116], and yet development of a more robust method is a remaining important research topic. [126, 155]

Following the machine learning study [107], deep learning-based ^{1}H-MRS (DL-MRS) has shown its potential as an alternative to the NLSF approaches. [108, 110, 111, 164] However, these previous studies used standard CNNs that do not provide uncertainty in the quantitative outcome. [94, 165] In the training of standard CNNs, the weights are directly optimized by minimizing the objective function. [74, 90] The set of the optimized weights are then used for inference where a single, best guess output is obtained but with no information as to how certain the prediction would be. [166] To be translated to clinical applications, deep learning-based methods should also provide how much the network is confident about its predicted outcome. In the case of the NLSF approaches, the CRLB have been used as representing lower bounds of the fit error (uncertainty). [123, 167]

In contrast, Bayesian neural networks (BNNs) provide a probabilistic interpretation of deep learning models by inferring distributions over the models' weights. [165, 168] That is, a trained BNN can be described in terms of the (posterior) probability distribution of weights instead of a single set of optimized, deterministic

The work presented in this chapter has been submitted to Magnetic Resonance in Medicine and is currently under review. Title: Bayesian deep learning-based 1H-MRS of the brain: metabolite quantification with uncertainty estimation using Monte Carlo dropout

weights in the standard CNNs. [165, 168] The distribution of weights results in a distribution of network outputs and thus provides information about the uncertainty therein. [166] The Bayesian approach requires highly complex computations and thus BNNs are difficult to generalize. [166, 168] However, recent studies showed that a CNN trained with dropout can be interpreted as a BNN and that such a Bayesian convolutional neural network (BCNN) in combination with Monte Carlo dropout (MCDO) sampling can provide uncertainty in the predicted outcome. [98, 124, 165]

This chapter deals with the applicability of the BCNN with MCDO sampling in ^{1}H-MRS of human brain at 3.0 T. Using simulated spectra, a BCNN was trained to generate a noise-free, line-narrowed, MMs signal-removed, metabolite-only spectrum from a human brain spectrum that is typically degraded with limited SNR, line-broadening, MMs-signal, and unknown baseline. (described in Chapter 2) Both metabolite content and corresponding uncertainty are quantified from MCDO sampled spectra. The method was tested first on the simulated spectra and further on modified in vivo spectra. For the latter, the BCNN-predicted results were compared with the metabolite content and CRLB from the linear combination of model spectra (LCModel) analysis [55] which is the current gold-standard in the NLSF approaches.

4.2. Methods and Materials

4.2.1. Theory

Outline of Bayesian approach

In the Bayesian approach, one seeks the most probable functions for a given data set consisting of input $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ and output $\mathbf{Y} = \{y_1, y_2, \ldots, y_N\}$. In BNN, the functions are defined in terms of network weights w, and one seeks a distribution of weights.[94, 168] Initially, a prior distribution of weights, p(w), such as a Gaussian distribution, is assumed based on one's prior knowledge before observing any data. As one observes data, p(w) is optimized towards the most probable weight distribution for the given data set, namely, a posterior distribution, $\mathbf{p(w|X, Y)}$, upon the knowledge about the probability of the output being generated from the input for a set of w, namely, a likelihood distribution, $\mathrm{p}(\mathbf{Y}|\mathbf{X}, \mathbf{w})$, according to the Bayes' theorem given as

$$\mathbf{p}(\mathbf{w}|\mathbf{X},\mathbf{Y}) = \frac{\mathrm{p}(\mathbf{Y}|\mathbf{X},\mathbf{w})\mathrm{p}(\mathbf{w})}{\mathrm{p}(\mathbf{Y}|\mathbf{X})}. \quad (4.1)$$

where $\mathrm{p}(\mathbf{Y}|\mathbf{X}) = \int \mathrm{p}(\mathbf{Y}|\mathbf{X},\mathbf{w})\mathrm{p}(\mathbf{w})\mathrm{d}\mathbf{w}$ is the normalizer known as model evidence. [166, 168] Denoting the approximate variational distribution as q(w), the BNN output for an input x*, is calculated by [165, 168]

$$\mathrm{p}(\boldsymbol{y}^*|\boldsymbol{x}^*,\mathbf{X},\mathbf{Y}) = \int \mathrm{p}(\boldsymbol{y}^*|\boldsymbol{x}^*,\mathbf{w})\mathrm{p}(\mathbf{w}|\mathbf{X},\mathbf{Y})\mathrm{d}\mathbf{w}. \quad (4.2)$$

However, the posterior is computationally intractable and thus modelled using a simple variational distribution. Such approximation of the posterior is called variational inference.[94, 168] If the approximate variance distribution as q(**w**), then the BNN output for the inputs $\mathbf{x}^*$, $\mathrm{p}(\boldsymbol{y}^*|\boldsymbol{x}^*,\mathbf{X},\mathbf{Y})$ can be approximated as [94]

$$\mathrm{p}(\boldsymbol{y}^*|\boldsymbol{x}^*,\mathbf{X},\mathbf{Y}) \simeq \int \mathrm{p}(\boldsymbol{y}^*|\boldsymbol{x}^*,\mathbf{w})\mathrm{q}(\mathbf{w})\mathrm{d}\mathbf{w} \simeq \frac{1}{T}\sum_{t=1}^{T}\mathrm{p}\left(\boldsymbol{y}^*|\boldsymbol{x}^*,\mathbf{w}_t^{MC}\right). \quad (4.3)$$

where the integration in the second term in the Equation (4.3) was approximated with Monte Carlo integration in the third term and $\mathbf{w}_t^{MC}$ denotes the weights sampled from q(**w**) at the *t*-th draw. Recently, it was shown that sampling weights from $\mathrm{q}(\mathbf{w}_i)$ is equivalent to performing a dropout operation on the i-th layer of a neural network, which randomly sets $\mathrm{p}_{\mathrm{dropout}}$ proportion of the elements of the layer to zero ($\mathrm{p}_{\mathrm{dropout}}$: dropout probability) [124].

Given this interpretation of dropout as performing approximate variational inference [124, 168], the last term in Equation (4.3) is viewed as averaging the stochastic forward pass through a neural network with dropout at test time, which is called Monte Carlo dropout (MCDO) sampling. [94] In practice, MCDO sampling is performed first by training a neural network with dropout before every weight layer and then by performing dropout at test time as well. [124] It was further shown [168] that optimization of any neural network with dropout is equivalent to variational inference and that the optimized weights in a neural network with dropout can be viewed as the optimized approximate variational distribution in a BNN with the same network architecture as the neural network. Therefore, a CNN trained with dropout can be treated as a BCNN. [124, 168]

Classification of uncertainty

In BCNN, one deals with a distribution of network weights rather than a

deterministic single set of weights in a standard CNN. The uncertainty in weights results in the uncertainty in the model output. Such uncertainty resulting from the uncertainty in the model parameters is called epistemic (model) uncertainty. [124] In addition, noise inherent in the observation can also result in uncertainty in the network output. [124] Such uncertainty is called aleatoric (data) uncertainty and further classified into homoscedastic and heteroscedastic aleatoric uncertainty depending on the variability of the uncertainty for different inputs. [124]

Here, heteroscedastic aleatoric uncertainty is assumed, i.e., different observation noise for different inputs. Unlike aleatoric uncertainty, epistemic uncertainty is known to be reduced as the amount of training data increases [124, 169], and thus associated with the prediction errors in response to the input data that is out of distribution (OOD). [124]

Training and inference

To estimate the uncertainty using the MCDO approach, a BCNN is trained by using the following loss function. [124]

$$L = \frac{1}{D}\sum_{i}\frac{1}{2}\frac{||y_i - \hat{y}_i||^2}{\hat{\sigma}_i^2} + \frac{1}{2}\log\hat{\sigma}_i^2 \tag{4.4}$$

where, for an input $\boldsymbol{x}$, $\hat{y}_i$ is the output from the network, y_i is the corresponding GT, D is the number of elements in the output, and $\hat{\sigma}_i^2$ is the noise variance for output element i that corresponds to the amount of noise inherent in the data and is therefore associated with the aleatoric uncertainty. Note that $\hat{\sigma}_i^2$ is also obtained as a network output in the MCDO approach by making it data-dependent (heteroscedastic) and learned during training in an unsupervised manner. [124]

Once the BCNN is trained, the network output $\hat{\boldsymbol{y}}_t$ and $\hat{\boldsymbol{\sigma}}_t^2$ are sampled for T times. Then, the prediction is achieved by obtaining the predictive mean as

$$\hat{\boldsymbol{y}} \simeq \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{y}}_t \tag{4.5}$$

and the predictive uncertainty of $\hat{\boldsymbol{y}}$ from the variance of $\hat{\boldsymbol{y}}$, Var($\hat{\boldsymbol{y}}$), as:

$$\mathrm{Var}(\hat{\boldsymbol{y}}) \simeq \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{y}}_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{y}}_t\right)^2 + \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\sigma}}_t^2 \tag{4.6}$$

where the epistemic uncertainty is given by the sum of the first two terms and the

aleatoric uncertainty is given by the last term in Equation (4.6). [124] From here on, we refer to Var ($\hat{y}$) as total uncertainty.

Estimation of metabolite content and uncertainty

The outline of estimating metabolite content and uncertainty is shown in Figure 4-1. Our BCNN takes a brain spectrum as an input and is trained to predict a metabolite-only spectrum. After training of the BCNN, MCDO sampled spectra were obtained T times (T-MCDO) per an input spectrum. The predictive mean of the T-MCDO sampled spectra was calculated using Equation (4.5) and assigned as the final, BCNN-predicted spectrum.

The individual metabolite content were estimated from the BCNN-predicted spectrum by multiple regression. That is, $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{S}}\boldsymbol{B}_{pinv}$ where, for 17 metabolites and 1024 spectral data points, $\widehat{\boldsymbol{A}}$ is the estimated metabolite content (1 × 17 vector), $\widehat{\boldsymbol{S}}$ is the BCNN-predicted spectrum (1 × 1024 vector), and $\boldsymbol{B}_{pinv}$ is the pseudo-inverse of the metabolite basis set (1024 × 17 matrix).

To estimate uncertainty in the BCNN-predicted metabolite content, first, we defined a two-standard deviation spectrum (2SD spectrum) as 2 × SD-spectrum where SD-spectrum was obtained by (total uncertainty spectrum)$^{1/2}$ based on Equation (4.6). Then, the uncertainty was estimated from the 2SD spectrum also by the multiple regression. Given that the 2SD spectrum is in absolute mode, the metabolite basis set was used also in absolute mode in this process of uncertainty estimation.

Finally, the BCNN-predicted metabolite content was expressed in 'metabolite content (%uncertainty)' where %uncertainty was obtained by converting the estimated uncertainty into the percentage of the respective metabolite content.

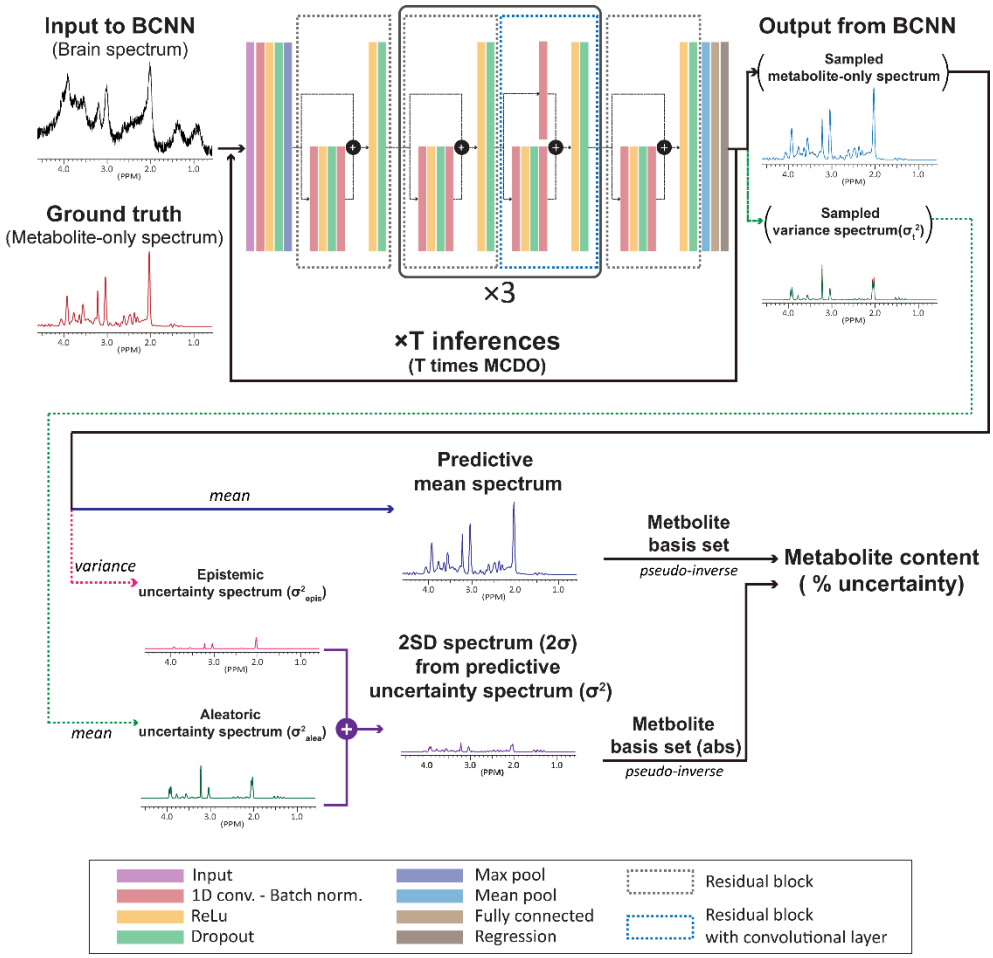


Figure 4-1. The simplified design of the Bayesian convolutional neural network (BCNN) and the workflow from training of the network to estimation of metabolite content and associated uncertainty. The input spectra are simulated human brain spectra at 3.0T, which are typically degraded by low SNR, linebroadening, MM signal, unknown baseline, and phase/frequency shift. The ground truth (GT) target spectra are metabolite-only spectra before adjusting SNR, linewidth and phase/frequency and adding the MM signal. The output of the network consists of a metabolite-only spectrum and a noise variance spectrum. A total of 50 output are obtained per an input spectrum by Monte Carlo dropout (50 times MCDO) sampling. The mean of the metabolite-only output spectra (predictive mean spectrum) is assigned as the BCNN-predicted spectrum. The variance of the metabolite-only output spectra is assigned as the epistemic (model) uncertainty spectrum. The mean of the noise variance spectra is assigned as the aleatoric (data) uncertainty spectrum. A 2SD spectrum is obtained from the total (= aleatoric + epistemic = predictive) uncertainty spectrum. Multiple regression is performed on both the BCNN-predicted spectrum and the 2SD spectrum using the metabolite basis set to estimate the individual metabolite content and associated uncertainty, respectively. Finally, the results are expressed as 'metabolite content (%uncertainty)'.

4.2.2. Preparation of Spectra

Simulation of spectra

Human brain spectra were simulated by including 17 metabolites [34, 58, 130, 131] and 9 resonance groups of MMs [132, 133] as previously described in Section 2.2.3. The 17 metabolites were: Ala, Asp, Cr, GABA, Glc, Glu, Gln, GSH, GPC, Lac, mI, NAA, NAAG, PCr, PCh, PE, and Tau. The SNR and linewidth as measured from tNAA at ~2.0 ppm ranged 7 – 20 and 10.2 – 20.6 Hz, respectively. The zeroth-order phase (± 5°) and frequency shift (± 10 Hz) were randomly varied. The real components of the spectra were extracted and cropped to 1024 points (~ 0.5 – 4.5 ppm). A total of 100,000 brain spectra were simulated (N = 80,000 (training), 10,000 (validation), and 10,000 (test)) (simulated data set I).

In addition, to investigate the individual behavior of aleatoric, epistemic, and total uncertainty as a function of SNR, linewidth or metabolite concentrations, three groups of spectra (N = 1000 for each group) were simulated by varying one of those three simulation parameters in turn (simulated data set II). The values of the rest of the simulation parameters were fixed at their mean values in the simulated data set I. For simplicity, only NAA was considered in this analysis, for which the most robust BCNN performance was obtained on the simulated data set I. The ranges of the SNR, linewidth, and concentrations of NAA were 3.5 – 40, 4.9 – 40.3 Hz, and 3.75 – 34 mmol/L, respectively, which were far wider than those in the simulated data set I such that the investigation can be conducted in OOD regions as well.

Modification of in vivo spectra

The proposed method was further tested on modified in vivo data. The original, unmodified, in vivo data were collected previously from the left frontal lobe (2 × 2 × 2 cm^3) of 5 healthy volunteers (all males; age, 24 – 36 years (30 ± 3 years)) using a PRESS [49, 50] sequence at 3.0T (Siemens MAGNETOM Tim Trio (VB17A), 32 channel head coil, TR/TE = 2000/30 ms, spectral bandwidth = 2 kHz, number of data points = 2048, frequency offset = -2.0 ppm, autoshimming; the same sequence used for the acquisition of the metabolite basis spectra in phantom). For each volunteer, eight water-suppressed spectra each with NSA = 8 were collected consecutively from the same voxel and combined into one spectrum (NSA ~ 64). Each of the five resulting spectra were zero-filled to 4096, Fourier transformed, zeroth-order phase-

corrected, and frequency-shifted by using jMRUI (v.5.2). [116] The residual water was removed (HLSVD filter [119]). Then, both SNR and linewidth were simultaneously adjusted to generate 10 groups of modified spectra with increasing severity of spectral degradation. For line-broadening, the same Lorentzian apodization function that was used for the training data set was used. Thus, there were one group of unmodified and 10 groups of modified in vivo spectra, each of which has 5 spectra from the 5 subjects. The spectra in real mode were extracted and cropped to 1024 points as with the simulated spectra.

4.2.3. BCNN

Implementation

A BCNN was implemented using Matlab deep learning toolbox (v.9.7 (R2019b); Mathworks Inc., Natick, MA) based on a ResNet [170] (Figure 4-1). It contained 5 residual blocks and 3 residual blocks with additional convolutional and batch normalization layers. Each residual block consists of one-dimensional convolution, batch normalization, activation (ReLU) and dropout. A dropout layer was placed before every convolutional layer (except for the 1st one) and the fully connected layer (or equivalently, placed immediately after every activation layer in our network design), and rendered to operate at test time as well. To learn heteroscedastic $\hat{\sigma}_i^2$ during training, the number of channels of the fully connected layer prior to the regression layer was doubled and the additional channel was assigned for $\hat{\sigma}_i^2$. [124]

Optimization and training

The BCNN was optimized and trained by using the training and validation sets of the simulated data set I on four graphics processing units (GPUs; NVIDIA TITAN RTX). The root-mean-square propagation (RMSprop) [171] optimizer was used. An L_2 regularizer (weight decay) was included in the loss function. The training set was normalized to a 0 – 255 scale followed by zero-center normalization.

The filter size and number of filters of the convolutional layers were set based on the Chapter 2.2.4**.** The one-dimensional convolution filter sizes were 15 × 1 for the first residual block (stride = 2 × 1) and 7 × 1 for the rest of the residual blocks (stride = 1 × 1). The numbers of filters were 32 and 512 for the first and the last residual blocks, respectively. For the pair of the residual blocks with three

repetitions between the first and the last residual blocks, they were 64, 128 and 256 for each repetition. The initial learning rate and L_2-regularization parameter were 7.524×10^{-4} and 9.393×10^{-5}, respectively, which were Bayesian-optimized **[135]** in the ranges of 1×10^{-5} to 1×10^{-3} and 1×10^{-6} to 1×10^{-3}, respectively. The dropout probability and mini batch size were determined to be 0.2 and 256, respectively, according to the root-mean-square-error (RMSE) between the BCNN-predicted and GT spectra in a simple grid search. The searching points of these two parameters were 0.1, 0.2, and 0.5 (3 points), and 32, 64, 128, 256, and 512 (5 points), respectively. The number of samples (T) in the T-MCDO sampling was optimized on the validation set. Five to 200 MCDO samples were obtained with a step size of 5 for each BCNN input. An optimal T of 50 was determined according to the overall MAPE of the 17 metabolite content.

Using the optimized hyperparameters, the BCNN was trained with the maximum epoch of 2000 and an early stopping rule (50 epoch patience). The learning rate was scheduled to drop at every 11th epoch by a factor of 0.9. The optimization and training of the BCNN took ~96 and ~18 hours, respectively.

4.2.4. Evaluation of Proposed Method

Evaluation on simulated spectra

The performance of the BCNN in the estimation of the metabolite content and associated uncertainty was evaluated on the test set of the simulated data set I.

The results of the metabolite quantification were presented in term of MAPE for individual metabolites (MAPE = $\frac{1}{\mathrm{N_{test}}}\sum_{n=1}^{\mathrm{N_{test}}}|GT_n - estimated_n|/GT_n \times 100$; $\mathrm{N_{test}}$ = 10,000).

In comparison between two spectra (e.g., GT spectrum vs. BCNN-predicted spectrum), the resulting residual spectrum is shown and the amount of the residual signal therein is presented in terms of the MSE in percent (%MSE = [mean square of (GT – predicted)]/[mean square of (GT)] × 100).

Uncertainty as a function of SNR, linewidth or metabolite concentrations

Using the BCNN trained on the simulated data set I, the BCNN-predicted (metabolite-only) spectra and the aleatoric, epistemic, and total uncertainty spectra

were obtained for the simulated data set II. Then, the areas of the NAA signal at ~2.0 ppm were measured from the three types of the uncertainty spectra and normalized to that from the BCNN-predicted (metabolite-only) spectra. Finally, the changes of the three types of the uncertainty were examined as a function of the SNR, linewidth or concentration of NAA.

Evaluation on modified in vivo spectra

Using the unmodified in vivo spectra as a reference, the robustness of the BCNN against varying SNR and linewidth was tested on the 10 groups of the modified in vivo spectra. The results from the BCNN were compared with the metabolite content and CRLB from the LCModel analysis (v.6.3-1J) **[55]** where the vendor-provided metabolite basis set was used. The analysis was performed for the 6 representative metabolites. That is, Glx, mI, tCho, total tCr, and tNAA were included as they are most frequently presented from LCModel analysis. GABA was also included for its well-known difficulty of quantification. The estimated metabolite content was normalized by that of tNAA. Those cases where metabolites are quantified with null amplitude and 999% CRLB were excluded in the LCModel analysis.

4.2.4. Statistical Analysis

Statistical analyses were performed using PRISM (v.6.01; GraphPad Software, Inc., La Jolla, CA, USA). For a pair-wise group comparison, a Student t-test was performed. For correlations between two variables, the Pearson's correlation coefficient was calculated. A p-value of less than 0.05 was considered statistically significant.

4.3. Results

4.3.1. Metabolite Content and Uncertainty Estimation on the Simulated Spectra

Representative spectra for metabolite quantification

Figure 4-2 shows the representative simulated brain spectra used as BCNN inputs (A-D) and each BCNN prediction spectrum (I-L). BCNN mapped the simulated brain spectra (A-D) into noise-free, line-narrowed, baseline-removed, phase/frequency shift-corrected metabolite-only spectra (I-L). The denosing, line-

narrowing and baseline removal functions of BCNN are clearly shown in Figure 4-2 (I-L). The GT (E-H), reconstructed (using the estimated metabolite content and the metabolite basis set) (M-P), residual spectra ([GT – predicted] (Q-T) and [predicted – reconstructed] (U-X)) are also shown. The residual signal is negligible ([GT – predicted] (Q-T)). The predicted spectra are almost fully accounted for by a linear combination of the metabolite basis set as demonstrated in the residual spectra ([predicted – reconstructed] (U-X).

Representative spectra for uncertainty estimation

Figure 4-3 (E)-(H) show the total (aleatoric + epistemic) uncertainty spectra for the BCNN-input spectra (A-D) which are the same spectra that were shown in Figure 4-2(A)-(D). The corresponding aleatoric ((I)-(L)) and epistemic ((M)-(P)) uncertainty spectra are also shown. The 2SD spectra are shown in Figure 4-3 (Q)-(T). The reconstructed 2SD spectra ((U)-(X)) are almost fully accounted for by the linear combination of the metabolite basis set (Figure 4-3 (Y)-(B')).

The mean %MSE between the 2SD spectra and reconstructed 2SD spectra for the test set of the simulated data set I was 1.54 ± 0.38 %.

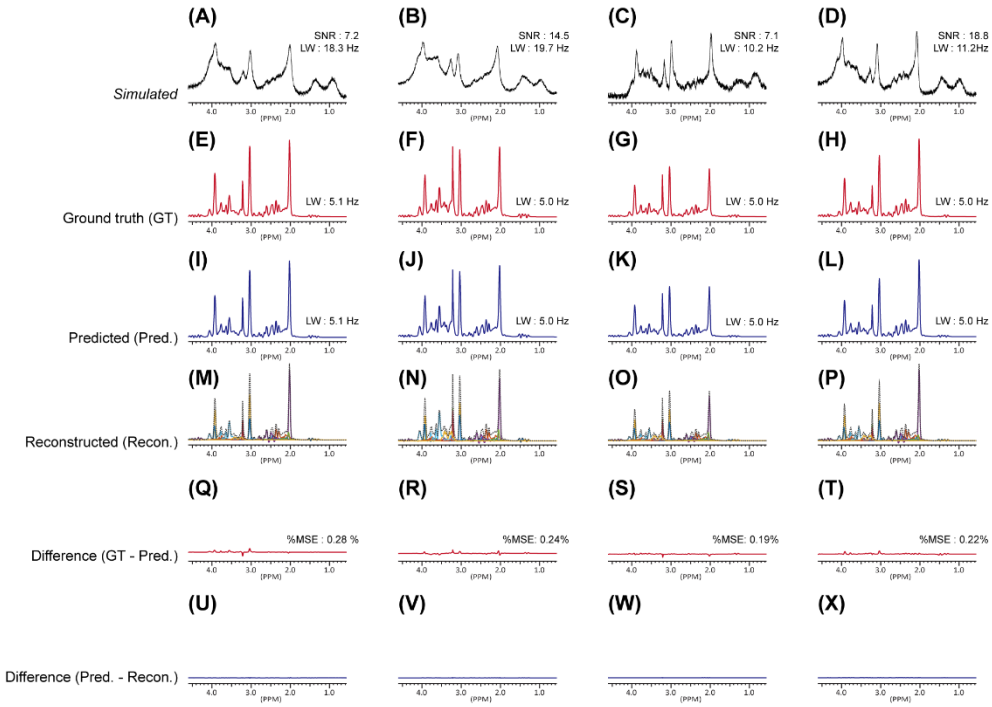


Figure 4-2. The representative BCNN input, output, and residual spectra. (A)-(D): Simulated brain spectra with different combinations of SNR and linewidth used as inputs to the BCNN. (E)-(H): Ground truth (GT) of the network output. (I)-(L): BCNN-predicted metabolite-only spectra obtained from the predictive mean of the 50-MCDO samples. The linewidths also well match those in GT. (M)-(P): Reconstructed spectra (dotted line) using the estimated metabolite content and basis set for 17 metabolites. The individual metabolite signals constituting the reconstructed spectra are also shown. (Q)-(T): Difference spectra obtained by subtracting BCNN-predicted spectra ((I)-(L)) from GT spectra ((E)-(H)). The mean squared error in percent (%MSE) between the two spectra are also shown. (U)-(X): Difference spectra obtained by subtracting the reconstructed spectra ((M)-(P)) from BCNN-predicted spectra ((I)-(L)).

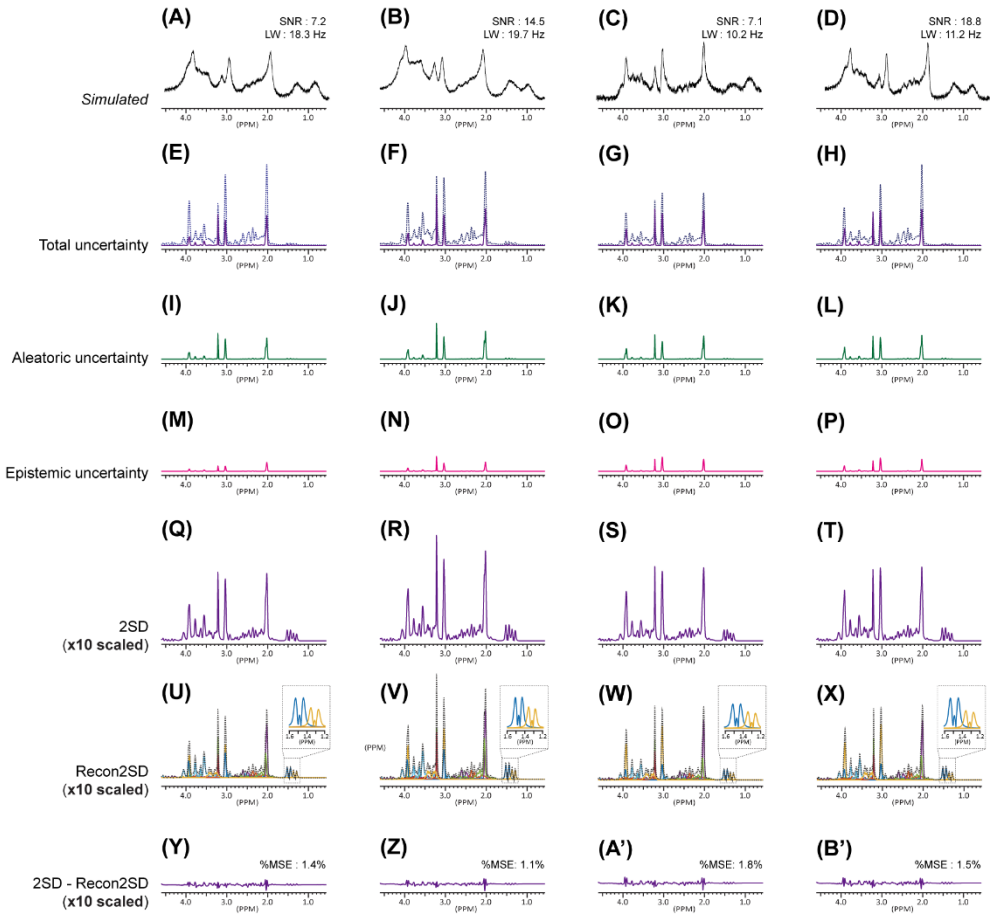


Figure 4-3. The representative uncertainty spectra used in the estimation of uncertainty in metabolite content for the input spectra shown in Figure 4-2(A)-(D). (A)-(D): The same simulated brain spectra shown in Figure 4-2(A)-(D). (E)-(H): Total uncertainty spectra (solid line) obtained by combining the aleatoric and epistemic uncertainty spectra. The BCNN-predicted spectra are also shown in dotted line. (I)-(L): Aleatoric uncertainty spectra. (M)-(P): Epistemic uncertainty spectra. (Q)-(T): 2SD spectra (scaled 10 times) obtained from the total uncertainty spectra. The 2SD spectra are used for the estimation of the uncertainty in metabolite content by multiple regression in combination with the metabolite basis set in absolute mode. (U)-(X): Reconstructed 2SD spectra (Recon2SD; scaled 10 times; dotted line) using the estimated uncertainty in metabolite content and basis set for 17 metabolites. The individual metabolite signals constituting the reconstructed spectra are also shown. (Y)-(B'): Difference spectra (scaled 10 times) obtained by subtracting reconstructed spectra ((U)-(X)) from 2SD spectra ((Q)-(T)). The mean squared error in percent (%MSE) between the two spectra are also shown.

Quantitative results

Figure 4-4 shows the actual errors in the estimated metabolite content in terms of MAPE over the 10,000 spectra in the test set of the simulated data set I. The actual errors (MAPE) are compared with the BCNN-predicted uncertainty expressed in mean %uncertainty. The MAPE of Cr, Glu, Gln, mI, NAA, and Tau are < 10% (< 5% for Glu, NAA, and mI). However, the MAPE of GABA is ~20%, and they are > 20% for Ala, GPC, Lac, NAAG, and PCh. For the majority of the metabolites, the mean %uncertainty are no less than MAPE (except for Gln and GPC).

Figure 4-5 shows the correlations between the GT and estimated metabolite content for the test set. The r ranges from 0.240 ($p < 0.001$; PE) to 0.996 ($p < 0.001$; NAA) (0.797 ± 0.209). For those 6 metabolites with MAPE ≤ 10%, the r ranges from 0.934 ($p < 0.001$; Gln) to 0.996 ($p < 0.001$; NAA) (0.973 ± 0.022).

Table 4-1 summarizes the correlations between the GT error and BCNN-predicted %uncertainty for the test set (0.72-0.94 (0.83 ± 0.06); $p < 0.001$ for all metabolites). For those 6 metabolites with MAPE ≤ 10%, the r ranges from 0.77 (Cr) to 0.94 (Glu) (0.88 ± 0.07). The cases where the predicted %uncertainty is no less than the GT error are also shown in percentage, which ranges 76.6-98.0 (85.0 ± 6.4).

Aleatoric and epistemic uncertainty as a function of SNR, linewidth or concentration

Figure 4-6 compares the aleatoric, epistemic, and total uncertainty of NAA as a function of SNR (A), linewidth (B), or concentration (C) using the simulated data set II. In Figure 4-6(A), the aleatoric uncertainty is far more strongly dependent on the SNR than the epistemic uncertainty. The total uncertainty increases gradually and smoothly as the SNR decreases throughout the entire SNR range. In Figure 4-6(B), overall, the aleatoric uncertainty tends to be increasing as the linewidth is increasing up to the upper bound of the training linewidth range (dotted vertical lines) and shows little change thereafter. The epistemic uncertainty below the lower bound of the training linewidth range is higher than the values inside the training linewidth range. In Figure 4-6(C), (for the given constant noise level) the aleatoric uncertainty is decreasing as the concentration is increasing, as it was decreasing with increasing SNR shown in Figure 4-6(A). The epistemic uncertainty does not change substantially not only inside but also outside the training concentration ranges except in the concentration range beyond ~23 mmol/ L.

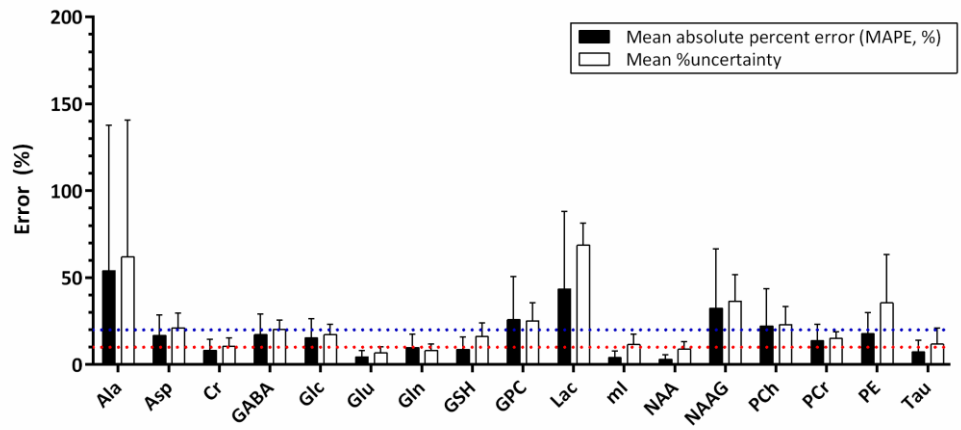


Figure 4-4. The comparison between the actual errors and the BCNN-predicted uncertainty in the estimated metabolite content for individual metabolites. The actual errors in the estimated metabolite content are expressed in terms of mean absolute percent error (MAPE) over the 10,000 simulated spectra in the test set of the simulated data set I. The BCNN-predicted uncertainty in the estimated metabolite content is expressed in mean %uncertainty (mean over the 10,000 simulated spectra). The 10% and 20% error levels (y-axis) are marked in dotted horizontal line

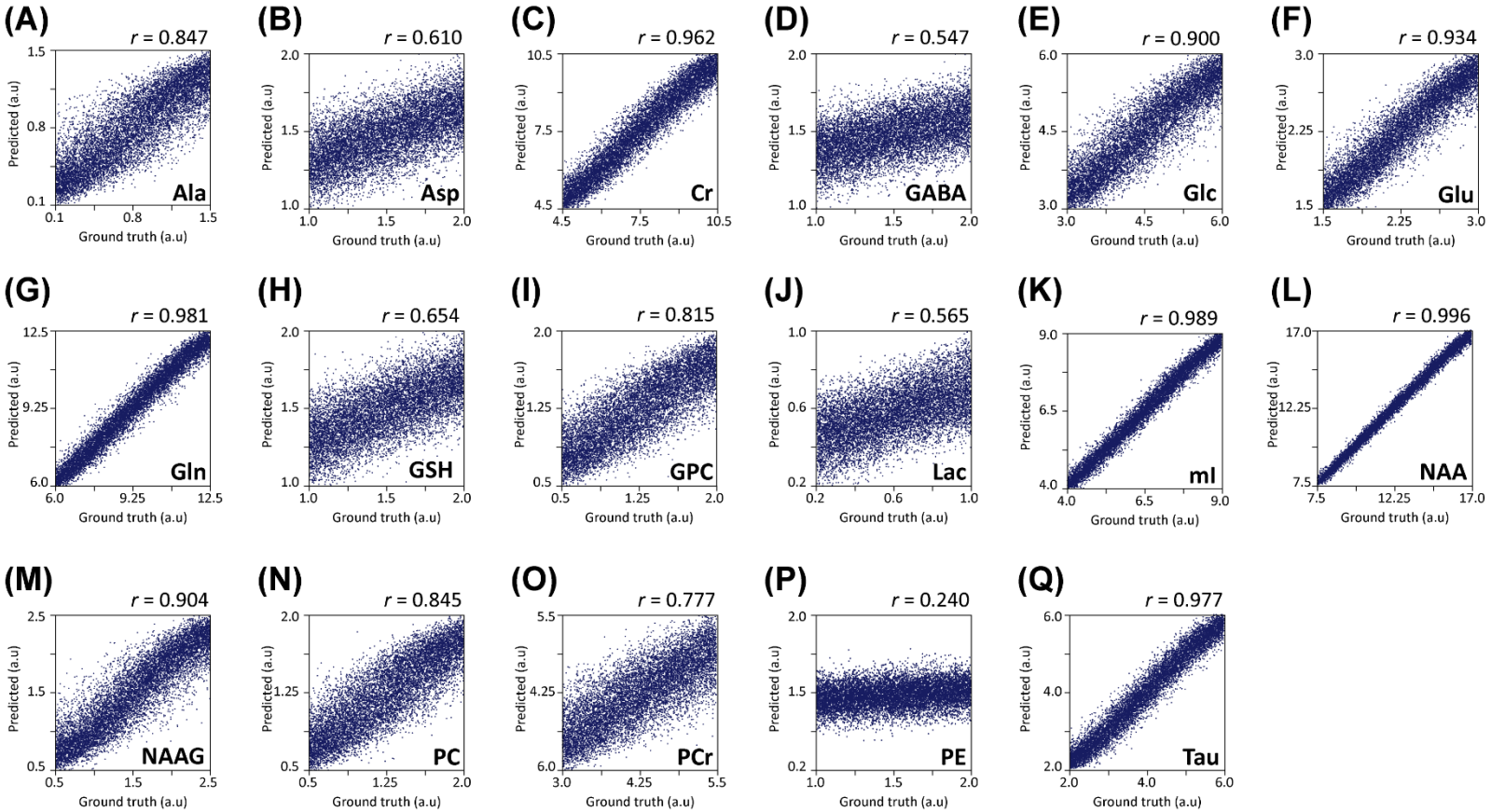


Figure 4-5. The correlations between the ground truth and estimated metabolite content for the 10,000 simulated spectra in the test set of the simulated data set I. The correlation coefficients (r) are shown on each panel. For all r's $p < 0.001$.

	Ala	Asp	Cr	GABA	Glc	Gln	Glu	GSH	GPC
Correlation (r^*)	0.83	0.79	0.77	0.84	0.81	0.83	0.94	0.75	0.72
Number of cases for %uncertainty ≥GT error (%)	82.9	87.9	79.4	84.4	77.0	85.8	95.8	87.0	81.4

	Lac	NAA	NAAG	mI	PCh	PCr	PE	Tau
Correlation (r^*)	0.85	0.93	0.79	0.93	0.76	0.80	0.87	0.87
Number of cases for %uncertainty ≥GT error (%)	83.5	98.0	81.3	97.2	76.6	78.7	81.6	87.1

Table 4-1. The correlations between the absolute error and BCNN-predicted %uncertainty of the individual metabolites for the test set of the simulated spectra set I.

* For all correlations, $p < 0.001$.

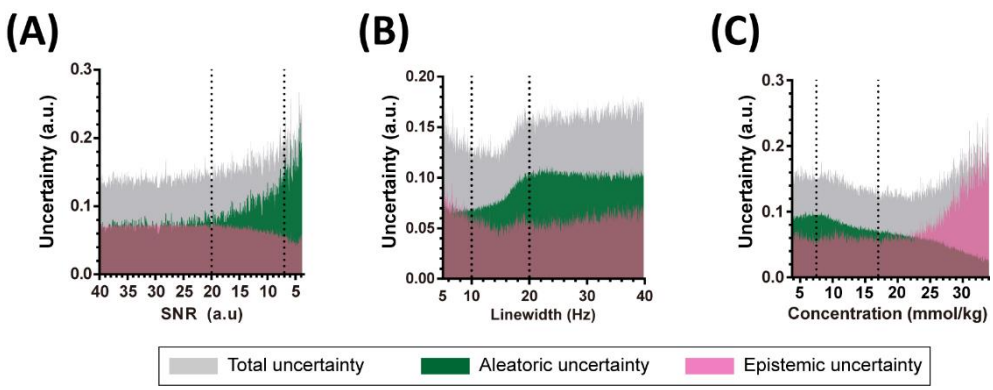


Figure 4-6. Aleatoric (green), epistemic (pink), and total (gray; aleatoric + epistemic) uncertainty of NAA as a function of (A) SNR, (B) linewidth, or (C) concentration using the BCNN and the simulated data set II. The BCNN was trained on the simulated data set I, of which the ranges of SNR, linewidth, and concentrations of NAA are indicated with the dotted vertical lines. The amount of uncertainty was obtained by calculating the area of the NAA signal at ~2.0 ppm from each of the three types of the uncertainty spectra and normalizing it to that from the respective BCNN-predicted (metabolite-only) spectrum. Note that in (A) the SNR is decreasing (from left to right) along the x-axis.

4.3.2. BCNN and LCModel on Modified in vivo Spectra

The representative modified in vivo spectra from one subject are shown in Figure 4-7 for the 3rd (B), 6th (C), and 9th (D) modified spectra groups and the unmodified reference spectra group (A). The increasing severity of spectral degradation due to the adjusted SNR and linewidth is clearly seen from the left to the right.

Table 4-2 summarizes the mean SNR and linewidth over the 5 volunteers for the spectra groups. The severity of spectral degradation is elevated towards the 10th spectra group. Those spectra in the #6 - #10 spectra groups are OOD data.

Figure 4-8 shows the mean variations (in %; mean over the 5 volunteers) in the BCNN-predicted metabolite content across the 10 modified in vivo spectra groups with respect to the metabolite content estimated from the reference spectra group for the 6 representative metabolites. The corresponding mean BCNN-estimated %uncertainty are also shown. The results are compared to the mean metabolite content and CRLB from the LCModel analysis. Overall, for both BCNN and LCModel, the variation in the estimated metabolite content tends to be increasing as the severity of spectral degradation increases. However, the extent of variation in the metabolite content tends to be smaller with BCNN than with LCModel for all metabolites both inside and outside the training ranges except for mI, for which LCModel outperforms BCNN inside the training range. For tCr and tNAA, the variation in metabolite content is more highly correlated with CRLB from LCModel than with %uncertainty from BCNN ($r = 0.942$ vs. 0.910 for tCr; 0.949 vs. 0.934 for tNAA). Overall, however, there is a trend towards higher r with BCNN than with LCModel (0.938 ± 0.019 vs. 0.881 ± 0.057 ($p = 0.115$)).

The original, unmodified in vivo spectra as well as the corresponding BCNN-predicted metabolite-only spectra and uncertainty spectra are shown for all 5 volunteers in Figure 4-9. For the same volunteer used in Figure 4-7, the BCNN-predicted metabolite-only spectra, uncertainty spectra and the LCModel fits are shown for all modified in vivo spectra (belonging to #1 – #10 spectra groups) as well as for the original, unmodified in vivo spectrum in Figure 4-10, Figure 4-11 and Figure 4-12, respectively.

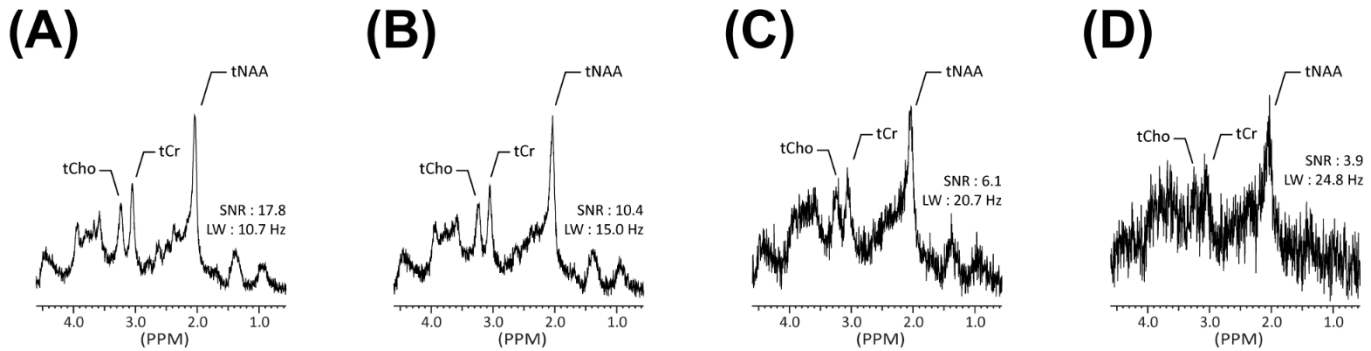


Figure 4-7. The representative modified in vivo spectra from one subject. (A): Unmodified reference (original in vivo) spectrum. (B)-(D): The SNR and linewidth of the reference spectrum shown in (A) were simultaneously modified with 10 different severities of spectral degradation (10 modified spectra groups) and the results are shown for the 3rd (B), 6th (C), and 9th modified spectra groups. The corresponding SNR and linewidth are shown.

Spectra group*	Reference**	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
SNR* (a.u)	16.3±1.1	15.6±1.0	12.7±1.4	10.8±1.5	10.4±0.7	8.1±0.8	6.2±0.5	5.7±0.7	4.5±0.4	3.5±0.3	2.3±0.2
Linewidth* (Hz)	11.4±0.8	13.0±1.4	14.6±1.7	15.8±1.1	17.1±0.8	19.0±0.8	21.7±0.6	24.3±1.5	25.6±1.7	27.8±2.4	30.2±1.7

Table 4-2. The SNR and linewidth of the modified in vivo spectra groups.

* The ranges of the SNR and linewidth are within, and out of, the training ranges of the BCNN for the spectra groups from reference to #5 and for the spectra groups from #6 to #10, respectively.

** Original in vivo spectra

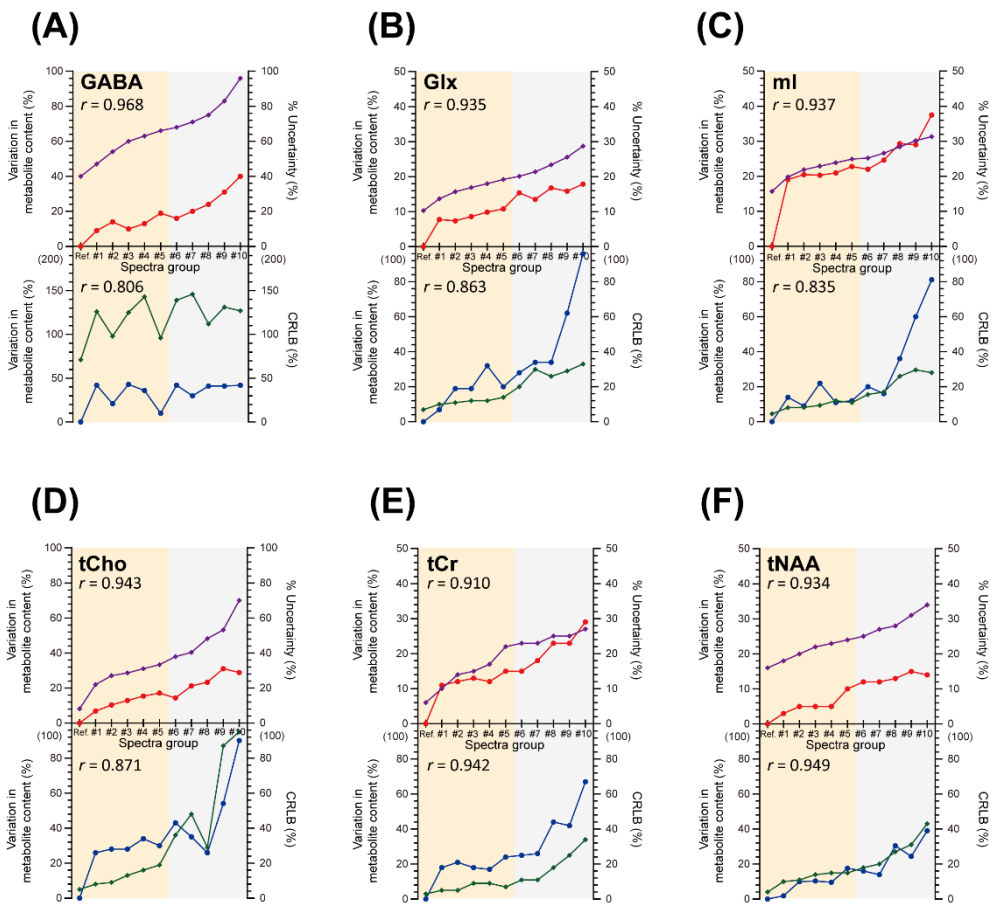


Figure 4-8. The mean variations (%) in the BCNN-predicted metabolite content (upper panel; red line with closed circle) across the 10 modified in vivo spectra groups (#1 - #10) with respect to the metabolite content estimated from the reference spectra group (Ref.) for the 6 representative metabolites of (A) GABA, (B) Glx, (C) mI, (D) tCho, (E) tCr, and (F) tNAA. The mean values of the variations were calculated over the 5 volunteers in each spectra group. The corresponding mean BCNN-estimated %uncertainty (upper panel; purple line with open circle) are also shown. The results from the BCNN are compared to the mean variations in metabolite content (lower panel; blue line with closed diamond) and CRLB (lower panel; green line with open diamond) from the LCModel analysis. The correlation coefficient (r) between the variation in metabolite content and the estimated uncertainty (%Uncertainty for BCNN and CRLB for LCModel) is also shown. Please note the varying scale of the y-axes. For both BCNN and LCModel, the metabolite content was normalized to that of tNAA. In the LCModel analysis, GABA was quantified with null amplitude and 999% CRLB for one subject in spectra group #1 and #3. These data were excluded in the final analysis.

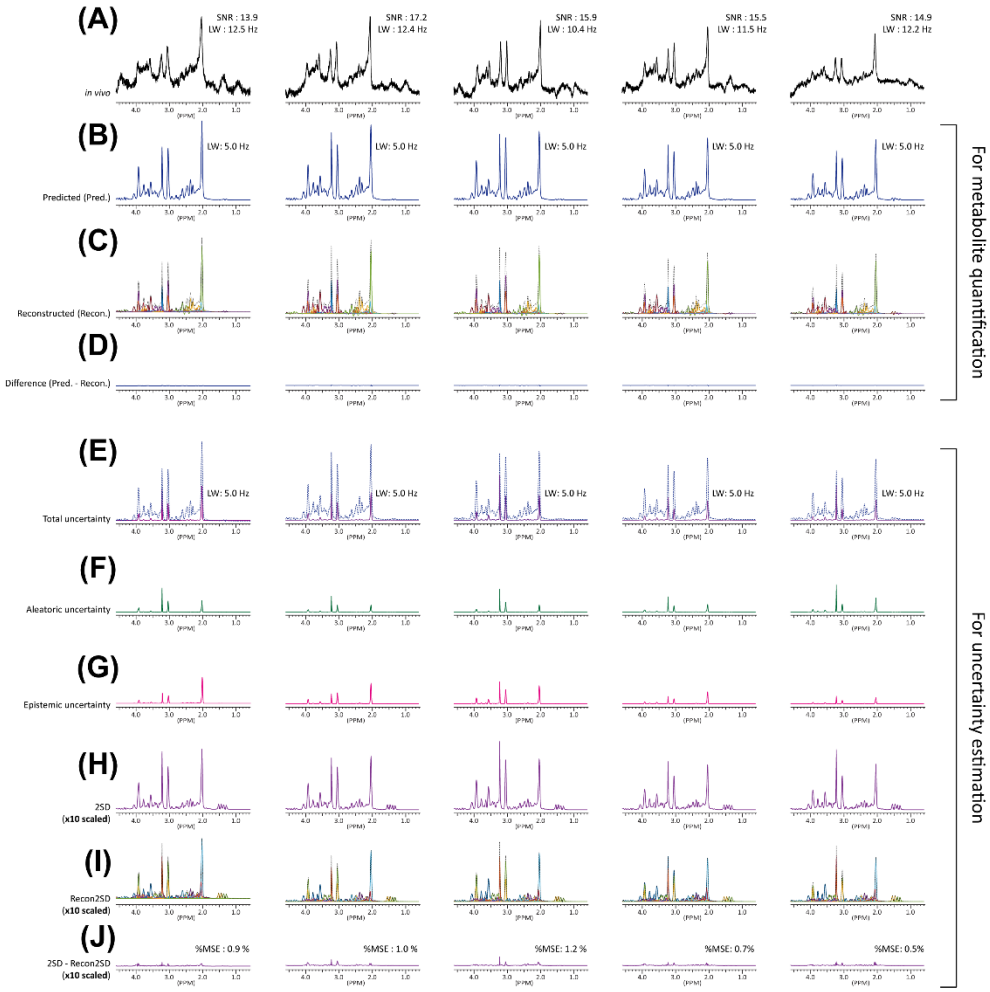


Figure 4-9. The original, unmodified in vivo spectra and corresponding BCNN-predicted metabolite-only spectra and uncertainty spectra for all 5 volunteers. (A) original, unmodified spectra. (B) BCNN-predicted metabolite-only spectra. (C) reconstructed spectra using the metabolite basis set. (D) difference spectra between (B) and (C). (E) total uncertainty spectra. (F) aleatoric uncertainty. (G) epistemic uncertainty. (H) 2SD spectra obtained from (E). (I) reconstructed 2SD spectra using the metabolite basis set. (J) difference spectra between (H) and (I).

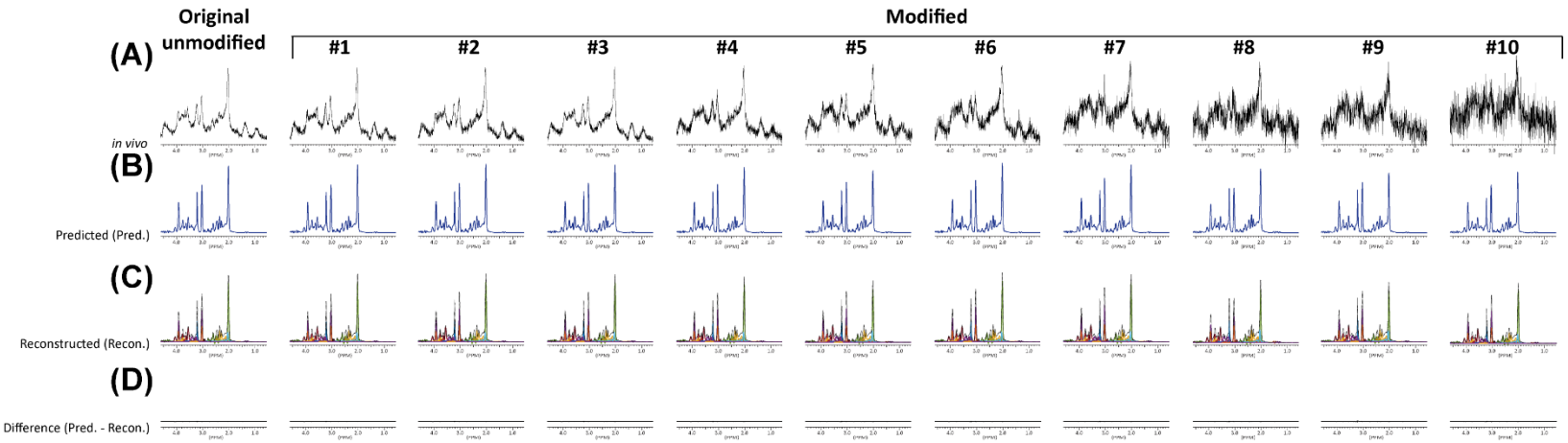


Figure 4-10. The BCNN-predicted metabolite-only spectra for all modified in vivo spectra (#1 - #10) from the same volunteer shown in Figure 4-7.

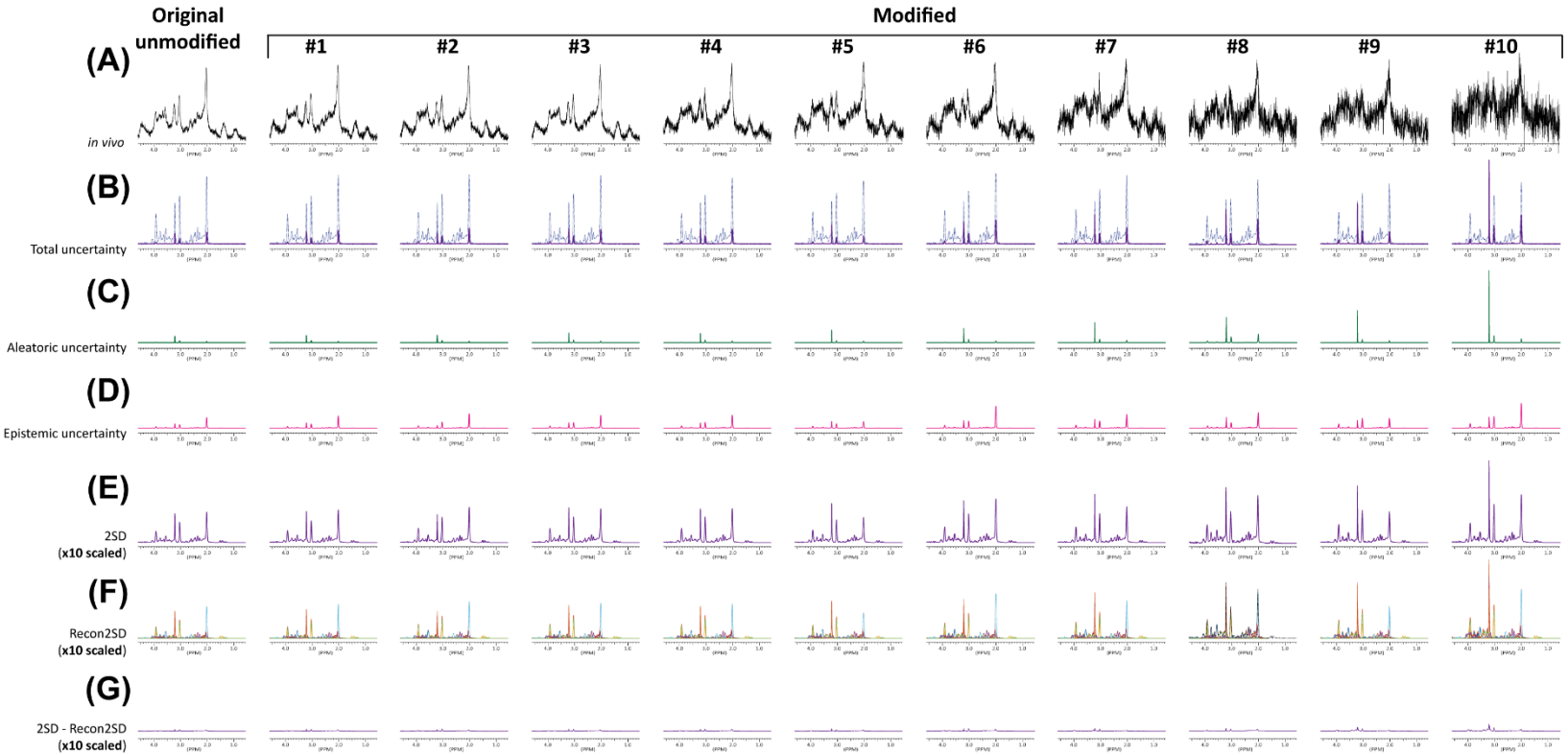


Figure 4-11. The BCNN-predicted uncertainty spectra for all modified in vivo spectra (#1 - #10) from the same volunteer shown in Figure 4-7.

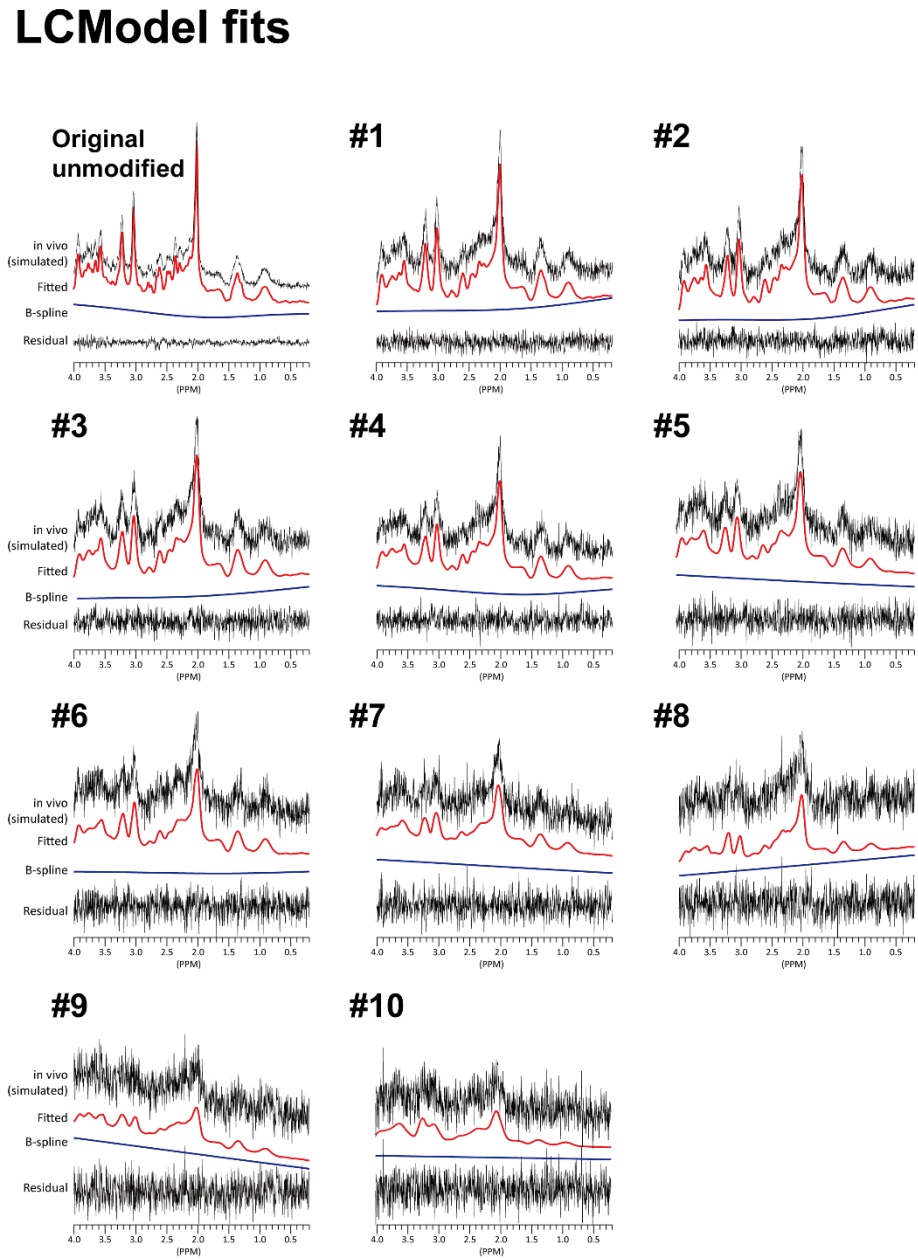


Figure 4-12. The LCModel fits for all modified in vivo spectra (#1 - #10) from the same volunteer shown in Figure 4-7.

4.4. Discussion

4.4.1. Motivation of Study

The recent accomplishment of deep learning is astonishing. [172, 173] However, estimation of uncertainty in the predictions of a neural network is an important prerequisite for more rigorous applications of deep learning, which is particularly so for its clinical applications. [174] According to the growing impact of deep learning on medical data analysis [173, 175], DL-MRS has also been reported for brain metabolite quantification [108, 110, 111, 164], and, indeed, as described in Chapter 3, achieved the estimation of the uncertainty in the CNN-predicted metabolite content by a data-driven approach. However, as the authors discussed, the approach was derived heuristically and suitable only if the SBRs are available for individual metabolites. On the other hand, the MCDO sampling approach [94, 124, 168] is theory-oriented and could be implemented for any CNN by simply adding dropout layers rendered to operate at test time as well. [124] It is increasingly used in deep learning-based MRI studies. [176-178]

4.4.2. Metabolite Quantification on Simulated Brain Spectra

Among the 17 metabolites, Cr, Glu, Gln, mI, NAA, and Tau had MAPE < 10% (< 5% for Glu, NAA, and mI). Tau is less commonly reported than Glx, mI, tCho, tCr, and tNAA from LCModel analysis. Although we are not able to fully understand the relatively low MAPE of Tau in our study due to the black box nature of artificial neural networks, it may likely result, at least in part, from the effective line-narrowing capability of our BCNN. Given the relatively low MAPE of both Glu and Gln, separation of these two metabolites may also be benefited from the line-narrowing, MMs signal-removing, and denoising capability of the BCNN. However, our BCNN should be tested further on more realistically simulated spectra (see the 4.6 study limitation Section) and a larger amount of in vivo data. The quantification of Ala, GPC, Lac, NAAG, and PC remains very challenging and requires further improvement as in Chapter 2. There is a trend towards slight improvement in the accuracy of metabolite quantification for several metabolites (Ala, GSH, Gln, Glu, mI, and Tau) in the current study with respect to the previous study. This may result from the use of a more advanced network architecture (ResNet vs. generic CNN) and

a larger amount of training data (80,000 vs. 40,000). The dependence of MAPE of the metabolites on the amount of training data is provided in Figure 4-13.

4.4.3. Uncertainty Estimation on Simulated Brain Spectra

The predicted errors obtained by the MCDO sampling were highly correlated with GT errors for the majority of the metabolites. We estimated the uncertainty in the BCNN-predicted metabolite content from the 2SD spectrum. Out of the 10,000 spectra in the test set of the simulated data set I, the percentage of the spectra, for which the estimated %uncertainty was no less than the corresponding GT error ranged 76.6 – 98.0% (85.0 ± 6.4%) over the 17 metabolites. It should be noted that this result would depend on the performance of neural networks and test data sets, and the factor of 2 multiplied to the SD spectrum could be adjusted such that the predicted errors are preferentially inclusive of, and as close as, the actual errors.

In the estimation of the uncertainty, the 2SD spectra were in absolute mode. As such, the multiple regression was performed using the metabolite basis set also in absolute mode. This means that those spectral regions where metabolites containing negative signal in real mode due to J-modulation are spectrally overlapped with other metabolites are subject to quantitative errors, if fitted in absolute mode. However, at the TE of 30 ms, at which both the metabolite basis set used for the spectra simulation and the in vivo spectra were acquired, the amount of such negative signal and the resulting quantitative errors are negligible for all metabolites. This is demonstrated in Table 4-3 using the 10,000 spectra in the test set of the simulated data set I. In this additional analysis, the MAPE of the metabolites obtained in absolute mode (for both BCNN-predicted spectra and basis set) were in close agreement with those obtained in real mode (Figure 4-4) with no statistical difference ($p > 0.901$ for all 17 metabolites).

Spectra mode	Ala	Asp	Cr	GABA	Glc	Glu	Gln	GSH	GPC
real*	54.0±83.7	16.8±11.9	8.1±6.4	17.2±12.0	15.4±11.0	4.4±3.7	9.7±7.8	8.7±7.2	25.7±24.9
absolute*	53.8±83.3	16.8±11.9	8.1±6.4	17.4±12.1	15.4±11.1	4.5±3.7	9.8±7.8	8.8±7.2	25.8±24.9
p-value	0.998	0.955	<0.999	0.993	0.989	0.993	0.975	0.902	<0.999

	Lac	mI	NAA	NAAG	PC	PCr	PE	Tau
real	43.4±44.7	4.2±3.5	3.1±2.5	32.4±34.3	22.2±21.4	13.6±9.6	17.7±12.3	7.4±6.6
absolute	43.0±44.2	4.2±3.5	3.1±2.5	32.4±34.3	22.2±21.5	13.6±9.6	17.7±12.3	7.4±6.6
p-value	0.988	0.989	0.996	0.997	0.998	0.996	0.968	0.991

Table 4-3. Comparison between the mean MAPE of the metabolites obtained in real and absolute spectral modes over the 10,000 spectra in the test set of the simulated data set I.

* The mean MAPE values over the 17 metabolites were 17.87±30.04% and 17.87±29.90% for the real and absolute spectral modes, respectively.

4.4.4. Aleatoric, Epistemic and Total Uncertainty as a Function of SNR, Linewidth or Concentration of NAA

In Figure 4-6(A), the substantially increased aleatoric uncertainty and the relatively stable epistemic uncertainty in response to the reduced SNR are in line with the characteristics of aleatoric uncertainty that is associated with inherent noise in the input data. [124, 168] The relatively low total uncertainty beyond the upper bound of the training SNR range (> 20) indicates that the corresponding spectra could also be used for NAA quantification without additional training of the BCNN.

Linebroadening renders peak-to-peak distinction difficult and therefore might be recognized to a certain extent as another form of noise in the point of view of the BCNN. This might explain, at least in part, the increasing aleatoric uncertainty as a function of linewidth in the ~5 – 20 Hz range in Figure 4-6 (B). In general, a lower CRLB is expected for those spectra with a narrower linewidth in the NLSF approach. However, the epistemic uncertainty appears higher below the lower bound of the training linewidth range (< 10 Hz) in Figure 4-6 (B). Therefore, our BCNN requires more training if it is to be used also in that narrower linewidth range.

In Figure 4-6 (C), the noise levels of the spectra were fixed. Therefore, the decrease in aleatoric uncertainty in combination with the relatively stable epistemic uncertainty as the concentration of NAA increases indicate again that aleatoric uncertainty is associated with SNR. [124, 168] The epistemic uncertainty did not change substantially against concentration changes even outside the training range, but was rapidly increased beyond ~23 mmol/L. Therefore, our BCNN might also be used for those input spectra where NAA is present with higher concentrations than the training range to a certain extent, which is in line with the observation in Figure 4-6 (A). However, as the NAA level deviates from that extent of tolerance, the spectra may finally be recognized as the OOD data by the BCNN (i.e., too dominating a signal at ~2.0 ppm), and consequently the epistemic uncertainty starts rapidly increasing. This result together with the continuous monotonic decrease in aleatoric uncertainty are also in line with the characteristics of the two types of uncertainty. [124, 168]

As discussed above, the current approach where the total uncertainty can be resolved into the data and model uncertainty provides an insight into the effective training strategies.

4.4.5. Robustness of BCNN against SNR and Linewidth Tested on Modified in vivo Spectra

The robustness of the proposed method was compared with that of the LCModel analysis on the modified in vivo spectra. For both BCNN and LCModel, the variation in the estimated metabolite content tended to be increasing as the severity of spectral degradation increases, but to a lesser extent with BCNN. For mI, the variation obtained from the BCNN was larger than expected from Figure 4-4, which requires further investigation on a larger data set. Nonetheless, for the rest of the metabolites, the variation obtained with the BCNN was less than 20%, and, for tNAA and Glx, it was ~10% or less in the training range even including those spectra in the 5th modified spectra group (#5 in Table 4-2), the SNR and linewidth of which were degraded by a factor of ~2 with respect to the reference spectra. Even beyond the training ranges, the BCNN-predicted metabolite content was smoothly varying, which is in line with the findings in Figure 4-6. The finding herein that BCNN tends to be more robust than LCModel against spectral degradation is encouraging.

Overall, the correlation of the variation in the estimated metabolite content also tended to be higher with %uncertainty from BCNN than with CRLB from LCModel (0.938±0.019 vs. 0.881±0.057 ($p = 0.115$)). The strong correlation between the variation in the estimated metabolite content and the uncertainty obtained from BCNN seen in Figure 8 together with those findings from Figure 4-4 and Table 4-1 support that the uncertainty estimated by the proposed method is indicative of the accuracy of the estimated metabolite content rather than precision. As is well known and as discussed extensively in the Chapter 3, CRLB provides minimum possible errors rather than actual errors as a measure of the level of uncertainty. Therefore, the relatively low correlation of the CRLB values with the variation in the estimated metabolite content found in this study needs to be interpreted carefully.

4.4.6. Study Limitation

The accuracy of metabolite quantification with the BCNN in this study tended to be improved for several metabolites compared to that with a CNN in the Chapter 2. However, the BCNN still requires data preprocessing such as phase/frequency

correction, Fourier transform, removal of residual water signal, and cropping of input spectra. While the preprocessing may not require a considerable amount of effort for MR spectroscopists, fully automated quantitative analysis directly from the raw data as in the LCModel analysis should be the next step. It would facilitate rapid processing of large data sets such as those from magnetic resonance spectroscopic imaging (MRSI) (Table 4-4). In the development of such a fully automated analysis tool, the ranges of the simulation parameters also need to be extended accordingly (esp., phase (both zeroth- and first-order) and frequency shift).

Having mentioned a fully automated analysis, the analysis should also include handling of spectral artifacts such as ghosting. We investigated the performance of our BCNN in response to the spectra with varying degrees of contamination by residual water (for mimicking the baseline variation), lipid, or ghosting. We used residual water signal and ghosting artifact that were acquired in phantom in the previous study. [179] The lipid signal was modeled by using jMRUI. These different types of spectral contaminators were added to the original, unmodified in vivo spectrum shown in Figure 4-7 with varying degrees of contamination.

The performance of our BCNN in response to these spectra are shown in Figure 4-14. All spectra that are involved in the course of the metabolite quantification and uncertainty estimation by BCNN are shown. Note that these types of spectra that are contaminated by residual water, lipid, or ghosting artifact have never been shown to the BCNN during training. These spectra are also clearly distinguished from those out-of-distribution data set that we simulated for additional testing of the BCNN by simply extending the ranges of SNR and linewidth. Nonetheless, according to “Difference (Pred. in (A) – Pred.)” (the 4th row in the Figure 4-14; the 1st dotted red box), which were obtained by subtracting each BCNN-predicted spectrum from the BCNN-predicted spectrum of the original, unmodified spectrum in A, our BCNN appears robust against mild contamination (i.e., contamination level B). However, at contamination level C, residual signal (difference) starts noticeable for lipid in particular. At contamination level D, residual signal (difference) is noticeable not only for lipid but also for residual water.

For lipid, in line with the noticeable residual signal in “Difference (Pred. in (A) – Pred.)” at contamination levels C and D, the corresponding “2SD” spectra (the

3rd row from the bottom of the Figure 4-14; the 2nd dotted red box), which are used for the calculation of %uncertainty show higher intensity (larger uncertainty) of the peaks. On the other hand, for residual water, such higher peaks (larger uncertainty) are not clearly depicted in the "2SD" spectrum at contamination level D, despite the noticeable residual signal in the corresponding "Difference (Pred. in (A) – Pred.)".

In summary, our BCNN in its current form appears robust against mild spectral contamination by residual water (baseline variation), lipid, or ghosting artifact. For more severe contamination, however, it is subject to less accurate prediction of metabolite-only spectra and uncertainty spectra (2SD spectra). Therefore, a separate unit (machine) needs to be incorporated into the working pipeline, which is capable of detecting and even removing such unwanted signal from input spectra. Alternatively, the neural network should be trained such that metabolite-only spectra and uncertainty spectra could still be accurately retrieved from those contaminated spectra.

Recent studies have shown that such unwanted signal can be detected [128, 129, 179] and even removed [129] by deep learning approaches. Therefore, such a machine also needs to be incorporated in the fully automated working pipeline. Alternatively, our BCNN needs to be trained such that metabolite-only spectra and uncertainty spectra could still be accurately retrieved from those contaminated spectra, which should otherwise be discarded.

In this study, we used simulated spectra for both training and testing of the BCNN. The effectiveness of the simulation strategy and the utility of the simulated spectra were demonstrated in the previous study where, using the same simulation strategy, a CNN was trained solely on the simulated brain spectra, and yet the performance of the CNN on actual in vivo spectra were comparable to those from the LCModel analysis as described in Chapter 2. Nonetheless, the lack of incorporation of the baseline variation into the simulation is an additional limitation of our study. Although we modeled it simply using residual water in the above discussion, it results from many other factors including hardware imperfection [180], and can be difficult to recognize by visual inspection. Due to the difficulty of precisely characterizing and modeling the baseline variation, we did not include it explicitly in the simulation. Instead, we simulated MMs signal by randomly varying the linewidth and amplitude of the 9 MMs resonance groups (M1-M9) individually

(within ±20% and ±10%, respectively) as described in Chapter 2. The relative amount of MMs signal to metabolite signal was also randomly varied (within ±25%). Therefore, our BCNN is somewhat tolerant against variation in the shape and amount of the background signal of the metabolites (i.e., the leftover excluding the metabolite signal), and may also likely be tolerant of the baseline variation to a certain extent. In additional analysis (Figure 4-15), the BCNN was tested on a spectrum without MMs signal at all. On the one hand, the overall quantification errors from the spectrum without MMs were more than twice as large as those from the spectrum with MMs. On the other hand, our BCNN did not totally fail, despite the fact that such a spectrum without MMs signal is a completely out-of-distribution sample. This observation supports, at least in part, that our BCNN may likely be capable of handling the baseline variation to a certain extent. This is not to say that our simulation approach can well account for the baseline variation. Ultimately, it should also be considered explicitly in the simulation. In the simulation of spectra, we used solely a Lorentzian apodization function, and additional analysis (Figure 4-16) shows that substantial line-broadening by different apodization functions such as Gaussian and Voigt functions can influence the performance of the BCNN. As depicted in "Difference (Pred. in (A) – Pred.)" (the 3rd row in the Figure 4-16; dotted red box), which were obtained by subtracting each BCNN-predicted spectrum from the BCNN-predicted spectrum of the reference spectrum in A, the influence of line-broadening is larger with different apodization functions than with the Lorentzian function that was used in our study. The corresponding quantitative results are summarized in the Table 4-5. Broadening the linewidth of the reference spectrum from the lower bound (~10 Hz) to the upper bound (~21 Hz) of the linewidth range of the training set resulted in the mean variation in metabolite content of 3.5% for the Lorentzian. For mI, Glx, tCr, tCho, and tNAA that are most commonly reported in MRS studies, the variations ranged 1.9-4.4%. However, for the Gaussian and Voigt functions, the mean variations increased to 6.7% and 4.9%, respectively. For mI, Glx, tCr, tCho, and tNAA, the variations ranged 0.6-14.7% for the Gaussian function and 1.4-9.3% for the Voigt function. Therefore, such sensitivity of neural networks to the choice of apodization functions should also be recalled in DL-MRS.

Lastly, this study was designed not to provide a lower bound on the target linewidth. Training a CNN with a narrower target linewidth (< 5 Hz) for reduced

spectral overlap could potentially improve the quantitative outcome to a certain extent on the one hand. On the other hand, development of such a CNN that should learn more spectral details could be more challenging in practice, given the limited performance of our CNN for the current target linewidth of ~5 Hz. For instance, the improved performance resulting from the narrower target linewidth might be traded off with a reduced upper limit of the linewidth of input spectra that can be processed with an acceptable quantitative error.

		single spectrum		40 spectra		1024 spectra[&]	
		BCNN[*]	LCModel[#]	BCNN[*]	LCModel[#]	BCNN[*]	LCModel[#]
processing time (sec)[$]	CPU	1.6	6.2	16.3	253	332	6742
	GPU	0.5	-	0.9	-	13.6	-

Table 4-4. Comparison of processing times between BCNN and LCModel.

$ A personal computer was used (Intel Core i5-9400F CPU @ 2.90Hz (4 cores), DDR4 RAM 16 GB, Ubuntu 20.04.2 LTS, Matlab (v.9.7 (R2019b), a single GPU (NVIDIA 1660ti)).

& a single MRSI data that consists of 1024 spectra (MRSI matrix size = 32 x 32)

* Unlike LCModel, the current version of our BCNN requires data preprocessing, which was not accounted for in the estimated processing time. The processing time was estimated for 50 MCDO samples (T = 50).

version v.6.3-1J

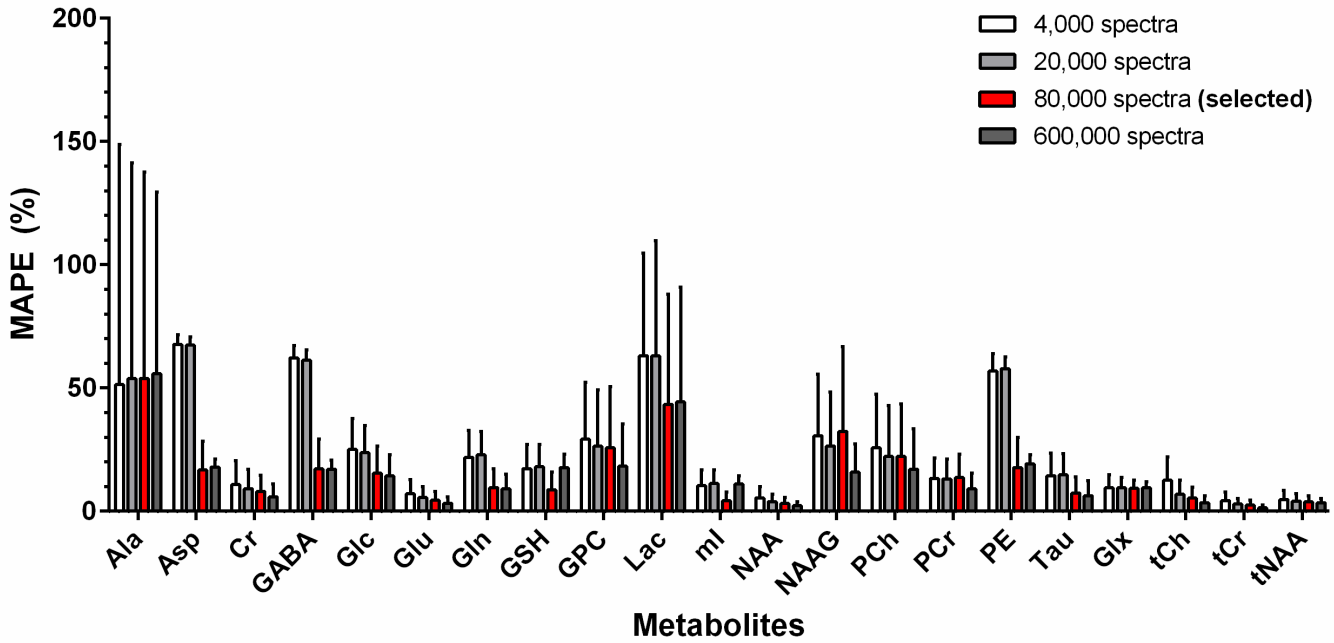
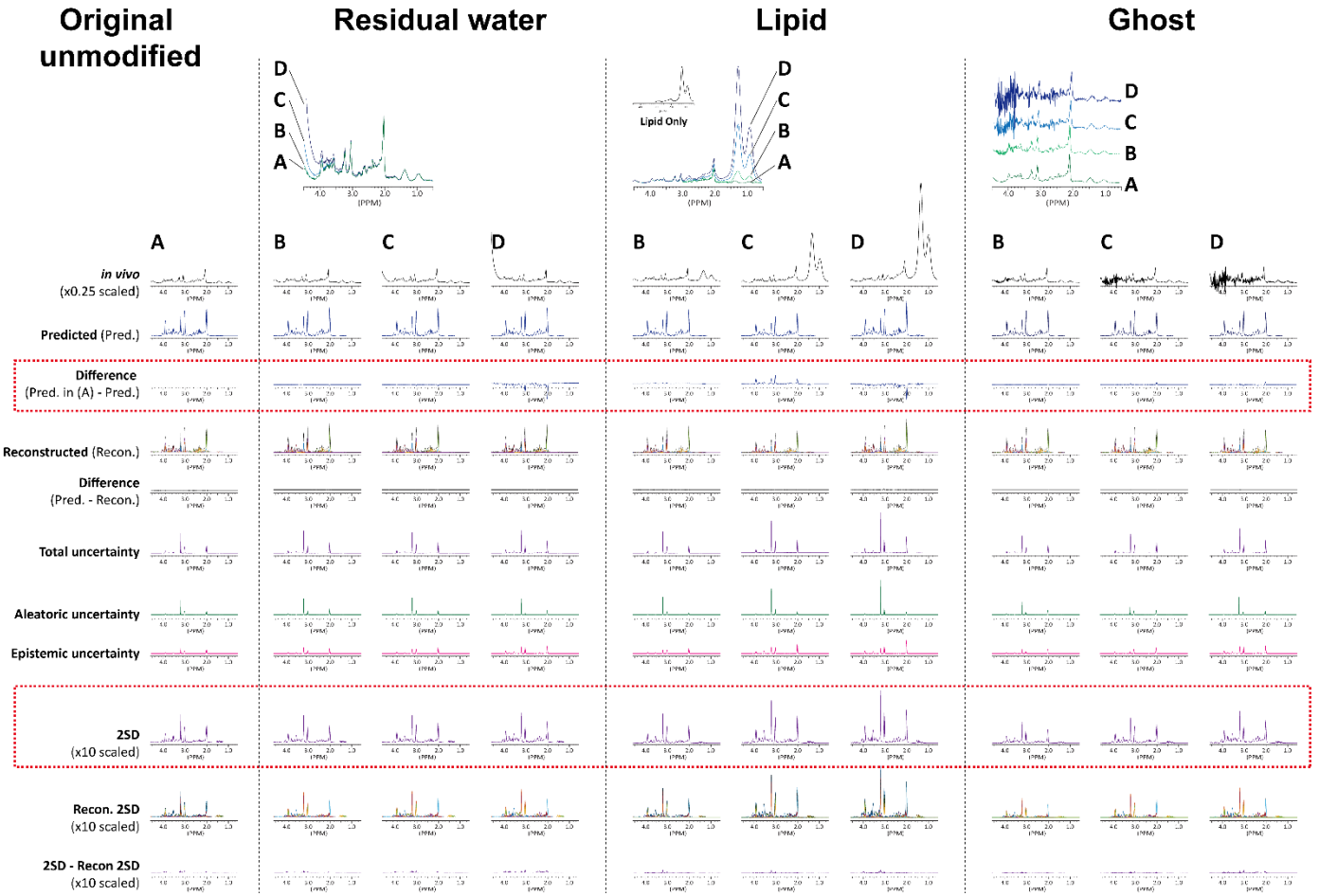


Figure 4-13. The dependence of MAPE of the metabolites on the amount of training data. The figure compares the mean absolute percent errors (MAPE) of the individual metabolites over the 10k spectra in the test set after training the BCNN on each of the 4k, 20k, 80k (finally selected in this study), and 600k training data sets. The same test set (10k spectra) was used for all training data sets. Overall, there is a trend towards decreasing MAPE as the number of training data increases. The MAPE tends to drop suddenly at 80k training set (for Asp, GABA, Glc, Gln, Lac, PE, and Tau). The MAPE of the metabolites for the 80k training set are comparable to those for the 600k training set, despite the substantial difference in the number of spectra between the two training sets. According to this result and in consideration of training efficiency, we had finally selected the 80k training set in this study.



(Figure caption on next page.)

Figure 4-14. The performance of BCNN in response to spectra with varying degrees of contamination by residual water, lipid, or ghosting artifact. Note that these types of spectra that are contaminated by residual water, lipid, or ghosting artifact have never been shown to the BCNN during training. These spectra are also clearly distinguished from those out-of-distribution data set that we simulated for additional testing of the BCNN by simply extending the ranges of SNR and linewidth. Nonetheless, according to "Difference (Pred. in (A) – Pred.)" (the 4th row in the figure; the 1st dotted red box), which were obtained by subtracting each BCNN-predicted spectrum from the BCNN-predicted spectrum of the original, unmodified spectrum in A, our BCNN appears robust against mild contamination (i.e., contamination level B). However, at contamination level C, residual signal (difference) starts noticeable for lipid in particular. At contamination level D, residual signal (difference) is noticeable not only for lipid but also for residual water.

For lipid, in line with the noticeable residual signal in "Difference (Pred. in (A) – Pred.)" at contamination levels C and D, the corresponding "2SD" spectra (the 3rd row from the bottom of the figure; the 2nd dotted red box), which are used for the calculation of %uncertainty show higher intensity (larger uncertainty) of the peaks. On the other hand, for residual water, such higher peaks (larger uncertainty) are not clearly depicted in the "2SD" spectrum at contamination level D, despite the noticeable residual signal in the corresponding "Difference (Pred. in (A) – Pred.)".

In summary, our BCNN in its current form appears robust against mild spectral contamination by residual water (baseline variation), lipid, or ghosting artifact. For more severe contamination, however, it is subject to less accurate prediction of metabolite-only spectra and uncertainty spectra (2SD spectra). As discussed in the main text (discussion section), therefore, a separate unit (machine) needs to be incorporated into the working pipeline, which is capable of detecting and even removing such unwanted signal from input spectra. Alternatively, the neural network should be trained such that metabolite-only spectra and uncertainty spectra could still be accurately retrieved from those contaminated spectra.

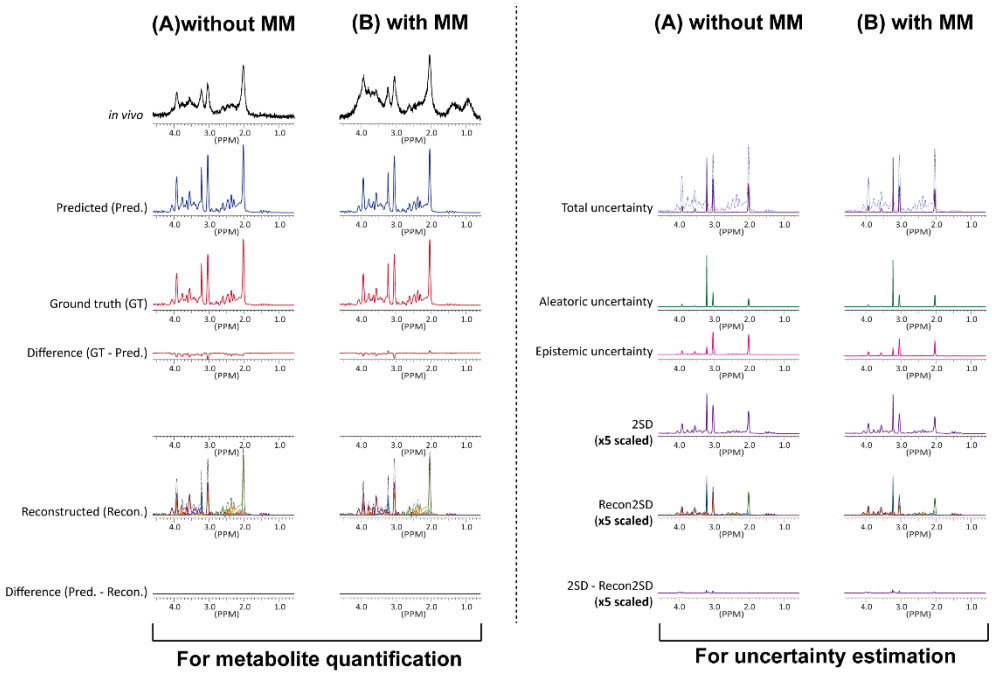


Figure 4-15. The performance of BCNN in response to a spectrum without MM signal. To examine the performance of BCNN in response to a spectrum without macromolecule signal, two identical metabolite spectra was simulated and added macromolecule signal only to one of them. To isolate the impact of the presence/absence of macromolecule signal on the performance of the BCNN, the SNR and linewidth of the spectra were adjusted to the mean values of the training data set. The BCNN-generated spectra in the course of metabolite quantification and uncertainty estimation are shown below for the simulated spectra without (A) and with (B) macromolecule (MM) signal.

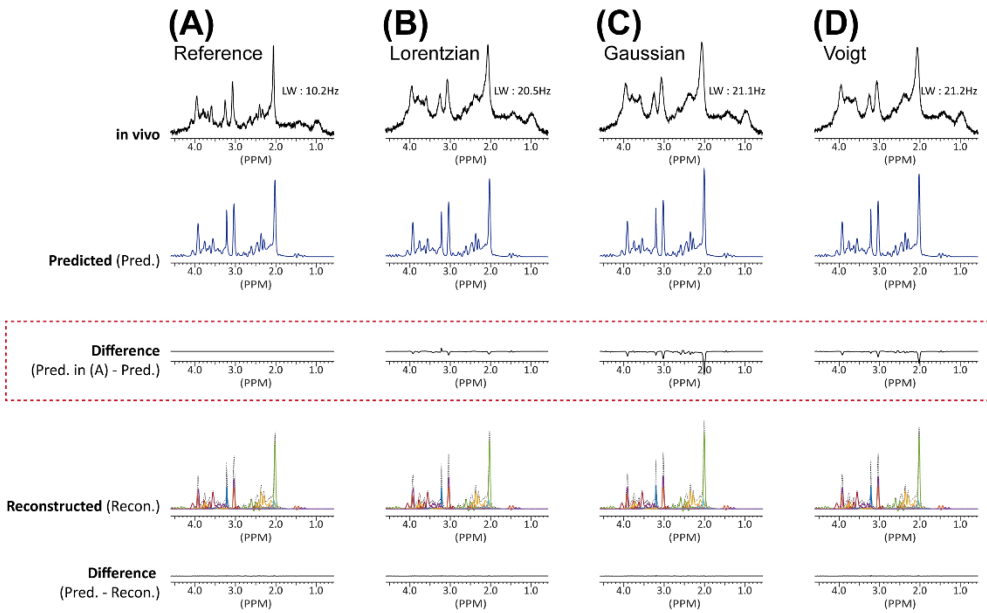


Figure 4-16. The performance of the BCNN on the spectra broadened with different apodization functions. ((A) reference, (B) Lorentzian function, (C) Gaussian function, and (D) Voigt function)

Metabolite	Apodization function		
	Lorentzian	Gaussian	Voigt
Ala	6.5	14.2	12.4
Asp	1.2	6.7	2.8
Cr	6.2	9.0	7.7
GABA	0.8	2.1	0.5
GPC	3.9	5.6	1.9
GSH	0.7	5.2	0.1
Glc	5.0	5.9	8.5
Gln	2.0	1.3	3.9
Glu	2.1	7.7	3.9
Lac	4.4	6.0	4.7
NAA	2.8	15.1	8.8
NAAG	3.9	11.6	12.9
mI	1.9	0.6	2.4
PC	3.7	7.1	0.8
PCr	1.8	2.3	1.8
PE	3.7	1.6	3.1
Tau	10.1	6.3	8.6
Glx	2.1	5.1	1.7
tCho	3.8	6.4	1.4
tCr	4.4	6.2	5.2
tNAA	2.9	14.7	9.3
Mean (%)	**3.5**	**6.7**	**4.9**
SD (%)	**2.0**	**3.8**	**3.4**

Table 4-5. The variation (%) in the estimated metabolite content with respect to the metabolite content estimated from the reference spectrum.

Chapter 5. Conclusion

5.1. Study summary

This study utilized deep learning to method developed for quantifying in vivo metabolites using MRS and obtained results showing improved accuracy and performance than the existing NLSF method. In the deep learning approach, the metabolite-only spectrum was set as GT and trained to extract the metabolite information inherent in the MR spectrum as much as possible. That is, we proposed a method for predicting all metabolites (Chapter 2 and 4) or isolating and extracting only a specific target metabolite with metabolites serving as an internal reference (Chapter 3). As a result, both methods showed high accuracy and low variability in metabolite quantification performance compared to the NLSF methods.

Moreover, this study shown a quantitative uncertainty for the predicted results in deep learning. That is, we developed an uncertainty quantification technique based on a big data-driven heuristic approach (Chapter 3) and a theoretical approach through BCNN (Chapter 4) that approximates variational inference (Chapter 4) and applied it in both in vivo and simulated brain MR spectrum. As a result, it provides uncertainty information closer to the actual accuracy than the CRLB used in the NLSF.

Therefore, the metabolite quantification method through deep learning-based MRS presented in this study is expected to increase clinical applicability by accurately providing in vivo metabolite information.

5.2. Future works

GABA and Glu are important metabolites in various diseases, but are known to be difficult to quantify with conventional methods. Nonetheless, MAPEs of less than 15% and 10% were obtained for GABA and Glu on the simulated spectra. Using true in vivo data, it was also shown that the proposed deep learning-based methods were more robust against poor SNR (Chapter 2 and 4) and spectral dispersion (Chapter 4). These results clearly demonstrate the potential applicability of the proposed methods. For animal studies at high field, the advantage of the higher SNR can easily be compensated with the far smaller voxel size, especially, for mice.

In this circumstance, the robust performance of our methods against poor SNR could be advantageous over the LCModel. However, since the current study was conducted only on healthy human/animals, all of the methods developed in this study should extensively be verified on a larger amount of actual in vivo data representing various diseases.

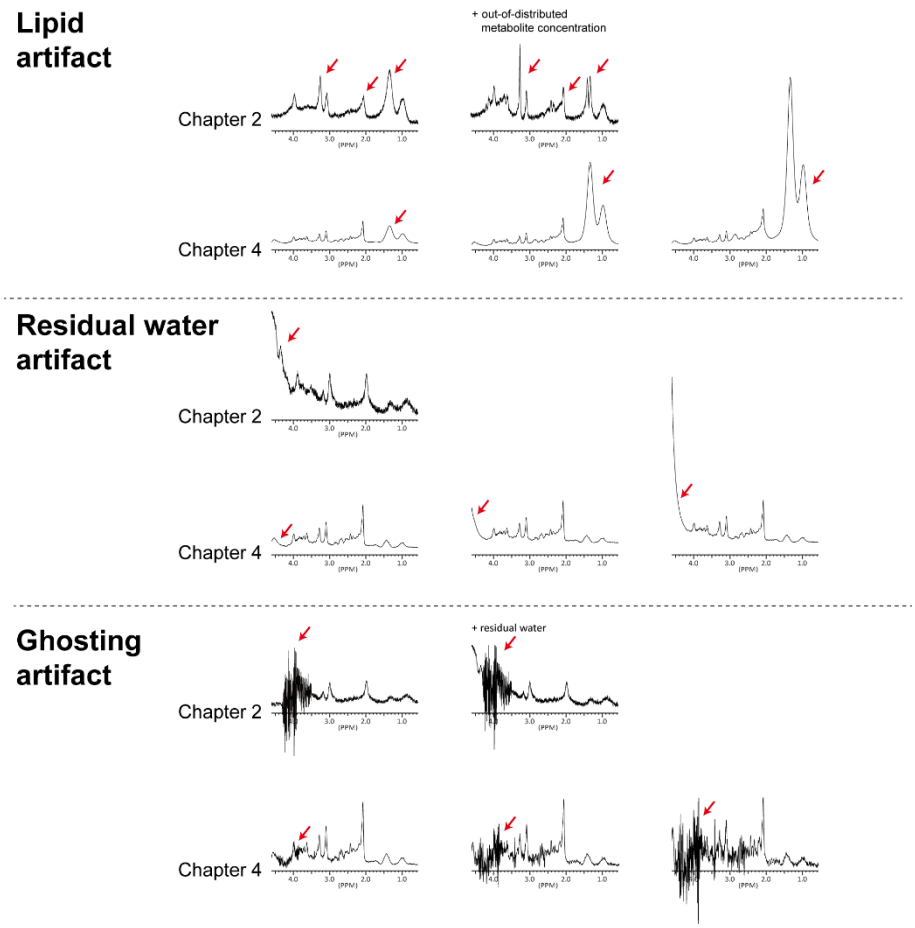


Figure 5-1. Abnormal MRS spectra including spectral artifact used for additional verification of the method presented in the current study.

Spectra contaminated by unwanted signal such as ghosting, residual water signal, and lipid signal are encountered more often than not. (Figure 5-1) As discussed in Section 2.4.4 and Section 4.4.6, the methods developed in this research are tolerant only of mild contamination (Figure 4-14). Therefore, an additional technical development is mandatory prior to the practical application of the methods This is also true for those spectra, in which the relative metabolite concentrations are

outside the training range, such as those from tumor patients as discussed in Section 2.4.4.

Furthermore, the current BCNN-based uncertainty estimation method in Section 4.4.6 show that a local perturbation (artifact) can influence the uncertainty for all metabolites. This is clearly an additional limitation of the current method. Such a spectrum may still be identified and rejected prior to BCNN processing, for instance, by defining an inclusion criterion in terms of the sum of the uncertainty over all metabolites. However, to clearly understand as to why the input spectrum has to be rejected, an alternative approach should also be considered. For instance, a generative adversarial network (GAN)-based detection of anomaly in brain spectra has recently been reported. [179] The big data-driven approach proposed in Chapter 3 might also be considered for the estimation of uncertainty for individual metabolites in response to a localized spectral contaminator. However, unlike the GAN-based approach, it would require a tremendous amount of training data in order to account for a variety of different types, locations, and amplitude of spectral artifact.

The separation of uncertainty into the data and model uncertainty as employed in Chapter 4 can potentially provide an intuition into the training strategy. However, Figure 4-6 also implicitly shows the difficulty of clearly distinguishing between these two different types of uncertainty, which could be possible only upon a perfectly trained BCNN. This limitation should also be recalled in the application of our BCNN in its current form.

Magnetic resonance spectroscopic imaging (MRSI) is a technique that combines the single voxel MRS with the spatial encoding in MRI such that it allows spatial mapping of multiple metabolites across a slice or over the whole volume of the brain. FIDs are collected in k-space to form a k-t space data set. Therefore, the current deep learning methods should be applied to each individual FIDs in the k-t space in the case of MRSI data. This means that an MRSI-specific method should be developed for more efficient data processing.

Finally, a number of data repositories are available for deep learning in MRI. However, even as of today, such data sources are very limited in MRS. Initiation of MRS data repositories by cooperative effort of the International Society as well as local research groups would be a major asset in the development of deep learning-based MRS.

Bibliography

1. Bloch, F., *Nuclear induction.* Physical review, 1946. **70**(7-8): p. 460.
2. Purcell, E.M., H.C. Torrey, and R.V. Pound, *Resonance absorption by nuclear magnetic moments in a solid.* Physical review, 1946. **69**(1-2): p. 37.
3. Lauterbur, P.C., *Image formation by induced local interactions: examples employing nuclear magnetic resonance.* nature, 1973. **242**(5394): p. 190-191.
4. Mansfield, P. and P.K. Grannell, *NMR'diffraction'in solids?* Journal of Physics C: solid state physics, 1973. **6**(22): p. L422.
5. Mansfield, P. and A.A. Maudsley, *Medical imaging by NMR.* The British journal of radiology, 1977. **50**(591): p. 188-194.
6. Bose, S.N., *Plancks gesetz und lichtquantenhypothese.* 1924.
7. Gerlach, W. and O. Stern, *Der experimentelle nachweis der richtungsquantelung im magnetfeld.* Zeitschrift für Physik, 1922. **9**(1): p. 349-352.
8. Pauli, W., *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren.* Zeitschrift für Physik, 1925. **31**(1): p. 765-783.
9. Uhlenbeck, G.E. and S. Goudsmit, *Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons.* Die Naturwissenschaften, 1925. **13**(47): p. 953-954.
10. Zeeman, P., *XXXII. On the influence of magnetism on the nature of the light emitted by a substance.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1897. **43**(262): p. 226-239.
11. Krane, K.S. and D. Halliday, *Introductory nuclear physics*. Vol. 465. 1988: Wiley New York.
12. Frisch, R. and O. Stern, *Über die magnetische Ablenkung von Wasserstoffmolekülen und das magnetische Moment des Protons. I.* Zeitschrift für Physik, 1933. **85**(1): p. 4-16.
13. Boltzmann, L., *Studien uber das Gleichgewicht der lebenden Kraft.* Wissenschafiliche Abhandlungen, 1868. **1**: p. 49-96.
14. Bloch, F. and A. Siegert, *Magnetic resonance for nonrotating fields.* Physical Review, 1940. **57**(6): p. 522.
15. Deelchand, D.K., et al., *Across-vendor standardization of semi-LASER for single-voxel MRS at 3T.* NMR in Biomedicine, 2021. **34**(5): p. e4218.
16. Bendall, M.R. and D.T. Pegg, *Uniform sample excitation with surface coils for in vivo spectroscopy by adiabatic rapid half passage.* Journal of Magnetic Resonance (1969), 1986. **67**(2): p. 376-381.
17. Kim, D.H., et al., *Regularized higher-order in vivo shimming.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2002. **48**(4): p. 715-722.
18. Träber, F., et al., *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.* Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2004. **19**(5): p. 537-545.
19. Barker, P.B., D.O. Hearshen, and M.D. Boska, *Single-voxel proton MRS of the*

human brain at 1.5 T and 3.0 T. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2001. **45**(5): p. 765-769.
20. Ethofer, T., et al., *Comparison of longitudinal metabolite relaxation times in different regions of the human brain at 1.5 and 3 Tesla.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2003. **50**(6): p. 1296-1301.
21. Freeman, R. and H. Hill, *Fourier transform study of NMR spin–lattice relaxation by "progressive saturation".* The Journal of Chemical Physics, 1971. **54**(8): p. 3367-3377.
22. Gasparovic, C., H. Chen, and P.G. Mullins, *Errors in 1H-MRS estimates of brain metabolite concentrations caused by failing to take into account tissue-specific signal relaxation.* NMR in Biomedicine, 2018. **31**(6): p. e3914.
23. Hahn, E.L., *Spin echoes.* Physical review, 1950. **80**(4): p. 580.
24. Dickinson, W., *Dependence of the F 19 nuclear resonance position on chemical compound.* Physical Review, 1950. **77**(5): p. 736.
25. Proctor, W. and F. Yu, *The dependence of a nuclear magnetic resonance frequency upon chemical compound.* Physical Review, 1950. **77**(5): p. 717.
26. Wishart, D.S., et al., *1 H, 13 C and 15 N chemical shift referencing in biomolecular NMR.* Journal of biomolecular NMR, 1995. **6**(2): p. 135-140.
27. Gutowsky, H., D. McCall, and C. Slichter, *Coupling among nuclear magnetic dipoles in molecules.* Physical Review, 1951. **84**(3): p. 589.
28. Hahn, E. and D. Maxwell, *Spin echo measurements of nuclear spin coupling in molecules.* Physical Review, 1952. **88**(5): p. 1070.
29. Fermi, E., *Über die magnetischen Momente der Atomkerne.* Zeitschrift für Physik, 1930. **60**(5-6): p. 320-333.
30. Hahn, E. and D. Maxwell, *Chemical shift and field independent frequency modulation of the spin echo envelope.* Physical Review, 1951. **84**(6): p. 1246.
31. Allen, P.S., R.B. Thompson, and A.H. Wilman, *Metabolite-specific NMR spectroscopy in vivo.* NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo, 1997. **10**(8): p. 435-444.
32. Thompson, R.B. and P.S. Allen, *Response of metabolites with coupled spins to the STEAM sequence.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2001. **45**(6): p. 955-965.
33. Yüksel, C. and D. Öngür, *Magnetic resonance spectroscopy studies of glutamate-related abnormalities in mood disorders.* Biological psychiatry, 2010. **68**(9): p. 785-794.
34. Govindaraju, V., K. Young, and A.A. Maudsley, *Proton NMR chemical shifts and coupling constants for brain metabolites.* NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo, 2000. **13**(3): p. 129-153.
35. Behar, K.L., et al., *Analysis of macromolecule resonances in 1H NMR spectra of human brain.* Magnetic Resonance in Medicine, 1994. **32**(3): p. 294-302.
36. Cudalbu, C., et al., *Contribution of macromolecules to brain 1H MR spectra: Experts' consensus recommendations.* NMR in Biomedicine, 2021. **34**(5): p. e4393.
37. Graham, G.D., et al., *Spectroscopic assessment of alterations in macromolecule and*

small-molecule metabolites in human brain after stroke. Stroke, 2001. **32**(12): p. 2797-2802.
38. Mader, I., et al., *Proton MR spectroscopy with metabolite-nulling reveals elevated macromolecules in acute multiple sclerosis.* Brain, 2001. **124**(5): p. 953-961.
39. Seeger, U., et al., *Parameterized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2003. **49**(1): p. 19-28.
40. de Graaf, R.A., et al., *High magnetic field water and metabolite proton T1 and T2 relaxation in rat brain in vivo.* Magnetic resonance in medicine: An official journal of the international society for magnetic resonance in medicine, 2006. **56**(2): p. 386-394.
41. Penner, J. and R. Bartha, *Semi-LASER 1H MR spectroscopy at 7 Tesla in human brain: metabolite quantification incorporating subject-specific macromolecule removal.* Magnetic resonance in medicine, 2015. **74**(1): p. 4-12.
42. Klose, U., *In vivo proton spectroscopy in presence of eddy currents.* Magnetic resonance in medicine, 1990. **14**(1): p. 26-30.
43. Lin, C., et al., *Eddy current correction in volume-localized MR spectroscopy.* Journal of Magnetic Resonance Imaging, 1994. **4**(6): p. 823-827.
44. Haase, A., et al., *1H NMR chemical shift selective (CHESS) imaging.* Physics in Medicine & Biology, 1985. **30**(4): p. 341.
45. Tkáć, I. and R. Gruetter, *Methodology of 1 H NMR spectroscopy of the human brain at very high magnetic fields.* Applied magnetic resonance, 2005. **29**(1): p. 139-157.
46. Cabanes, E., et al., *Optimization of residual water signal removal by HLSVD on simulated short echo time proton MR spectra of the human brain.* Journal of Magnetic Resonance, 2001. **150**(2): p. 116-125.
47. Ordidge, R.J., A. Connelly, and J.A. Lohman, *Image-selected in vivo spectroscopy (ISIS). A new technique for spatially selective NMR spectroscopy.* Journal of Magnetic Resonance (1969), 1986. **66**(2): p. 283-294.
48. Frahm, J., K.-D. Merboldt, and W. Hänicke, *Localized proton spectroscopy using stimulated echoes.* Journal of Magnetic Resonance (1969), 1987. **72**(3): p. 502-508.
49. Bottomley, P.A., *Selective volume method for performing localized NMR spectroscopy*. 1984, Google Patents.
50. Bottomley, P.A., *Spatial localization in NMR spectroscopy in vivo.* Annals of the New York Academy of Sciences, 1987. **508**(1): p. 333-348.
51. Mlynárik, V., et al., *Localized short-echo-time proton MR spectroscopy with full signal-intensity acquisition.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2006. **56**(5): p. 965-970.
52. Burstein, D., *Stimulated echoes: description, applications, practical hints.* Concepts in Magnetic Resonance, 1996. **8**(4): p. 269-278.
53. Singh, S. and R. Deslauriers, *Tailored selective NMR excitation by projection presaturation.* Concepts in Magnetic Resonance, 1995. **7**(1): p. 1-27.
54. Ratiney, H., et al., *Time-domain semi-parametric estimation based on a metabolite basis set.* NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo, 2005. **18**(1): p. 1-13.
55. Provencher, S.W., *Estimation of metabolite concentrations from localized in vivo*

proton NMR spectra. Magnetic resonance in medicine, 1993. **30**(6): p. 672-679.
56. Wilson, M., et al., *A constrained least-squares approach to the automated quantitation of in vivo 1H magnetic resonance spectroscopy data.* Magnetic resonance in medicine, 2011. **65**(1): p. 1-12.
57. Oeltzschner, G., et al., *Osprey: Open-source processing, reconstruction & estimation of magnetic resonance spectroscopy data.* Journal of Neuroscience Methods, 2020. **343**: p. 108827.
58. De Graaf, R.A., *In vivo NMR spectroscopy: principles and techniques*. 2019: John Wiley & Sons.
59. Smith, S., et al., *Computer simulations in magnetic resonance. An object-oriented programming approach.* Journal of Magnetic Resonance, Series A, 1994. **106**(1): p. 75-105.
60. Moré, J.J., *The Levenberg-Marquardt algorithm: implementation and theory*, in *Numerical analysis*. 1978, Springer. p. 105-116.
61. Cavassila, S., et al., *Cramer-Rao bound expressions for parametric estimation of overlapping peaks: influence of prior knowledge.* Journal of Magnetic Resonance, 2000. **143**(2): p. 311-320.
62. Van Trees, H.L., *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. 2004: John Wiley & Sons.
63. Near, J., et al., *Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations.* NMR in Biomedicine, 2021. **34**(5): p. e4257.
64. Barker, P.B., et al., *Quantitation of proton NMR spectra of the human brain using tissue water as an internal concentration reference.* NMR in Biomedicine, 1993. **6**(1): p. 89-94.
65. Saunders, D.E., et al., *Aging of the adult human brain: in vivo quantitation of metabolite content with proton magnetic resonance spectroscopy.* Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 1999. **9**(5): p. 711-716.
66. Hebb, D.O., *The organisation of behaviour: a neuropsychological theory*. 1949: Science Editions New York.
67. Newell, A., J.C. Shaw, and H.A. Simon. *Report on a general problem solving program*. in *IFIP congress*. 1959. Pittsburgh, PA.
68. Rosenblatt, F., *The perceptron: a probabilistic model for information storage and organization in the brain.* Psychological review, 1958. **65**(6): p. 386.
69. Turing, A.M. and J. Haugeland, *Computing machinery and intelligence*. 1950: MIT Press Cambridge, MA.
70. Newell, A., J.C. Shaw, and H.A. Simon, *Elements of a theory of human problem solving.* Psychological review, 1958. **65**(3): p. 151.
71. Newell, A. and H. Simon, *The logic theory machine--A complex information processing system.* IRE Transactions on information theory, 1956. **2**(3): p. 61-79.
72. Samuel, A.L., *Some studies in machine learning using the game of checkers.* IBM Journal of research and development, 1959. **3**(3): p. 210-229.
73. Krizhevsky, A. and G. Hinton, *Learning multiple layers of features from tiny images.* 2009.
74. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436-444.

75. Bottou, L., *Large-scale machine learning with stochastic gradient descent*, in *Proceedings of COMPSTAT'2010*. 2010, Springer. p. 177-186.
76. LeCun, Y., et al., *Gradient-based learning applied to document recognition.* Proceedings of the IEEE, 1998. **86**(11): p. 2278-2324.
77. Bergstra, J., et al., *Algorithms for hyper-parameter optimization.* Advances in neural information processing systems, 2011. **24**.
78. Hecht-Nielsen, R., *Theory of the backpropagation neural network*, in *Neural networks for perception*. 1992, Elsevier. p. 65-93.
79. Werbos, P., *Beyond regression:" new tools for prediction and analysis in the behavioral sciences.* Ph. D. dissertation, Harvard University, 1974.
80. Shannon, C.E., *A mathematical theory of communication.* The Bell system technical journal, 1948. **27**(3): p. 379-423.
81. Akaike, H., *Information theory and an extension of the maximum likelihood principle*, in *Selected papers of hirotugu akaike*. 1998, Springer. p. 199-213.
82. Nwankpa, C., et al., *Activation functions: Comparison of trends in practice and research for deep learning.* arXiv preprint arXiv:1811.03378, 2018.
83. Hornik, K., M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximators.* Neural networks, 1989. **2**(5): p. 359-366.
84. Kingma, D.P. and J. Ba, *Adam: A method for stochastic optimization.* arXiv preprint arXiv:1412.6980, 2014.
85. Riedmiller, M. and H. Braun. *A direct adaptive method for faster backpropagation learning: The RPROP algorithm*. in *IEEE international conference on neural networks*. 1993. IEEE.
86. Ciaburro, G. and B. Venkateswaran, *Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles*. 2017: Packt Publishing Ltd.
87. Hinton, G.E., S. Osindero, and Y.-W. Teh, *A fast learning algorithm for deep belief nets.* Neural computation, 2006. **18**(7): p. 1527-1554.
88. LeCun, Y. and Y. Bengio, *Convolutional networks for images, speech, and time series.* The handbook of brain theory and neural networks, 1995. **3361**(10): p. 1995.
89. Salakhutdinov, R. and G. Hinton. *Deep boltzmann machines*. in *Artificial intelligence and statistics*. 2009. PMLR.
90. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks.* Advances in neural information processing systems, 2012. **25**: p. 1097-1105.
91. Ying, X. *An overview of overfitting and its solutions*. in *Journal of Physics: Conference Series*. 2019. IOP Publishing.
92. Srivastava, N., et al., *Dropout: a simple way to prevent neural networks from overfitting.* The journal of machine learning research, 2014. **15**(1): p. 1929-1958.
93. Baldi, P. and P. Sadowski, *The dropout learning algorithm.* Artificial intelligence, 2014. **210**: p. 78-122.
94. Gal, Y. and Z. Ghahramani. *Dropout as a bayesian approximation: Representing model uncertainty in deep learning*. in *international conference on machine learning*. 2016. PMLR.
95. Stone, M., *Cross-validatory choice and assessment of statistical predictions.* Journal of the royal statistical society: Series B (Methodological), 1974. **36**(2): p. 111-133.
96. Bengio, Y., P. Simard, and P. Frasconi, *Learning long-term dependencies with*

gradient descent is difficult. IEEE transactions on neural networks, 1994. **5**(2): p. 157-166.
97. Pascanu, R., T. Mikolov, and Y. Bengio. *On the difficulty of training recurrent neural networks*. in *International conference on machine learning*. 2013. PMLR.
98. Mikolov, T., et al., *Subword language modeling with neural networks.* preprint (http://www. fit. vutbr. cz/imikolov/rnnlm/char. pdf), 2012. **8**: p. 67.
99. LeCun, Y.A., et al., *Efficient backprop*, in *Neural networks: Tricks of the trade*. 2012, Springer. p. 9-48.
100. Glorot, X. and Y. Bengio. *Understanding the difficulty of training deep feedforward neural networks*. in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010. JMLR Workshop and Conference Proceedings.
101. He, K., et al. *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*. in *Proceedings of the IEEE international conference on computer vision*. 2015.
102. Ioffe, S. and C. Szegedy. *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. in *International conference on machine learning*. 2015. PMLR.
103. Hubel, D.H. and T.N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.* The Journal of physiology, 1962. **160**(1): p. 106-154.
104. Boureau, Y.-L., J. Ponce, and Y. LeCun. *A theoretical analysis of feature pooling in visual recognition*. in *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010.
105. Chen, C.-C., et al., *An Improved Fault Diagnosis Using 1D-Convolutional Neural Network Model.* Electronics, 2021. **10**(1): p. 59.
106. Malek, S., F. Melgani, and Y. Bazi, *One-dimensional convolutional neural networks for spectroscopic signal regression.* Journal of Chemometrics, 2018. **32**(5): p. e2977.
107. Das, D., et al. *Quantification of metabolites in magnetic resonance spectroscopic imaging using machine learning*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2017. Springer.
108. Gurbani, S.S., et al., *Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting.* Magnetic resonance in medicine, 2019. **81**(5): p. 3346-3357.
109. Hanstock, C., et al., *Localized in vivo proton spectroscopy using depth-pulse spectral editing.* Journal of Magnetic Resonance (1969), 1987. **71**(2): p. 349-354.
110. Hatami, N., M. Sdika, and H. Ratiney. *Magnetic resonance spectroscopy quantification using deep learning*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018. Springer.
111. Iqbal, Z., et al., *Super-resolution 1h magnetic resonance spectroscopic imaging utilizing deep learning.* Frontiers in oncology, 2019. **9**: p. 1010.
112. Kim, H., J.M. Wild, and P.S. Allen, *Strategy for the spectral filtering of myo-inositol and other strongly coupled spins.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2004. **51**(2): p. 263-272.
113. Mescher, M., et al., *Simultaneous in vivo spectral editing and water suppression.* NMR in Biomedicine: An International Journal Devoted to the Development and

Application of Magnetic Resonance In Vivo, 1998. **11**(6): p. 266-272.
114. Naressi, A., et al., *Java-based graphical user interface for the MRUI quantitation package.* Magnetic resonance materials in physics, biology and medicine, 2001. **12**(2-3): p. 141.
115. Prichard, J.W. and R.G. Shulman, *NMR spectroscopy of brain metabolism in vivo.* Annual review of neuroscience, 1986. **9**(1): p. 61-85.
116. Ratiney, H., et al., *Time-domain quantitation of 1 H short echo-time signals: background accommodation.* Magnetic Resonance Materials in Physics, Biology and Medicine, 2004. **16**(6): p. 284-296.
117. Thompson, R.B. and P.S. Allen, *A new multiple quantum filter design procedure for use on strongly coupled spin systems found in vivo: its application to glutamate.* Magnetic resonance in medicine, 1998. **39**(5): p. 762-771.
118. Trabesinger, A.H., et al., *Detection of glutathione in the human brain in vivo by means of double quantum coherence filtering.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 1999. **42**(2): p. 283-289.
119. Van den Boogaart, A., et al., *Removal of the water resonance from 1H magnetic resonance spectra.* Mathematics in signal processing III, 1994: p. 175-195.
120. Vanhamme, L., A. van den Boogaart, and S. Van Huffel, *Improved method for accurate and efficient quantification of MRS data with use of prior knowledge.* Journal of magnetic resonance, 1997. **129**(1): p. 35-43.
121. Wilman, A.H. and P.S. Allen, *Double-quantum filtering of citrate for in vivo observation.* Journal of Magnetic Resonance, Series B, 1994. **105**(1): p. 58-60.
122. Wilman, A.H. and P.S. Allen, *Yield enhancement of a double-quantum filter sequence designed for the edited detection of GABA.* Journal of Magnetic Resonance, Series B, 1995. **109**(2): p. 169-174.
123. Kreis, R., *The trouble with quality filtering based on relative C ramér-R ao lower bounds.* Magnetic resonance in medicine, 2016. **75**(1): p. 15-18.
124. Kendall, A. and Y. Gal, *What uncertainties do we need in bayesian deep learning for computer vision?* arXiv preprint arXiv:1703.04977, 2017.
125. Jiru, F., *Introduction to post-processing techniques.* European journal of radiology, 2008. **67**(2): p. 202-217.
126. Bhogal, A.A., et al., *1H–MRS processing parameters affect metabolite quantification: The urgent need for uniform and transparent standardization.* NMR in Biomedicine, 2017. **30**(11): p. e3804.
127. Pereira, S., et al., *Brain tumor segmentation using convolutional neural networks in MRI images.* IEEE transactions on medical imaging, 2016. **35**(5): p. 1240-1251.
128. Gurbani, S.S., et al., *A convolutional neural network to filter artifacts in spectroscopic MRI.* Magnetic resonance in medicine, 2018. **80**(5): p. 1765-1775.
129. Kyathanahally, S.P., A. Döring, and R. Kreis, *Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy.* Magnetic resonance in medicine, 2018. **80**(3): p. 851-863.
130. Perry, T., et al., *Regional distribution of amino acids in human brain obtained at autopsy.* Journal of neurochemistry, 1971. **18**(3): p. 513-519.
131. Perry, T., et al., *Free amino acids and related compounds in biopsies of human brain.* Journal of neurochemistry, 1971. **18**(3): p. 521-528.
132. Birch, R., et al., *Influence of macromolecule baseline on 1H MR spectroscopic*

imaging reproducibility. Magnetic resonance in medicine, 2017. **77**(1): p. 34-43.
133. Opstad, K.S., et al., *Toward accurate quantification of metabolites, lipids, and macromolecules in HRMAS spectra of human brain tumor biopsies using LCModel.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2008. **60**(5): p. 1237-1242.
134. Gottschalk, M., L. Lamalle, and C. Segebarth, *Short-TE localised 1H MRS of the human brain at 3 T: quantification of the metabolite signals using two approaches to account for macromolecular signal contributions.* NMR in Biomedicine, 2008. **21**(5): p. 507-517.
135. Snoek, J., H. Larochelle, and R.P. Adams, *Practical bayesian optimization of machine learning algorithms.* Advances in neural information processing systems, 2012. **25**.
136. Choi, C., et al., *Measurement of N-acetylaspartylglutamate in the human frontal brain by 1H-MRS at 7 T.* Magnetic resonance in medicine, 2010. **64**(5): p. 1247-1251.
137. Ganji, S.K., et al., *Measurement of regional variation of GABA in the human brain by optimized point-resolved spectroscopy at 7 T in vivo.* NMR in Biomedicine, 2014. **27**(10): p. 1167-1175.
138. Hetherington, H., et al., *Quantitative 31P spectroscopic imaging of human brain at 4 Tesla: assessment of gray and white matter differences of phosphocreatine and ATP.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2001. **45**(1): p. 46-52.
139. Kaiser, L.G., et al., *Age-related glutamate and glutamine concentration changes in normal human brain: 1H MR spectroscopy study at 4 T.* Neurobiology of aging, 2005. **26**(5): p. 665-672.
140. Komoroski, R.A., J.M. Pearce, and R.E. Mrak, *31P NMR spectroscopy of phospholipid metabolites in postmortem schizophrenic brain.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2008. **59**(3): p. 469-474.
141. Pouwels, P.J. and J. Frahm, *Regional metabolite concentrations in human brain as determined by quantitative localized proton MRS.* Magnetic resonance in medicine, 1998. **39**(1): p. 53-60.
142. Kyathanahally, S.P., et al., *Quality of clinical brain tumor MR spectra judged by humans and machine learning tools.* Magnetic resonance in medicine, 2018. **79**(5): p. 2500-2510.
143. Rothman, D.L., et al., *Localized 1H NMR measurements of gamma-aminobutyric acid in human brain in vivo.* Proceedings of the national academy of sciences, 1993. **90**(12): p. 5662-5666.
144. Zhu, H. and P.B. Barker, *MR spectroscopy and spectroscopic imaging of the brain.* Magnetic resonance neuroimaging, 2011: p. 203-226.
145. Edden, R.A., M.G. Pomper, and P.B. Barker, *In vivo differentiation of N-acetyl aspartyl glutamate from N-acetyl aspartate at 3 Tesla.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2007. **57**(6): p. 977-982.
146. Ramadan, S., A. Lin, and P. Stanwell, *Glutamate and glutamine: a review of in vivo MRS in the human brain.* NMR in Biomedicine, 2013. **26**(12): p. 1630-1646.
147. Haris, M., et al., *In vivo mapping of brain myo-inositol.* Neuroimage, 2011. **54**(3):

p. 2079-2085.
148. Pedrosa de Barros, N., et al., *Automatic quality control in clinical 1H MRSI of brain cancer.* NMR in Biomedicine, 2016. **29**(5): p. 563-575.
149. Howe, F., et al., *Metabolic profiles of human brain tumors using quantitative in vivo 1H magnetic resonance spectroscopy.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2003. **49**(2): p. 223-232.
150. Saunders, D.E., et al., *Discrimination of metabolite from lipid and macromolecule resonances in cerebral infarction in humans using short echo proton spectroscopy.* Journal of Magnetic Resonance Imaging, 1997. **7**(6): p. 1116-1121.
151. Cobbold, J.F., et al., *Hepatic lipid profiling in chronic hepatitis C: an in vitro and in vivo proton magnetic resonance spectroscopy study.* Journal of hepatology, 2010. **52**(1): p. 16-24.
152. Cudalbu, C., V. Mlynárik, and R. Gruetter, *Handling macromolecule signals in the quantification of the neurochemical profile.* Journal of Alzheimer's Disease, 2012. **31**(s3): p. S101-S115.
153. Castelvecchi, D., *Can we open the black box of AI?* Nature News, 2016. **538**(7623): p. 20.
154. Jin, K.H., et al., *Deep convolutional neural network for inverse problems in imaging.* IEEE Transactions on Image Processing, 2017. **26**(9): p. 4509-4522.
155. Wenger, K.J., et al., *Fitting algorithms and baseline correction influence the results of non-invasive in vivo quantitation of 2-hydroxyglutarate with 1H-MRS.* NMR in Biomedicine, 2019. **32**(1): p. e4027.
156. Lee, H.H. and H. Kim, *Parameterization of spectral baseline directly from short echo time full spectra in 1H-MRS.* Magnetic resonance in medicine, 2017. **78**(3): p. 836-847.
157. Slotboom, J., et al., *Spatially selective RF pulses and the effects of digitization on their performance.* Magnetic resonance in medicine, 1993. **30**(6): p. 732-740.
158. Tkáč, I., et al., *In vivo 1H NMR spectroscopy of rat brain at 1 ms echo time.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 1999. **41**(4): p. 649-656.
159. Lopez-Kolkovsky, A.L., S. Mériaux, and F. Boumezbeur, *Metabolite and macromolecule T1 and T2 relaxation times in the rat brain in vivo at 17.2 T.* Magnetic resonance in medicine, 2016. **75**(2): p. 503-514.
160. Wilson, M., et al., *Methodological consensus on clinical proton MRS of the brain: Review and recommendations.* Magnetic resonance in medicine, 2019. **82**(2): p. 527-550.
161. Raschka, S., *Model evaluation, model selection, and algorithm selection in machine learning.* arXiv preprint arXiv:1811.12808, 2018.
162. Hong, S.T., D.Z. Balla, and R. Pohmann, *Determination of regional variations and reproducibility in in vivo 1H NMR spectroscopy of the rat brain at 16.4 T.* Magnetic resonance in medicine, 2011. **66**(1): p. 11-17.
163. Pope, W.B., et al., *Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy.* Journal of neuro-oncology, 2012. **107**(1): p. 197-205.
164. Chandler, M., et al., *MRSNet: Metabolite quantification from edited magnetic resonance spectra with convolutional neural networks.* arXiv preprint

arXiv:1909.03836, 2019.
165. Gal, Y. and Z. Ghahramani, *Bayesian convolutional neural networks with Bernoulli approximate variational inference.* arXiv preprint arXiv:1506.02158, 2015.
166. Bishop, C.M., *Pattern recognition.* Machine learning, 2006. **128**(9).
167. Kreis, R., et al., *Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: Background and experts' consensus recommendations.* NMR in Biomedicine, 2021. **34**(5): p. e4347.
168. Gal, Y., *Uncertainty in deep learning. University of Cambridge*. 2016, PhD Thesis. 2016. URL: http://mlg. eng. cam. ac. uk/yarin/thesis/thesis. pdf ….
169. Der Kiureghian, A. and O. Ditlevsen, *Aleatory or epistemic? Does it matter?* Structural safety, 2009. **31**(2): p. 105-112.
170. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
171. Tieleman, T. and G. Hinton, *Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude.* COURSERA: Neural networks for machine learning, 2012. **4**(2): p. 26-31.
172. Ngiam, J., et al. *Multimodal deep learning*. in *ICML*. 2011.
173. Shen, D., G. Wu, and H.-I. Suk, *Deep learning in medical image analysis.* Annual review of biomedical engineering, 2017. **19**: p. 221-248.
174. Leibig, C., et al., *Leveraging uncertainty information from deep neural networks for disease detection.* Scientific reports, 2017. **7**(1): p. 1-14.
175. Litjens, G., et al., *A survey on deep learning in medical image analysis.* Medical image analysis, 2017. **42**: p. 60-88.
176. Glang, F., et al., *DeepCEST 3T: Robust MRI parameter determination and uncertainty quantification with neural networks—application to CEST imaging of the human brain at 3T.* Magnetic resonance in medicine, 2020. **84**(1): p. 450-466.
177. Herzog, L., et al., *Integrating uncertainty in deep neural networks for MRI based stroke analysis.* Medical Image Analysis, 2020. **65**: p. 101790.
178. Kwon, Y., et al., *Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation.* Computational Statistics & Data Analysis, 2020. **142**: p. 106816.
179. Jang, J., et al., *Unsupervised anomaly detection using generative adversarial networks in 1H-MRS of the brain.* Journal of Magnetic Resonance, 2021. **325**: p. 106936.
180. Kreis, R., *Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts.* NMR in Biomedicine, 2004. **17**(6): p. 361-381.

Abstract in Korean (국문 초록)

두뇌 내 특정한 부위에 대한 대사체들의 종류와 농도 정보를 획득할 수 있는 자기공명분광 (MRS) 분야에서 일반적으로 활용하고 있는 비선형 최소제곱피팅 (Nonlinear least squares fitting; NSLF)은 주어진 사전 정보 (Prior knowledge)에 의존한 정량화 결과 변동 특성을 나타낸다. NLSF 기반한 두뇌 대사체 정량화는 MRS 신호품질에 민감하게 성능 변화를 나타낸다. 무엇 보다, NLSF를 통한 정량화 결과의 신뢰 지표인 크라메르-라오 하한 (Cramer-Rao lower Bound; CRLB)은 정량화 결과에 대한 오차정보를 반영하는 정확도가 아닌, 정밀도를 표현하므로, 이를 주의하여 활용하지 않으면 통계적 편향성을 나타낼 위험이 있다. 이러한 문제들로 인해 MRS는 현재까지도 제한적으로만 임상 활용되고 있는 상황이다.

따라서 본 연구는 자기공명분광법을 이용한 두뇌 대사체 정량화 과정에 있어서 딥 러닝 기술을 접목하여, 정량화 정확도를 개선하는 점에 주 목적을 두고 있다. 구체적으로 본 연구에서는 다음과 같이 두 부분에 대한 방법을 제안하였다.

첫번째로는 깊은 인공신경망을 통해 MRS 신호내의 두뇌 대사체 공명 신호만을 추출하여, 이를 간단한 선형 회귀 후처리를 통해 정량화를 할 수 있는 분석 기술을 개발하였다. 두번째로는 딥 러닝에서 예측하는 결과들에 대한 불확실성 지표를 표현하는 방법에 대해 개발하였다. 구체적으로는 빅데이터 기반의 경험적 불확실성 지표와,

베이지안 접근법에 기반한 정규분포를 따르는 불확실성 지표 표현 방법을 개발하였다.

결과적으로 제안된 방법들은 NLSF 대비 MRS 신호 품질에 덜 영향을 받으면서 낮은 정량화 결과 변동성을 나타내는 동시에, NLSF의 정량화 결과에 대한 신뢰지표인 CRLB에 비해 더 실제 오차와 상관성이 높은 불확실성 지표 성능을 보였다.

따라서 본 연구는, MRS를 활용한 두뇌 대사체 정량화에 대한 정확도 개선을 위해 딥 러닝 기술들을 활용한다면, MRS의 임상 적용 가능성을 높일 수 있음을 시사한다.

주요어: 자기공명분광법 (MRS), 대사체, 딥 러닝

학번: 2014-22018