

Comparison of Convolutional Neural Networks Based Method and LCmodel on the Quantification of in Vivo Magnetic Resonance Spectroscopy

Yu-Long Huang

National Taiwan University of Science and Technology

Yi-Ru Lin

National Taiwan University of Science and Technology

Shang-Yueh Tsai (✉ sytsai@nccu.edu.tw)

National Chengchi University

Research Article

Keywords: Convolutional neural network, magnetic resonance spectroscopy, metabolite concentrations

Posted Date: January 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1190469/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Quantification of metabolites concentrations in institutional unit (IU) for between subject and long-term comparison is considered important strategy in the applications of magnetic resonance spectroscopy (MRS). The aim of this study is to investigate if metabolite concentrations quantified by convolutional neuronal network (CNN) based method associated with a proposed scaling procedure can reflect variations of the metabolite concentrations in institution unit (IU) at different brain regions with different signal-to-noise-ratio (SNR) and linewidth (LW). An error index based on standard error (SE) is proposed to indicate the confidence levels on the prediction for metabolites. *In vivo* MRS spectra were collected at 3 brain regions from 44 subjects at 3T system. Metabolite concentrations in IU quantified by LCModel and CNN from 44 subjects were compared. For *in vivo* spectra characterized under different spectral quality in terms of SNR and LW, line narrowing and noise free spectra were successfully exported by CNN.

Concentrations of five metabolites quantified by CNN and LCModel are in similar range with statistically significant Pearson's correlation coefficients (0.28~0.70). SE of the metabolites show positive correlation with Cramer-Rao lower bound (CRLB) ($r=0.60$) and with absolute CRLB ($r=0.84$). In conclusion, the CNN based method with the proposed scaling procedures can be used to quantify *in vivo* MRS spectra. The concentrations of five major metabolites were reported in IU, which are in the same range as those quantified using a routine MRS quantification procedures by LCModel. The SE can be used as error index indicating predicted uncertainties for metabolites with the information similar to the absolute CRLB.

Introduction

Magnetic resonance spectroscopy (MRS) is a noninvasive tool to measure the biochemical information in the brain. It has been widely applied to study various neurological and psychiatric disorders¹⁻⁴. The challenges on the quantification of MRS are the low signal to noise ratio (SNR), overlapping metabolic signals, contamination of macromolecules, lipids and artifacts⁵⁻⁷. Therefore, comprehensive spectral processing procedures are developed for the quantification of metabolites⁸⁻¹¹. Even though, the development of methods to detect metabolites characterized with low concentration, complex spectral shape, overlapped spectral line is still necessary for MRS at clinical field strength.

Recent growths in the deep learning (DL) algorithms have shown that models based on convolutional neuronal network (CNN) have a variety of applications on MRS¹²⁻¹⁶. Studies have shown that CNN can be used to sort the spectra according to spectral quality^{14,17}. Instead of passively filtering spectra, CNN can be also used to improve the spectra such as removal of artifacts¹³ or improvement of spectral quality^{18,19}. It has been shown that DL method based on CNN model has great potential on extracting highly overlapped metabolites¹⁸, such as γ -aminobutyric acid (GABA), N-Acetylaspartylglutamate (NAAG), Glutamate (Glu) and glutamine (Gln). However, this CNN based method was only demonstrated using relative quantification, by which only ratios between metabolites were reported. It is not known if the spectra processed by CNN based method can be properly quantified in institutional units (IU)^{17,20-22}, which are generally used for between subject and long-term comparison.

Another important property of spectral quantification is to estimate proper error index for each quantified metabolite. The error index is used to exclude unreliable quantified metabolite concentrations in the statistical analysis, which is considered essential in clinical MRS data. In one of the commonly used software packages, LCModel, the error metrics are Cramer-Rao lower bound (CRLB), which indicates the uncertainty in spectral fitting. The CRLB of individual metabolite, associated with SNR and LW, have been already widely used for quality assurance in the analysis of MRS data for decades. A recent study has further suggested that absolute CRLB (abs CRLB) is a better choice than CRLB for quality assurance to avoid possible bias on the statistic test especially in the clinical studies²³. For DL method, an error index based on SNR, LW and signal to background ratio (SBR) had recently been demonstrated to successfully reflect quantification uncertainties as other nonlinear based model fitting methods¹⁸. Here, an error metric estimated from individual metabolite was proposed for the DL based method and we compared the proposed error metrics to CRLB based error index reported by LCModel.

In this study, a CNN based method was implemented to process the spectra¹⁹. To increase the clinical feasibility of this CNN based method, a scaling method was proposed to quantify the processed spectra in IU²⁴. We investigated if metabolite concentrations quantified by the CNN based method can reflect variations of the metabolite concentrations among a group of subjects as those quantified by LCModel in IU. The performance of this CNN based method was evaluated on *in vivo* spectra collected from 3 brain regions for 44 subjects, characterized by different level of linewidth (LW) and SNR. As the multivariable regression was implemented in the CNN based method of this study, we proposed to use the standard error (SE) of each estimator as an error index to indicate the confidence levels on the CNN prediction for metabolites. The performance of SE of each metabolite was evaluated by correlating SE with the CRLB and abs CRLB.

Results

Simulation and CNN prediction

The distributions of SNR and LW of the simulated spectra were shown in the figure 1. For all data (N=50000) and test data (N=5000), SNR distributed in the range of 17~192 (67.2 ± 29.1) and LW ranged from 3.5~19.1 Hz (7.9 ± 2.9 Hz). Representative simulated spectra with different SNR, LW and relative metabolic amplitudes were shown in the figure 2, along with CNN predicted spectra, ground truth (GT) and residue. The capability of CNN on noise reduction and LW narrowing can be clearly seen on the predicted spectra. The predicted spectra are very close to the GT in terms of LW and peak heights. The difference between the predicted spectra and GT shows observable residues at spectral range of major metabolites such as tNAA, Cre, tCho and Glu. The %MSE is 0.65% for spectra with relative low SNR and large LW (SNR=32, LW=15.2 Hz) and is as low as 0.07% for spectra with relative good spectral quality (SNR=126, LW=3.8 Hz). In 10-fold cross validation, the mean and standard deviation of MSE among 10 folds are $2.91 \times 10^{10} \pm 5.58 \times 10^8$ (%MSE=0.20% \pm 0.17%) in the training/validation data (N=45000), and are

$2.47 \times 10^{10} \pm 8.69 \times 10^9$ (%MSE=0.21% \pm 0.19%) in the test data (N=5000). The 1st training model of CNN was adopted for the evaluation of CNN and for *in vivo* spectra.

For the test sets (N=5000), the scatter plots of the predicted metabolic amounts and GT were shown in the figure 3. All metabolites show positive correlations between the predicted metabolic amounts and GT. For all metabolites except NAAG, Gln and GABA, the predicted values spread evenly over the GT line (slope=1) with the Pearson's correlation coefficient, *r*, in the range of 0.92 ~0.99. For NAAG and Gln, the predicted values spread over a larger range than other metabolites associate with lower correlations (*r*=0.73 for Gln, *r*=0.81 for NAAG). For GABA, the predicted values only spread in a small metabolic range (1.3~1.7 mM) compared to GT (1.0~2.0mM), leading to low correlation (*r*=0.21). The mean absolute percentage errors (MAPEs) of all metabolites are shown in the figure 4. Overall, the metabolites with MAPEs less than 5% are NAA (2.2% \pm 1.8%), Cr (1.9% \pm 1.6%), Glu (4.8% \pm 3.9%), tNAA (2.1% \pm 1.7%), tCho (1.6% \pm 1.5%), Glx (3.1% \pm 2.3%), ml (2.7% \pm 2.3%). They are major metabolites typically considered quantifiable on spectra from the clinical MRI system. The metabolites with MAPEs around 10% are Gln (9.1% \pm 7.3%), GPC (9.4% \pm 9.7%), PCh (9.4% \pm 10.3%). The MAPEs over 10% are GABA (16.9% \pm 11.8%), NAAG (19.2% \pm 23.3%), Asp (16.2% \pm 11.1%), Glc (15.5% \pm 10.7%). The mean absolute errors (MAEs) of these metabolites were shown in figure S4. The distribution of MAPEs of the metabolites at different SNR and LW range were shown in figure S5 and figure S6. The mean and standard deviation of MAPE of all metabolites except GABA increase as decreasing SNR and as increasing LW. Based on the observation on MAPEs, this CNN model is suggested to process the spectra with LW < 13 Hz and with SNR > 25.

In vivo results

The fitting of metabolite, macromolecules (MMs) and baseline using LCModel for all subjects were shown in figure 5. Using default LCModel setting, the MMs at 0.9~3.11 ppm can be well fitted and the fitted MMs is similar to those in the previous reports^{25,26}. For the baselines estimated by spline, we found relatively large distortion across subjects in 3.5 ~ 4.0 ppm, which account for the residue water signals and contribution of MM37, MM38 and MM40 listed in table S2. Representative *in vivo* spectra fitted by LCModel from 3 brain regions were shown figure S1. Among three regions, DACC shows lower variations in fitted lines across subjects, where DSTRI shows largest variations, especially in the fitted baseline. The SNR and LW calculated based on NAA peak and by LCModel were summarized in Table 1. Among these 3 brain regions, DACC located in a relative homogeneous brain region has lowest LW, and DSTRI located in the deep brain has largest LW and lowest SNR. DLPFC with largest VOI has highest SNR among three regions. Representative *in vivo* spectra under different SNR and LW were shown in the figure 6 with the corresponding CNN predicted spectra. *In vivo* spectra exported as line narrowing, noise free spectra can be clearly seen on the CNN predicted spectra.

Table 1

SNR and LW of *in vivo* spectra for 3 brain regions from 44 subjects. SNR_{NAA} and LW_{NAA} were calculated in the frequency domain based on the peak of N-acetyl Aspartate (NAA) at 2.0 ppm. Signal is defined as the peak height of NAA without removing baseline and overlapping peaks. Noise is defined as the standard deviation of spectra in the range of -1.0 ~ -3.0 ppm. The LW of spectra is defined as full width half maximum of the NAA peak. $SNR_{LCModel}$ and $LW_{LCModel}$ were exported from LCModel. Note, despite different SNR and LW values exported by NAA based calculation and by LCModel, same spectral quality have been found for 3 regions. DSTRI have largest LW and lowest SNR among 3 regions. DACC has lowest LW and DLPFC has highest SNR among 3 regions.

	LW_{NAA}	$LW_{LCModel}$	SNR_{NAA}	$SNR_{LCModel}$
DACC	4.3~13.4 (6.5±1.8)	3.5~10.4 (5.6±1.5)	91~259 (162.8±34.4)	20~43 (32.5±5.0)
DLPFC	4.9~10.6 (7.0±1.3)	3.9~8.7 (6.4±1.1)	147~346 (219.4±43.0)	17~48 (36.4±6.9)
DSTRI	6.0~16.9 (8.6±2.0)	4.9~9.7 (6.7±1.1)	49~143 (90.5±21.2)	10~33 (19.8±5.6)

Metabolite concentrations quantified using CNN and LCModel, with corresponding SEs and CRLBs were summarized in Table 2. The between subject variations were calculated as standard deviation divided by mean from all subjects. Overall, the concentrations of CNN and LCModel are in similar range for tNAA, Cr, tCho, ml, Glu. Metabolite concentrations of CNN show smaller between subject variations than those of LCModel. In CNN quantification, Gln are at the level of 2~3 IU, which is around half of Glu. For LCModel, the Gln are in the range of 5~ 7 IU, which leads to discrepancy in the Glx concentrations by LCModel and by CNN. The SEs were in the order of 10^{-6} ~ 10^{-7} . For 3 brain regions, CRLBs are less than 5% for three singlets (tNAA, Cr, tCho) and less than 10% for ml and Glx. Scatter plots of metabolite concentrations quantified by CNN and by LCModel were shown in figure 7. For all metabolites, positive correlations were found between two methods ($r=0.28$ ~ 0.70). The concentrations of major metabolite except Gln and Glx from CNN and LCModel distributed over similar range. Scatter plots of SE versus CRLB and abs CRLB were shown in figure 8. Both CRLB and abs CRLB versus SE show positive correlation. The correlation coefficient between SE and abs CRLB ($r=0.84$) is higher than that between SE and CRLB ($r=0.60$).

Table 2

Metabolite concentrations of tNAA, Cr, tCho, ml, Glu, Gln, Glx quantified using CNN and LCModel from 3 brain regions. Concentrations are shown as mean \pm standard deviation (standard deviation/mean) from 44 subjects. Metabolites concentrations were reported in institutional unit (IU). The standard error (SE) of estimators were estimated in multiple linear regression. SE is less than the CNN predicted concentrations by an order of 10^{-7} . CRLBs were exported by LCModel. N is the number of subjects included in the calculation of CRLBs using the criterion: CRLB > 10% for tNAA, Cr, tCho; CRLB > 20% for ml, Glu, Gln, Glx.

		Concentration (IU)				
		CNN	LCModel	SE($\cdot 10^{-7}$)	CRLB	N
DACC	tNAA	12.12 \pm 1.35 (11.1%)	13.20 \pm 1.57 (11.9%)	2.50 \pm 0.61	2.16% \pm 0.48%	44
	Cr	9.23 \pm 0.89 (9.6%)	9.81 \pm 1.16 (11.8%)	1.80 \pm 0.44	2.05% \pm 0.21%	44
	tCho	2.53 \pm 0.27 (10.7%)	2.42 \pm 0.39 (16.0%)	0.74 \pm 0.19	2.48% \pm 0.55%	44
	ml	3.79 \pm 0.50 (13.2%)	4.54 \pm 0.55 (12.2%)	4.11 \pm 0.99	4.34% \pm 0.64%	44
	Glu	5.29 \pm 0.56 (10.6%)	5.91 \pm 0.85 (14.5%)	7.06 \pm 1.71	9.11% \pm 1.20%	44
	Gln	2.08 \pm 0.24 (11.6%)	6.22 \pm 1.09 (17.5%)	7.65 \pm 1.85	10.18% \pm 1.37%	44
	Glx	7.37 \pm 0.74 (10.0%)	12.13 \pm 1.58 (13.1%)	8.27 \pm 2.00	6.34% \pm 0.83%	44
DLPFC	tNAA	11.87 \pm 1.56 (13.1%)	10.52 \pm 1.36 (13.0%)	2.51 \pm 0.60	2.45% \pm 0.66%	44
	Cr	8.81 \pm 0.94 (10.7%)	7.60 \pm 1.05 (13.9%)	1.81 \pm 0.43	2.23% \pm 0.42%	44
	tCho	1.95 \pm 0.24 (12.4%)	1.61 \pm 0.24 (14.8%)	0.74 \pm 0.20	3.02% \pm 0.59%	44
	ml	3.44 \pm 0.69 (20.2%)	3.19 \pm 0.58 (18.2%)	4.11 \pm 0.99	5.30% \pm 1.50%	44
	Glu	4.69 \pm 0.50 (10.7%)	4.11 \pm 0.66 (16.1%)	7.06 \pm 1.71	12.25% \pm 3.70%	43
	Gln	1.65 \pm 0.30 (18.4%)	5.56 \pm 1.04 (18.8%)	7.66 \pm 1.75	11.09% \pm 3.62%	42
	Glx	6.37 \pm 0.65 (10.3%)	9.52 \pm 1.63 (17.1%)	8.27 \pm 1.99	7.80% \pm 2.50%	44
DSTRI	tNAA	9.52 \pm 1.07 (11.2%)	8.76 \pm 1.70 (19.4%)	2.52 \pm 0.58	3.55% \pm 1.21%	44
	Cr	8.13 \pm 0.52 (6.3%)	7.97 \pm 0.95 (12.0%)	1.82 \pm	2.48% \pm 0.66%	44

			0.42		
tCho	2.17 ± 0.18 (8.3%)	1.95 ± 0.23 (11.8%)	0.75 ± 0.17	3.09% ± 0.47%	44
ml	3.64 ± 0.51 (14.1%)	3.40 ± 0.58 (16.9%)	4.14 ± 0.95	6.27% ± 1.76%	44
Glu	4.83 ± 0.53 (11.0%)	4.95 ± 0.80 (16.1%)	7.11 ± 1.63	11.93% ± 2.94%	44
Gln	1.84 ± 0.31 (16.7%)	5.61 ± 0.90 (16.0%)	7.71 ± 1.83	14.14% ± 5.68%	40
Glx	6.73 ± 0.77 (11.4%)	10.35 ± 1.57 (15.1%)	8.33 ± 1.87	8.95% ± 2.77%	44

For GABA and NAAG, fitting on these two metabolites using LCMoel was not reliable (CRLB > 20% for GABA; CRLB>40% for NAAG) and NAAG can only be quantified in about half of the subjects (Table 3). As for CNN, consistent GABA concentrations were quantified with the between subject variation around 10% in 3 brain regions. Low NAAG concentrations were reported by CNN (< 1 IU) with large between subject variations (>50%).

Table 3

Metabolite concentrations of GABA and NAAG quantified using CNN and LCMoel from 3 brain regions. Concentrations are shown as mean ± standard deviation from all subjects. Metabolites concentrations were reported in institutional unit (IU). The standard error (SE) of estimators were estimated in the multiple linear regression. CRLBs were exported by LCMoel. N is the number of subjects included in the calculation of CRLBs when LCMoel can fit the metabolites (CRLB ≠ 999%).

		Concentration (IU)				
		CNN	LCMoel	SE(·10 ⁻⁷)	CRLB	N
GABA	DACC	1.34 ± 0.13	2.00 ± 0.63	2.95 ± 0.71	25.98% ± 11.88%	44
	DLPFC	1.19 ± 0.11	1.56 ± 0.64	2.83 ± 0.72	36.07% ± 31.98%	43
	DSTRI	1.08 ± 0.08	1.48 ± 0.62	3.06 ± 0.67	42.37% ± 27.72%	43
NAAG	DACC	0.55 ± 0.34	0.82 ± 0.47	5.63 ± 1.36	90.79% ± 130.62%	24
	DLPFC	0.21 ± 0.60	0.85 ± 0.77	5.41 ± 1.37	115.56% ± 136.92%	9
	DSTRI	0.06 ± 0.27	1.49 ± 0.53	5.85 ± 1.28	40.48% ± 63.44%	29

Discussion

In this study, the feasibility of a CNN based method on the quantification of metabolite concentrations was demonstrated on *in vivo* MRS data collected at 3 brain regions from 44 subjects. With the proposed scaling procedure, we have shown that this CNN based method can quantify the metabolites

concentrations in IU with the values of 5 major metabolites tNAA, Cr, tCho, ml and Glu in compatible range with those quantified by LCModel (Table 2). The between subject variations of these metabolite are also in compatible range with the previous studies²⁷⁻²⁹. For the CNN model used in this study, the optimized parameters such as initial learning rate, decay factor, L2 regularization etc. differ from the model used in the previous study^{18,19}, which implies that this CNN based method can be adapted for different MRS protocols and in vivo data sets after optimization of the models.

For the subject by subject comparison between CNN and LCModel, positive correlation ($p < 0.005$) were found for these metabolites (figure 7). The findings on *in vivo* results are in agreement with the findings in the evaluation of CNN model using the test set, where these metabolites showed good prediction compared to GT (figure 3) and they had lower MAPEs (<5%) compared with other metabolites (figure 4). Discrepancies were found on the quantification of *in vivo* Gln concentrations between CNN and LCModel. For the CNN, the prediction of Gln show larger errors (Table 2) and lower correlation to GT (figure 3) compared to other 5 major metabolites. As for LCModel, the separation of Glu and Gln on the spectra collected at 3T is a well-known difficulty. Therefore, if the CNN can serve a better tool to differentiate Gln from Glu on 3T system as claimed by the previous studies¹⁹ needs further investigation. It is necessary to have a reliable quantification strategy on Gln. As Glx is the combination of Glu and Gln, Glx by CNN also differs from that by LCModel. Despite the discrepancy in concentrations, Glx has similar within subject variations as Glu (Table 2).

The metabolic fitting was carried out using multivariable linear regression and the SE of each estimator is proposed as an error index to indicate the confidence level on quantified metabolic signal from the predicted spectra. For all subjects, SEs of the metabolites were lower than the quantified concentrations by an order of 10^{-7} , which implies that the predicted spectra were very close to the basis in terms of spectral shape and LW. Even though, we found higher SEs on metabolites with complex spectral lines (Glu, Gln and Glx) than SEs of three singlets (NAA, Cr, tCho), which is in accordance with the distribution of CRLBs (Table 2). A further examination using the correlation shows that SE has higher correlation ($r=0.84$) with the abs CRLB than that with CRLB ($r=0.60$) (figure 8). This indicates that SEs could potentially serve as an error index indicating the confidence levels on the CNN prediction with information close to abs CRLB. For CNN based quantification, it is necessary to accumulate more data to set up a quality assurance standard in terms of SE, SBR, LW and SNR in further investigation. This could be helpful for the possible applications of CNN based MRS quantification.

One interested issue on the CNN based MRS quantification is the possibility to differentiate metabolites with low concentrations from the overlapping metabolites, such as NAAG and GABA. These two metabolites are not quantifiable using LCModel (Table 3). In the evaluation of CNN model, NAAG shows largest prediction errors among all metabolites in the basis (MAPEs=19.2%, $r=0.81$). For *in vivo* data, low NAAG concentrations were reported (less than 1 IU) with within subject variations over 50%. Therefore, we think differentiation of NAAG from NAA is not successful using CNN. Similar findings have been reported with the MAPEs of NAAG over 30% at 3T¹⁹ and around 25% at 9.4T¹⁸. As for GABA, consistent GABA

concentrations were quantified (1.0~1.3 I.U.) with about 10% within subject variation for *in vivo* data. However, according to the evaluation of CNN model, the GABA show relative large prediction errors (MAPEs=16.9%) across all groups of SNR and LW (figure S5 and figure S6). At 3.0T, the MAPEs of GABA was reported to be around 20%¹⁹, which is similar to our results. A low slope of GABA ($r=0.21$) indicates that CNN is not able to catch the variations of GABA concentrations but exported GABA concentrations at a constant level (figure 3). Although good GABA prediction has been shown in a previous study¹⁸, it is reasonable as the results were reported at 9.4T. As GABA is a very important neurotransmitter, a further modification of this CNN model can thus be expected to improve the quantification of GABA at 3T. The possibility to quantify GABA on the spectra collected with PRESS sequence instead of the spectra editing sequence³⁰⁻³² will be a great advantage for the CNN based MRS method.

The limitation of this study is on the estimation of MMs on *in vivo* spectra. Recently, it is suggested that experimental MM model can improve the accuracy and stability on the estimation of MMs for *in vivo* spectra^{25,26}. However, experimental MM models were not available in our site. Therefore, default LCModel setting was adopted, where only MMs estimated in the spectral range less than 3.2 ppm can be acceptable. The MMs at 3.5 to 4.0 ppm were then handled only by the spline baseline fitting. Large variations in baseline at this spectral range were reported in the previous reports^{5,25,26}, which can be also seen in our results (figure 5 and figure S1). The uncertainty in the estimation of MMs can lead to incorrect metabolite-to-MM ratio even we have corrected the contribution of MMs according to their relative amplitude (table S2). Nevertheless, the simulated MMs actually vary in amplitude and metabolite-to-MM ratio also alters in a range in the preparation of simulated spectra. CNN model is therefore trained using the simulated spectra with a wide range of MMs. We think the CNN model used in this study can still deal with *in vivo* spectra even under some bias on the estimation of metabolite-to-MM ratio. Another issue is that *in vivo* spectra fitted with default LCModel setting have significant influence in the estimation of baselines and MMs, which leads to errors in the quantified concentrations²⁶, especially for ml, Glu, Gln. Among 7 metabolites shown in figure 7, the low correlation coefficients were indeed found in ml and Gln (0.29 and 0.28). The proper model on MMs and baseline setting in LCModel fitting can certainly improve the metabolite quantification using LCModel, which should be enrolled in the future study.

In conclusion, we have shown that CNN based method have potential to serve as spectral processing tool as LCModel on the quantification of *in vivo* MRS spectra. The proposed scaling procedure can be integrated with the CNN model to report the metabolite concentrations in IU, which are in the same range as those quantified using a routine MRS quantification procedures, LCModel. The SE of the estimator can be used as error index indicating predicted uncertainties for metabolites with the information similar to abs CRLB. The ability to isolate GABA and Gln needs further investigation with improved MM models and baseline fitting in LCModel.

Materials And Methods

Simulation of brain spectra

For the training and testing of the CNN, spectrum of each metabolite was extracted from the vendor provided basis in LCMoel ("press_te30_3t_01a.basis"). The basis was generated using PRESS sequence at 3T with parameters: TE=30 msec., TR=10 sec., bandwidth=2500 Hz, sample points =4096. The basis includes 15 metabolites: alanine (Ala), aspartate (Asp), creatine (Cr), γ -aminobutyric acid (GABA), glucose (Glc), glutamine (Gln), glutamate (Glu), glycerophosphocholine (GPC), phosphocholine (PCh), lactate (Lac), myo-inositol (ml), N-acetyl-aspartate (NAA), N-acetyl-aspartylglutamate (NAAG), scyllo-Inositol (Scyllo), and taurine (Tau). Scyllo and Tau were excluded due to limited information about *in vivo* concentrations in the brain. The concentrations of metabolites are 200 mM for GABA, Glc, ml, Tau; 100 mM for Ala, Gln, Lac; 66.69 mM for Scyllo; 50 mM for Cr, Glu, NAA; 40 mM for Asp; 24.73mM for GPC; 20.23mM for PCh; and 18.5mM for NAAG.

The simulation of brain spectra was carried out by varying metabolite concentrations, levels of macromolecule (MM), spectral linewidth (LW), and SNR using in-house script written in Python (v3.6). Firstly, simulated brain spectra (N=50000) were generated by combining 13 simulated metabolites with the concentrations evenly distributed between upper and lower bounds of metabolite concentrations listed in table S1^{19,33}. These metabolite-only spectra were used as the ground truth (GT) in the training and testing of CNN. Secondly, MMs were simulated using 13 Gaussian model functions at spectral range of 0.5 ~4.5 ppm^{5,34}. The simulation parameters of MMs are listed in table S2. The MMs was generated by varying relative amplitude and LW of 13 MMs within 10% and 20% range, respectively. MMs were added to the metabolite-only spectra according to metabolite-to-MM ratio, which is determined by the *in vivo* results of LCMoel using default setting (figure S1). MMs from LCMoel accounts for the contribution of MMs at range from 0.90 ppm to 3.11ppm²⁵. The total contribution of MM was then corrected for the contribution of MMs in 3.5 ppm to 4.0 ppm. The metabolite-to-MM for all subjects in three brain regions is 2.2~4.2 (3.2 \pm 0.58). In the simulation, the MMs were added to the metabolite-only spectra by varying the metabolite-to-MM ratio evenly distributed in the range of 2.2~4.2. Finally, LW broadening was applied by varying time constant of the exponential function such that the LW of the spectra was in the range of 3.5 ~19.1 Hz (7.9 \pm 2.9 Hz). Random noise was added to the spectra such that SNR of the spectra was in the range of 17~ 192 (67.2 \pm 29.1). The LW of spectra was calculated as full width half maximum of the NAA peak. SNR was calculated in the frequency domain. Signal is defined as the peak height of NAA without removing baseline and overlapping peaks. Noise is defined as the standard deviation of spectra in the range of -1.0 ~ -3.0 ppm. The simulated spectra were cropped at spectra range 0.5 ~ 4.2 ppm for the input of CNN. As the simulated spectrum is 2500 Hz (~19.57 ppm), the cropped simulated spectra corresponds to 774 spectra points (4096 \times 3.7/19.57).

Design and evaluation of CNN

The CNN model was constructed in python (v3.6); Keras (v2.2.4) with Tensorflow (v1.13.1, CUDA 10.1) on graphic processing unit (GPU; NVIDIA Titan RTX). In brief, CNN consists of one input layer, three convolution blocks, two max pooling layer, one fully connected layer with linear activation. Each convolution block was composed of 4 repetitions of convolution filter, batch normalization and activation layer^{18,19}. The size of convolution filter is 15x1 with pad size of 7. The number of convolution filter is 32

for 1st block, 64 for 2nd block and 128 for 3rd block. The rectified linear unit (ReLU) function was used in the convolution blocks as activation. The parameters for CNN structure and training were initially set as the implementation in other studies¹⁹. We optimized these parameters based on the loss of training and validation. For max pooling, the size and stride is 2x1 and 2. Stochastic gradient descent with momentum (SGDM) algorithm was used for the training with initial learning rate of 0.01 and momentum of 0.8. The loss function was mean square error (MSE) with L2 regularization parameter of 0.001. The schematic of the CNN is shown in figure S2. The batch size is 32 and maximum epoch is 110. The learning rate is reduced on plateau by a decay factor of 0.8 with a minimum of 10^{-7} when the validation loss did not improve through 5 epochs. The training time is around 70 minutes.

The evaluation of CNN was performed using 10-fold cross validation. Simulated brain spectra (N=50000) were randomly split into 10 groups. Each group was assigned as a test set (N=5000) and remaining spectra (N=45000) were assigned as validation set (N=4500) and training set (N=40500). The dimension of the input data is $40500 \times 774 \times 1$ (array \times height \times width) for the training set. The CNN was trained using training set and validation set. Errors between the GT and CNN predicted spectra on the test set were calculated. The performance of CNN model was evaluated by MSE between the predicted spectra and GT, and %MSE is the MSE normalized by mean square values of GT spectra. In addition, the accuracy on the estimation of the metabolic amounts was evaluated using mean absolute error (MAE) and mean absolute percentage error (MAPE) between CNN predicted metabolic amounts and GT.

In vivo Experiments

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Before being included in this study, all participants gave their informed consent to undergo a protocol, which was approved by the Research Ethics Committee of National Chengchi University.

MRS experiments were performed on 44 healthy volunteers using a 3T scanner (Skyra, SIEMENS Medical Solutions, Erlangen, Germany) with a 64-channel head coil array. T₁-weighted images, using a gradient echo sequence (TR/TE/flip angle: 250ms/2.61ms/70°; FOV: 256 x 256 mm²; MAT: 128 x 128), were acquired for localizing the MRS volume of interest (VOI) and for segmentation of tissue types. An automatic localization toolbox developed by ourselves³⁵ was used for the localization of VOIs. In summary, a VOI was defined on the template in standard space. A spatial normalization was performed between the template and subject's T1 images and the transformation matrix was applied to transform designed VOI from the standard space to the subject space³⁶. Three VOIs were defined in dorsal anterior cingulate cortex (DACC), dorsal striatum (DSTRI), dorsal lateral prefrontal cortex (DLPFC) in the left hemisphere. The size of the VOIs are 30×15×25 mm³ for DACC, 30×20×20 mm³ for DSTRI, 35×20×25 mm³ for DLPFC. MRS data were acquired using PRESS sequence with the following parameters: TR = 2 s, TE = 30 ms, flip angle = 90°, sample point = 2048, bandwidth = 2000 Hz, NEX = 128. A non-water

suppressed (NWS) reference scan was acquired using 4 averages as the reference for phase correction, calibration of metabolic signal. The total acquisition time for each subject is around 50 minutes.

Scaling the spectra for CNN model

For *in vivo* data, zero order phase correction was applied. All spectra were resampled to 2500 Hz using linear interpolation in the spectral domain, and then zero-padding to 4096 points in the time domain. The water scaling was performed before the input to CNN. The scaling factor (Fac) for the simulated spectra and *in vivo* spectra is calculated according to the known concentration of the simulated spectrum of NAA

$$Fac = \frac{Conc_{water} NH_{water}}{Conc_{NAA} NH_{NAA}} S_{NAA} \frac{1}{S_{water-invivo}} \quad [1]$$

Where S_{NAA} is the area of NAA peak at 2.0 ppm in the basis; $Conc_{water}$ is the concentration of water, which is set to 55556 mM; $Conc_{NAA}$ is the NAA concentration of the basis, which is 50 mM; NH_{water} and NH_{NAA} is the number of protons of water and NAA, which is 2 and 3; $S_{water-invivo}$ is water area of *in vivo* spectra, which can be calculated from the NWS spectra. *In vivo* spectra were multiplied by the Fac, and cropped at spectra range 0.5-4.2 ppm (774 spectral points) for the input of CNN. The diagram of water scaling procedures and CNN for *in vivo* data is illustrated in figure S3.

Quantification of metabolite concentrations

The CNN predicted spectra can be taken as linear combination of the spectra in the basis without line broadening, noise and baseline. The amounts of 13 metabolites were solved using multiple linear regression:

$$Y = X\beta \quad [2]$$

Where Y is a 774×1 matrix containing CNN predicted spectra; X is a 13×774 matrix containing the spectra of 13 metabolites in the basis; β is a 13×1 matrix representing relative amounts of 13 metabolites. These estimated metabolite amounts, β , were scaled to metabolite concentrations according to known metabolite concentrations in the basis and the standard error (SE) of the estimated β was exported. *In vivo* spectra were quantified by LCModel (<http://s-provencher.com/lcmodel.shtml>) using the same basis as CNN. Spectra were fitted between 0.5 ~ 4.2 ppm. Metabolic signals were calibrated to NWS data using the water scaling method. Partial volume and relaxation correction was applied for quantified metabolites from both CNN and LCModel as described in the previous study^{1,31}. In brief, tissue probability maps of gray matter, white matter and cerebral spinal fluids were generated using the segmentation toolbox provided by SPM8 (www.fil.ion.ucl.ac.uk/spm) on T1 images. Corrections for water density in tissues and for relaxation effects were carried out using the published parameters²². The metabolite concentrations were reported in IU. Here, the comparison between CNN and LCModel was made on major metabolites typically reported on the spectra collected at clinical field strength. These metabolites are:

tNAA=NAA+NAAG; Cr; tCho= GPC+PCh; ml; Glu; Gln; Glx= Glu+Gln. The metabolites were excluded for the comparison using the Cramer-Rao lower bounds (CRLB) with the following criterion: CRLB>10% for tNAA, Cr, tCho; CRLB > 20% for ml, Glu, Gln, Glx. To further examine the potential implications of SE, SE was correlated with the CRLB and abs CRLB^{23,37}, which is calculated as the CRLBs multiplied by the quantified metabolite concentrations.

Declarations

Acknowledgements

The authors thank the Taiwan Mind & Brain Imaging Center (TMBIC) and National Chengchi University for the instrument availability. The TMBIC is supported by the Ministry of Science and Technology, Taiwan. This work was supported in part by grants from the Ministry of Science and Technology, Taiwan (MOST 108-2314-B-004-001-MY3, MOST 107-2221-E-011-053)

Conflict of interest

All authors have no conflict of interest to declare

Author contribution

S.-Y. T. conceived and designed the experiments. Y.-L. H. and S.-Y. T. performed the experiments. Y.-L. H. and Y.-R. L. analyzed the data. S.-Y. T, Y.-L. H. and Y.-R. L interpreted the data. S.-Y. T. and Y.-L. H. wrote the main manuscript text and prepared the tables and figures. S.-Y. T. and Y.-R. L revising the manuscript critically for important intellectual content. All authors reviewed the manuscript and approved the final version.

References

- 1 Niddam, D. M. *et al.* Brain metabolites in chronic migraine patients with medication overuse headache. *Cephalalgia*, 333102420908579, doi:10.1177/0333102420908579 (2020).
- 2 Niddam, D. M., Wang, S. J. & Tsai, S. Y. Pain sensitivity and the primary sensorimotor cortices: a multimodal neuroimaging study. *Pain* **162**, 846-855, doi:10.1097/j.pain.0000000000002074 (2021).
- 3 Niddam, D. M. *et al.* Neurochemical changes in the medial wall of the brain in chronic migraine. *Brain* **141**, 377-390, doi:10.1093/brain/awx331 (2018).
- 4 Schur, R. R. *et al.* Brain GABA levels across psychiatric disorders: A systematic literature review and meta-analysis of (1) H-MRS studies. *Human brain mapping* **37**, 3337-3352, doi:10.1002/hbm.23244 (2016).

- 5 Birch, R., Peet, A. C., Dehghani, H. & Wilson, M. Influence of macromolecule baseline on (1) H MR spectroscopic imaging reproducibility. *Magn Reson Med* **77**, 34-43, doi:10.1002/mrm.26103 (2017).
- 6 Tsai, S. Y., Lin, Y. R., Lin, H. Y. & Lin, F. H. Reduction of lipid contamination in MR spectroscopy imaging using signal space projection. *Magn Reson Med* **81**, 1486-1498, doi:10.1002/mrm.27496 (2019).
- 7 Tkac, I. *et al.* Water and lipid suppression techniques for advanced (1) H MRS and MRSI of the human brain: Experts' consensus recommendations. *NMR Biomed*, e4459, doi:10.1002/nbm.4459 (2020).
- 8 Jiru, F., Skoch, A., Wagnerova, D., Dezortova, M. & Hajek, M. jSIPRO - analysis tool for magnetic resonance spectroscopic imaging. *Comput Methods Programs Biomed* **112**, 173-188, doi:10.1016/j.cmpb.2013.06.018 (2013).
- 9 Borbath, T., Murali-Manohar, S., Dorst, J., Wright, A. M. & Henning, A. ProFit-1D-A 1D fitting software and open-source validation data sets. *Magn Reson Med* **86**, 2910-2929, doi:10.1002/mrm.28941 (2021).
- 10 Near, J. *et al.* Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR Biomed*, e4257, doi:10.1002/nbm.4257 (2020).
- 11 Edden, R. A., Puts, N. A., Harris, A. D., Barker, P. B. & Evans, C. J. Gannet: A batch-processing tool for the quantitative analysis of gamma-aminobutyric acid-edited MR spectroscopy spectra. *J Magn Reson Imaging* **40**, 1445-1452, doi:10.1002/jmri.24478 (2014).
- 12 Nagaraja, B. H. *et al.* Tensor-Based Method for Residual Water Suppression in (1)H Magnetic Resonance Spectroscopic Imaging. *IEEE Trans Biomed Eng* **66**, 584-594, doi:10.1109/TBME.2018.2850911 (2019).
- 13 Kyathanahally, S. P., Doring, A. & Kreis, R. Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy. *Magn Reson Med* **80**, 851-863, doi:10.1002/mrm.27096 (2018).
- 14 Gurbani, S. S. *et al.* A convolutional neural network to filter artifacts in spectroscopic MRI. *Magn Reson Med* **80**, 1765-1775, doi:10.1002/mrm.27166 (2018).
- 15 Das, D., Coello, E., Schulte, R. & Menze, B. in *In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention* 462– 470 (Quebec, Canada, 2017).
- 16 Hyun, C. M., Kim, H. P., Lee, S. M., Lee, S. & Seo, J. K. Deep learning for undersampled MRI reconstruction. *Phys Med Biol* **63**, 135007, doi:10.1088/1361-6560/aac71a (2018).
- 17 Gurbani, S. S., Sheriff, S., Maudsley, A. A., Shim, H. & Cooper, L. A. D. Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting. *Magn Reson Med* **81**, 3346-3357, doi:10.1002/mrm.27641 (2019).

- 18 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* **84**, 1689-1706, doi:10.1002/mrm.28234 (2020).
- 19 Lee, H. H. & Kim, H. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. *Magn Reson Med* **82**, 33-48, doi:10.1002/mrm.27727 (2019).
- 20 Maudsley, A. A. *et al.* Mapping of brain metabolite distributions by volumetric proton MR spectroscopic imaging (MRSI). *Magn Reson Med* **61**, 548-559, doi:10.1002/mrm.21875 (2009).
- 21 Hatami, N., Sdika, M. & Ratiney, H. in *In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention* 467– 475 (Granada, Spain, 2018).
- 22 Gasparovic, C. *et al.* Use of tissue water as a concentration reference for proton spectroscopic imaging. *Magn Reson Med* **55**, 1219-1226 (2006).
- 23 Kreis, R. The trouble with quality filtering based on relative Cramer-Rao lower bounds. *Magn Reson Med* **75**, 15-18, doi:10.1002/mrm.25568 (2016).
- 24 Huang, Y., Lin, Y., Huang, T., Ko, C. & Tsai, S. in *Proceedings of International Society for Magnetic Resonance in Medicine* (Virtual Meeting, 2020).
- 25 Marjanska, M. & Terpstra, M. Influence of fitting approaches in LCModel on MRS quantification focusing on age-specific macromolecules and the spline baseline. *NMR Biomed*, e4197, doi:10.1002/nbm.4197 (2019).
- 26 Giapitzakis, I. A., Borbath, T., Murali-Manohar, S., Avdievich, N. & Henning, A. Investigation of the influence of macromolecules and spline baseline in the fitting model of human brain spectra at 9.4T. *Magn Reson Med* **81**, 746-758, doi:10.1002/mrm.27467 (2019).
- 27 Dhamala, E. *et al.* Validation of in vivo MRS measures of metabolite concentrations in the human brain. *NMR Biomed* **32**, e4058, doi:10.1002/nbm.4058 (2019).
- 28 Bednarik, P. *et al.* Feasibility and reproducibility of neurochemical profile quantification in the human hippocampus at 3 T. *NMR Biomed* **28**, 685-693, doi:10.1002/nbm.3309 (2015).
- 29 Chiu, P. W. *et al.* Metabolic changes in the anterior and posterior cingulate cortices of the normal aging brain: proton magnetic resonance spectroscopy study at 3 T. *Age (Dordr)* **36**, 251-264, doi:10.1007/s11357-013-9545-8 (2014).
- 30 Saleh, M. G. *et al.* Multi-vendor standardized sequence for edited magnetic resonance spectroscopy. *NeuroImage* **189**, 425-431, doi:10.1016/j.neuroimage.2019.01.056 (2019).

- 31 Tsai, S. Y., Fang, C. H., Wu, T. Y. & Lin, Y. R. Effects of Frequency Drift on the Quantification of Gamma-Aminobutyric Acid Using MEGA-PRESS. *Scientific reports* **6**, 24564, doi:10.1038/srep24564 (2016).
- 32 Mikkelsen, M. *et al.* Big GABA: Edited MR spectroscopy at 24 research sites. *NeuroImage* **159**, 32-45, doi:10.1016/j.neuroimage.2017.07.021 (2017).
- 33 Govindaraju, V., Young, K. & Maudsley, A. A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed* **13**, 129-153 (2000).
- 34 Opstad, K. S., Bell, B. A., Griffiths, J. R. & Howe, F. A. Toward accurate quantification of metabolites, lipids, and macromolecules in HRMAS spectra of human brain tumor biopsies using LCModel. *Magn Reson Med* **60**, 1237-1242, doi:10.1002/mrm.21496 (2008).
- 35 Peng, P. Y., Tsai, S. Y. & Lin, Y. R. in *In Proceedings of the 24th Annual Meeting of ISMRM* 4009 (Singapore, 2016).
- 36 Park, Y. W. *et al.* AutoVOI: real-time automatic prescription of volume-of-interest for single voxel spectroscopy. *Magn Reson Med*, doi:10.1002/mrm.27203 (2018).
- 37 Oz, G. *et al.* Advanced single voxel (1) H magnetic resonance spectroscopy techniques in humans: Experts' consensus recommendations. *NMR Biomed*, e4236, doi:10.1002/nbm.4236 (2020).

Figures

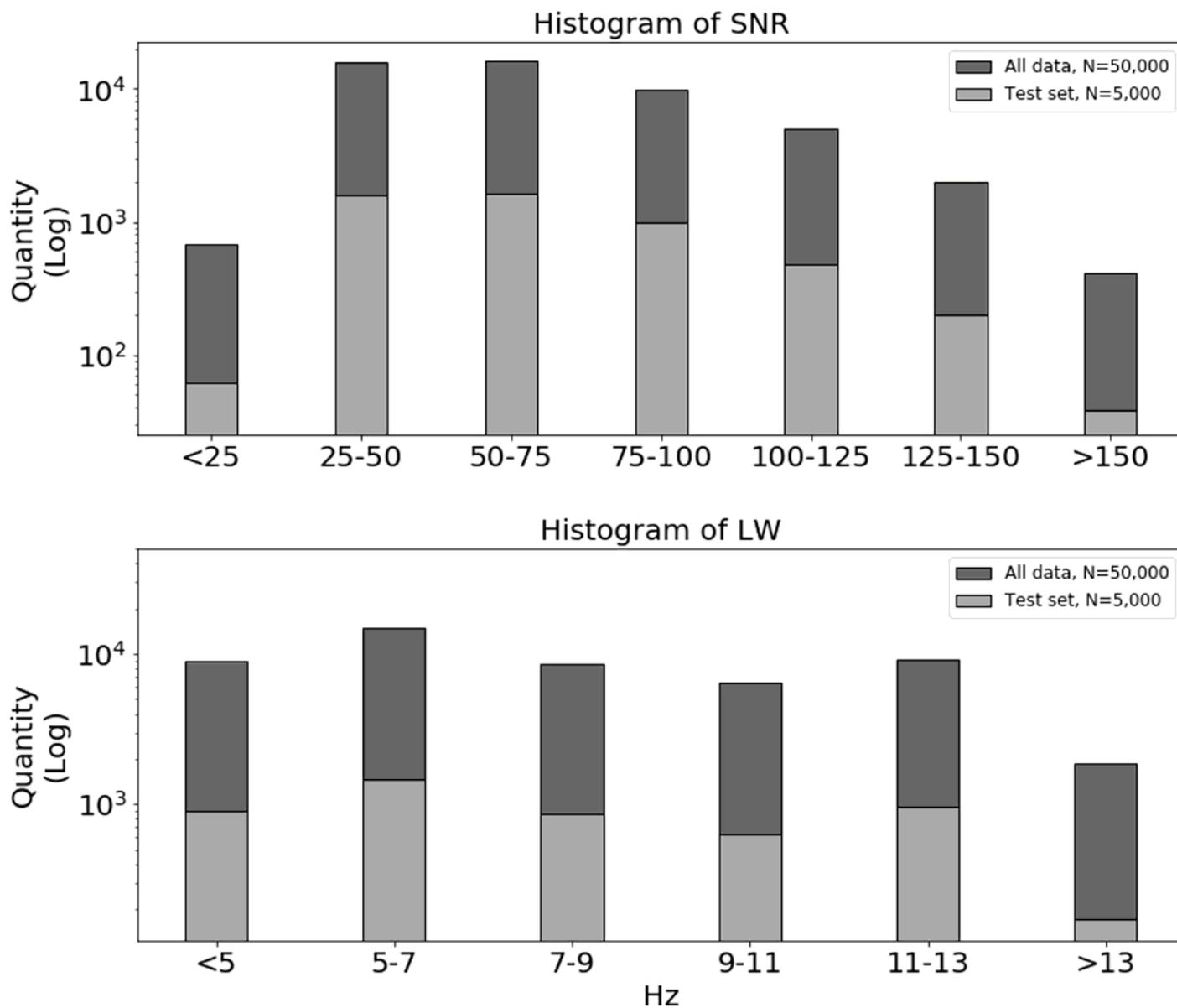


Figure 1

Histogram of SNR and LW of the simulated spectra for CNN training. The distribution of SNR and LW are in the range of 17~192 (67.2 ± 29.1) and 3.5~19.1 Hz (7.9 ± 2.9 Hz) for all data (N=50000). The SNR and LW of data used for test (N=5000) has similar distribution as all data.

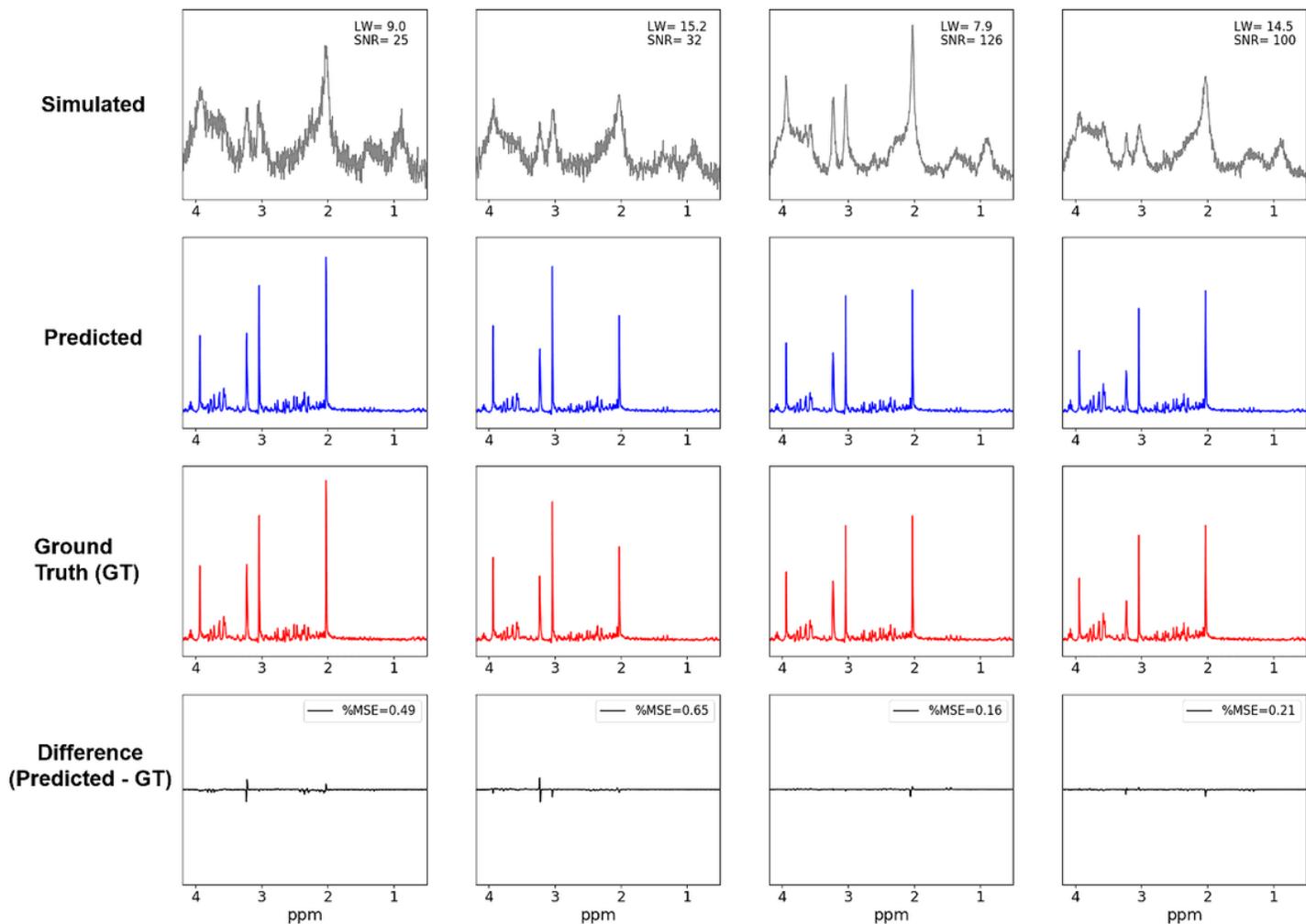


Figure 2

Representative spectra used for the training of CNN model. Simulated spectra (simulated) with different SNR, LW and relative metabolic amplitudes, CNN predicted spectra (predicted), Ground truth (GT), difference of CNN predicted spectra and GT were shown. The predicted spectra are all noise free, line narrowing and very close to the GT with observable residues at spectral range of major metabolites. The largest %MSE (0.65%) is in spectra with low SNR and large LW (SNR=32, LW=15.2 Hz) and lowest %MSE (0.16%) is in spectra with good spectral quality (SNR=126, LW=7.9 Hz).

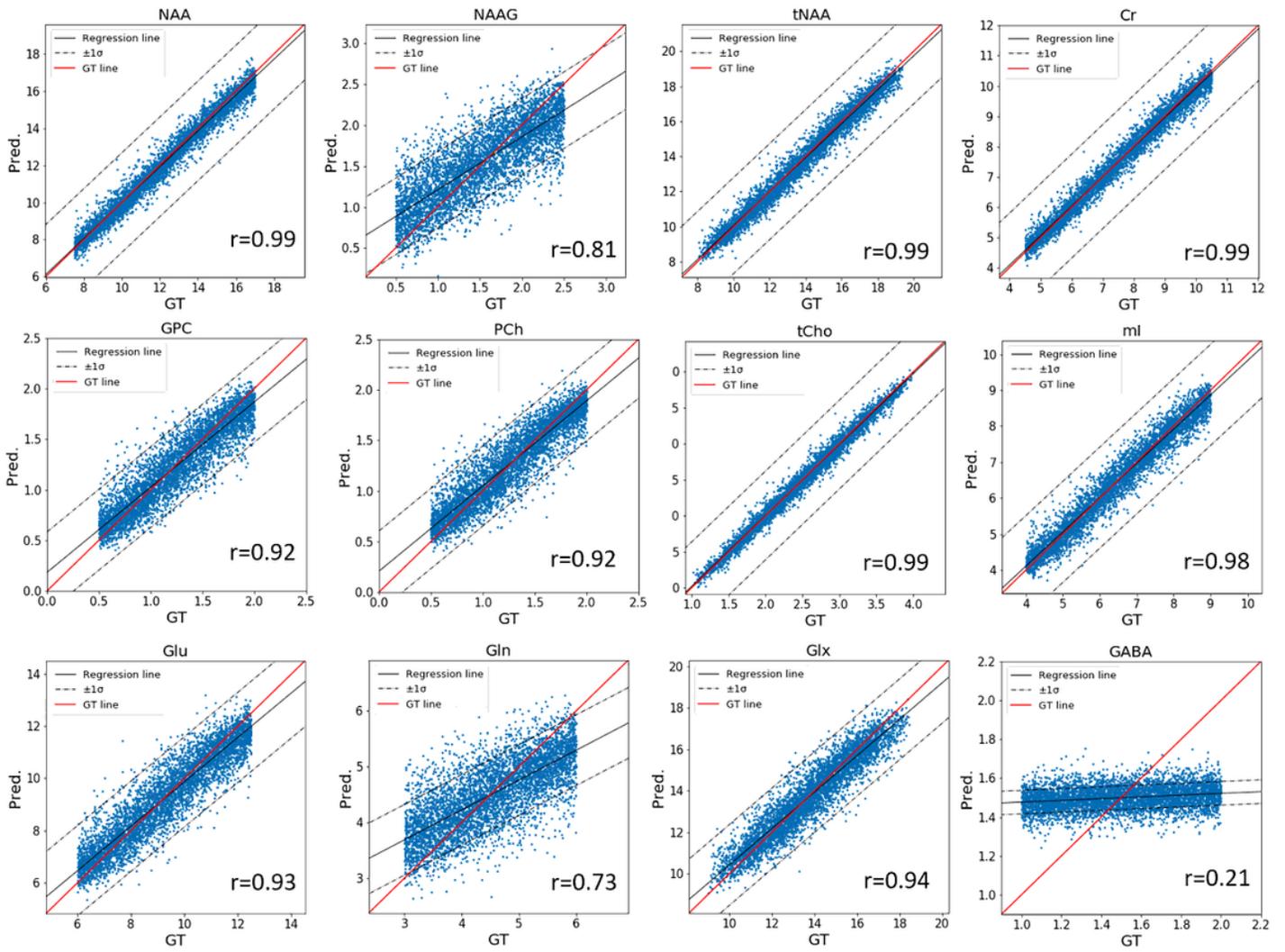


Figure 3

Scatter plots of the predicted metabolite concentrations (pred.) and GT for test data in CNN evaluation. For each plot, GT line (solid red) and regression line (solid black) with one standard deviation (dash black) were plotted. Pearson's correlation coefficient (r) were calculated for each metabolite. All metabolites have positive correlations. The lowest correlation is found in GABA ($r=0.21$). Other metabolites have correlations >0.9 except NAAG ($r=0.81$) and Gln ($r=0.73$).

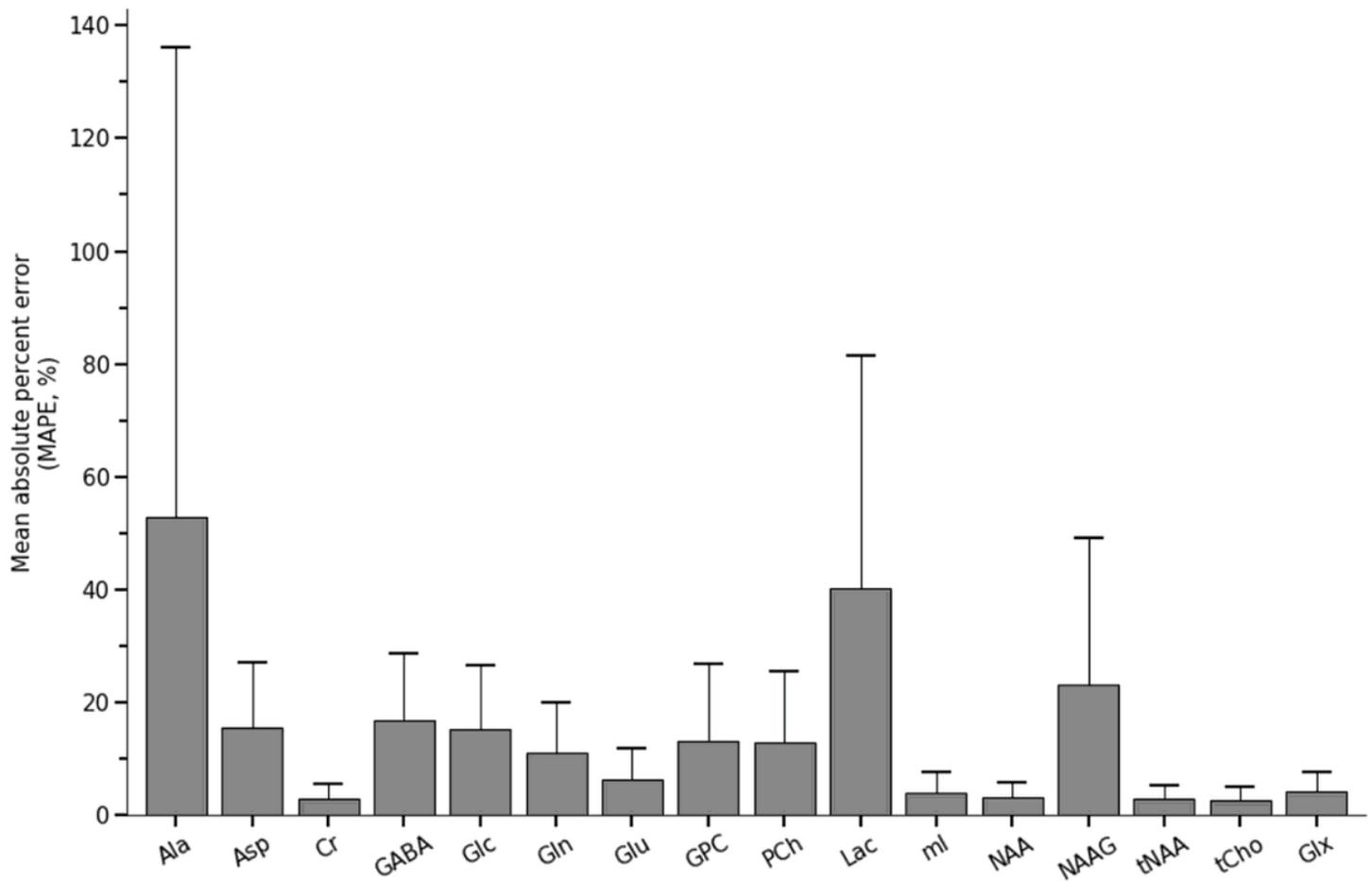


Figure 4

Mean absolute percent error (MAPE) of all metabolites in the basis. MAPEs were calculated by the errors between the predicted metabolic amounts and ground truth, where multiple linear regression was used to estimate the predicted metabolite amounts. The MAPEs and the error bars representing the standard deviation were estimated over the simulated brain spectra in the test set ($N = 5000$).

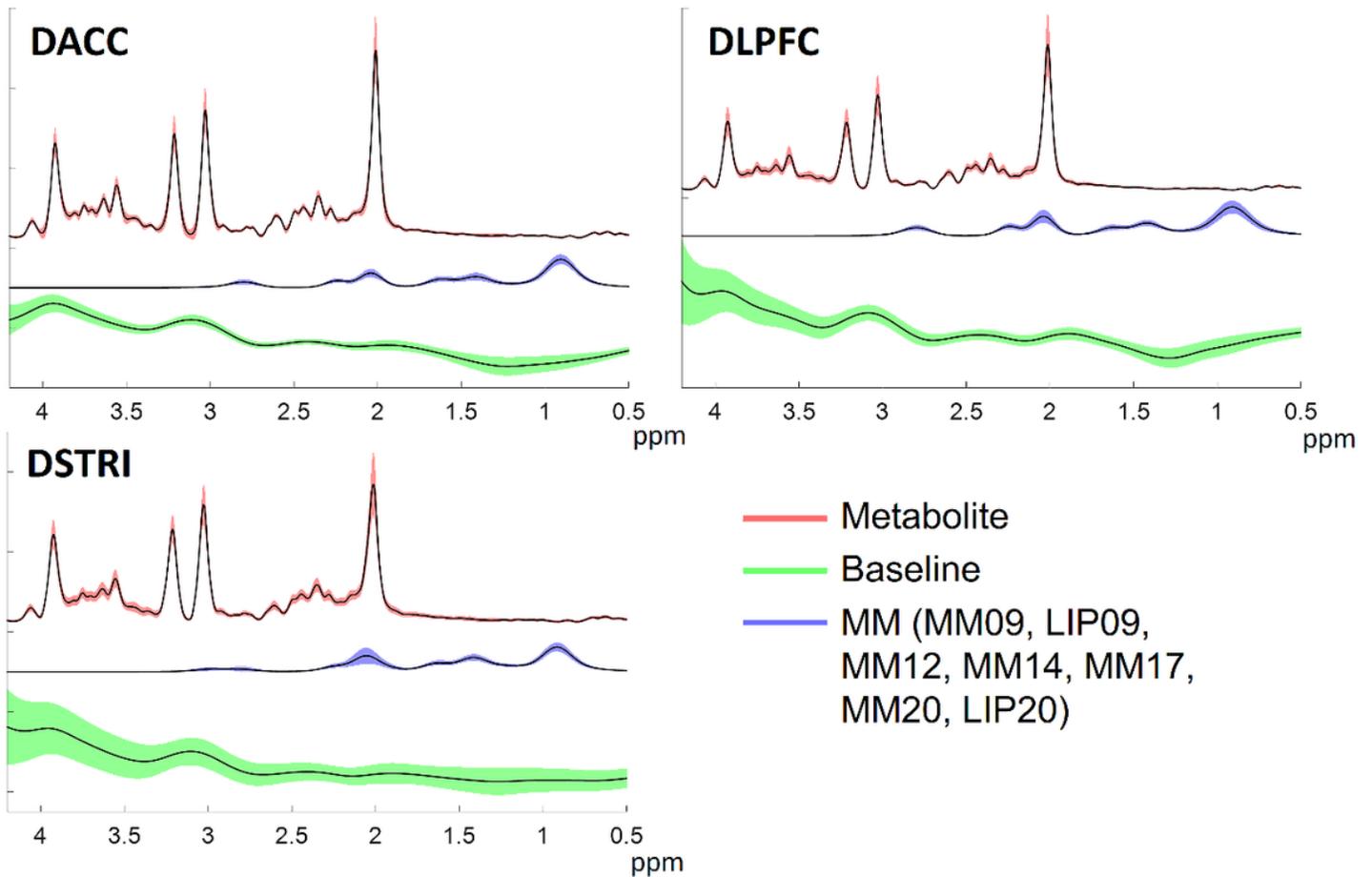


Figure 5

The LCMoDel fitting results for all subjects with mean (black line) and standard deviation (colored shaded area) for metabolites (red), baseline (green), and MMs (blue). The MM was the fitting lines of MM09, LIP09, MM12, MM14, MM17, MM20 and LIP20 from the LCMoDel, which is responsible for the contribution of MM09, MM12, MM14, MM16, MM20, MM21, MM23, MM26, MM30, MM31 in table S2. The fitted spline baselines were relatively distorted, with large deviations across volunteers in the range from 3.5 ppm to 4.0 ppm, which is responsible for the fitting of residue water and contribution of MM37, MM38 and MM40 in table S2.

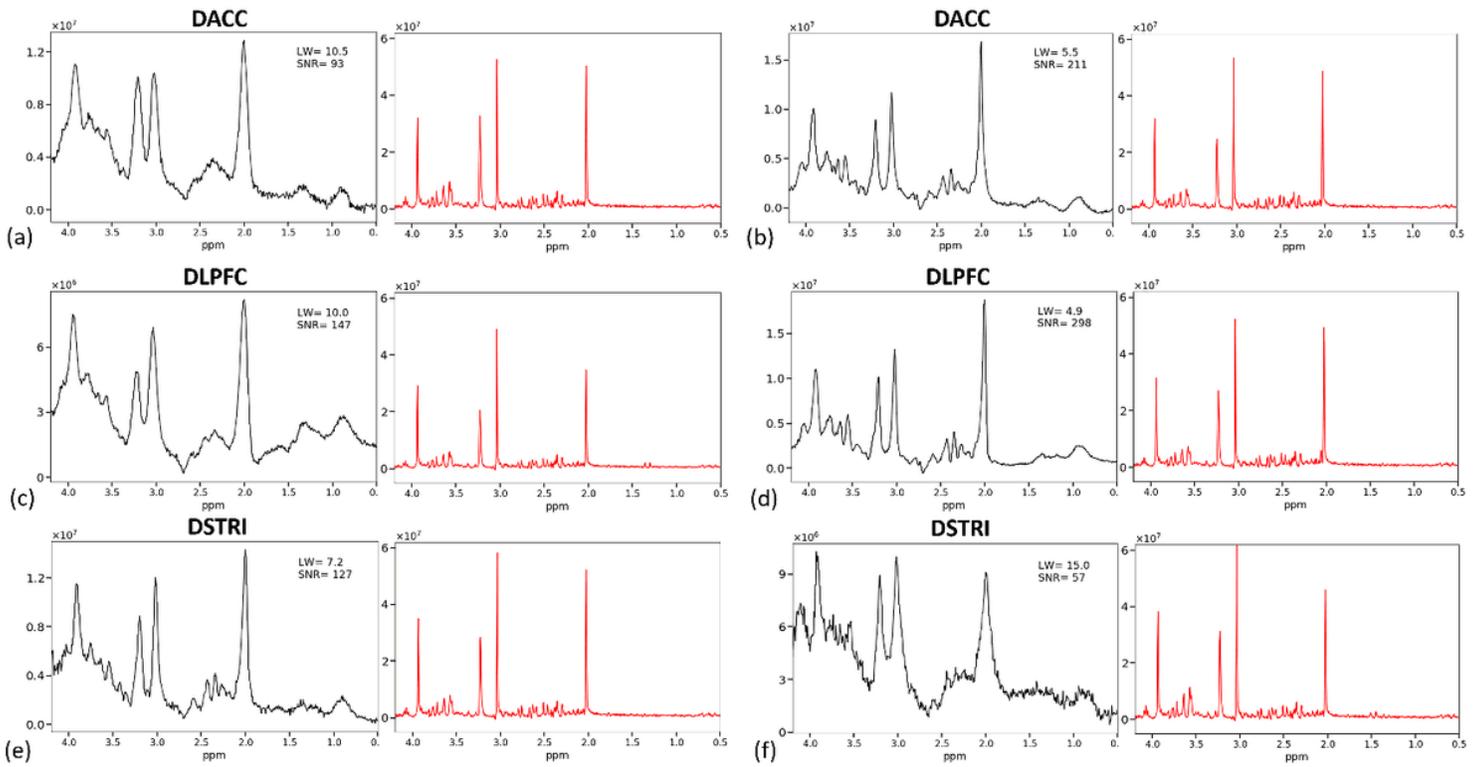


Figure 6

Representative *in vivo* spectra (black) and CNN predicted spectra (red) from (a) DACC, (b) DLPFC, (c) DSTRI. For each region, spectra under different spectral quality in term of SNR and LW were shown. The properties of noise reduction, LW narrowing, baseline removal can be clearly seen. The CNN predicted spectra have LW similar to the spectra in the basis.

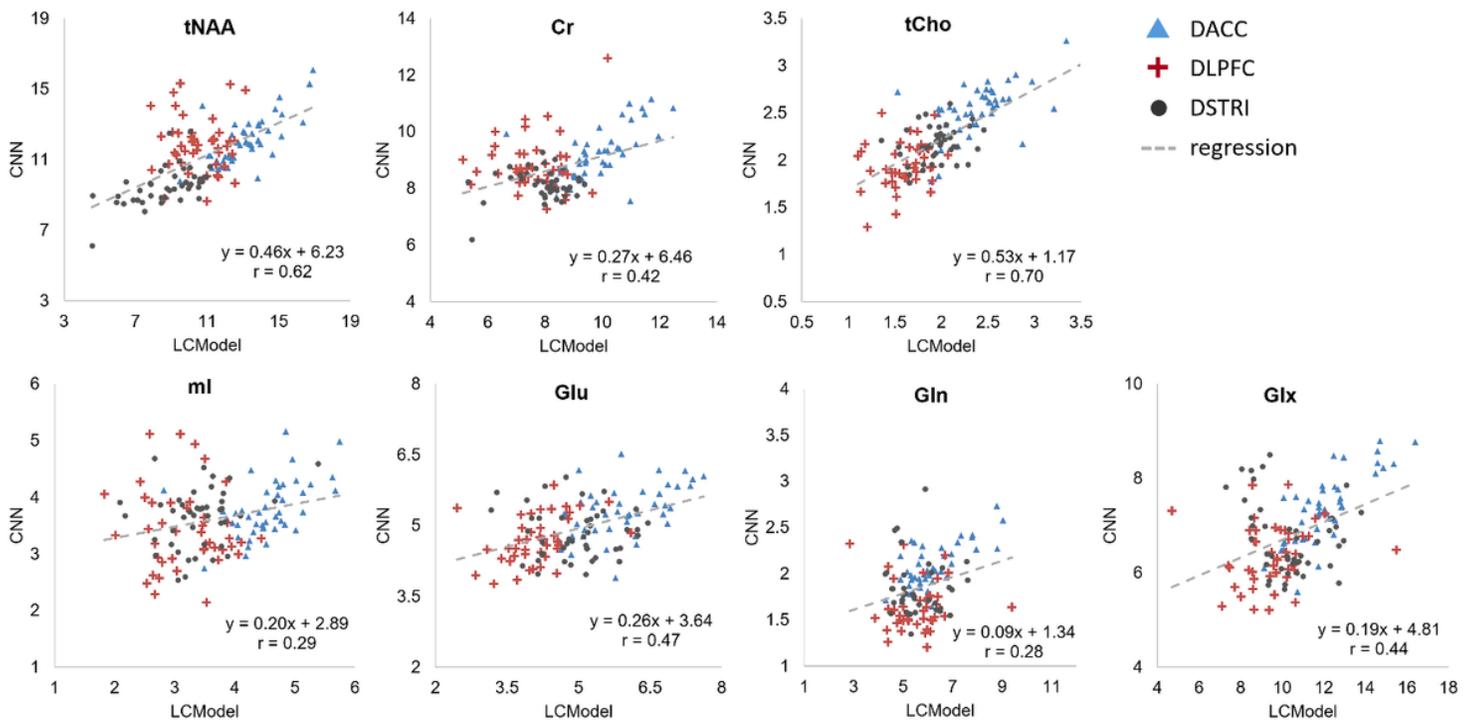


Figure 7

Scatter plots of the metabolite concentrations quantified by CNN and by LCMoel. For each plot, metabolite concentrations in IU were excluded by CRLB thresholds using the criterion: CRLB>10% for tNAA, Cr, tCho; CRLB > 20% for ml, Glu, Gln, Glx. For Gln, 4 subjects were excluded in DSTRI and 2 subjects were excluded in DLPFC. For Glu, 1 subject was excluded in DLPFC. For each plot, regression line (dash gray) was plotted and Pearson's correlation coefficient (r) was calculated. All metabolites have statistically significant correlation (p<0.005). Among 7 metabolites, lowest correlation is found in Gln (r=0.21). The concentrations of tNAA, Cr, tCho, ml, Glu from CNN and from LCMoel are distributed over similar range.

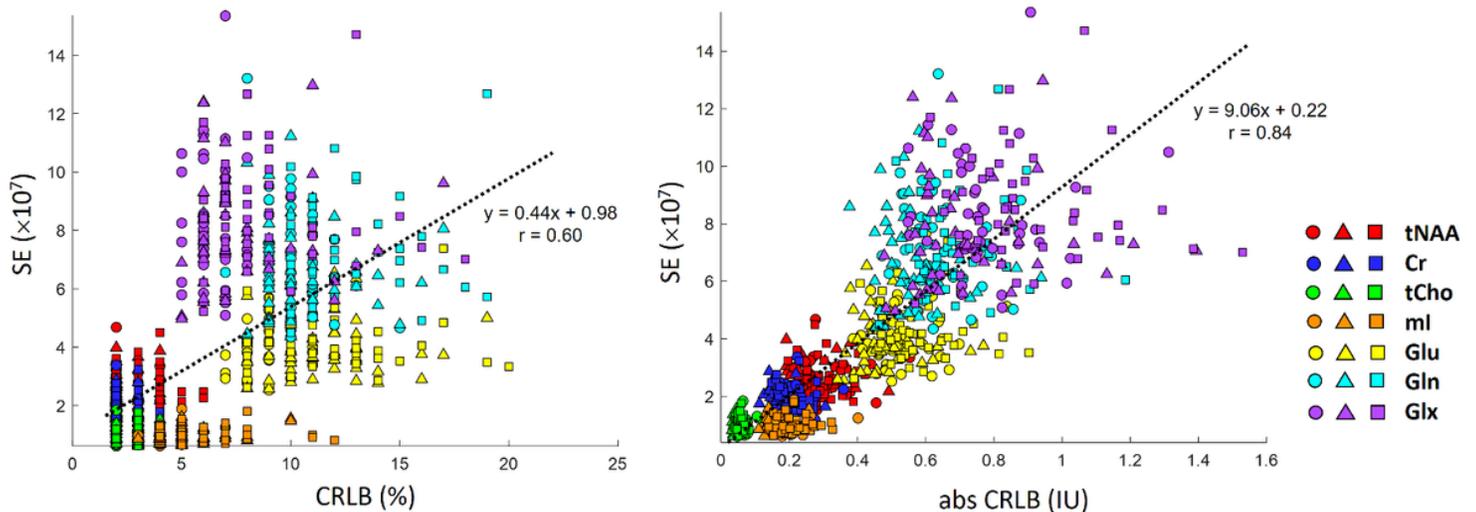


Figure 8

Scatter plot of standard error (SE) from CNN predicted spectra and Cramer-Rao lower bounds from LCMoel results for NAA, Cre, Cho, ml, Glu, Gln, Glx. Data from DACC (circle), DLPFC (triangle) and DSTRI (square) for all subjects were shown together. Both CRLB and abs CRLB show positive correlation with SE (p<0.01). Correlation coefficient between SE and abs CRLB is higher than correlation coefficient between SE and CRLB. Note, the SE in the plot was shown and analyzed with values multiplied by 10^7 .

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [DeeplearnMRSsupp.docx](#)