


ORIGINAL RESEARCH

Open Access



Prediction of multiple pH compartments by deep learning in magnetic resonance spectroscopy with hyperpolarized ^{13}C -labelled zymonic acid

Wai-Yan Ryana Fok^{1†}, Martin Grashei^{2†}, Jason G. Skinner², Bjoern H. Menze¹ and Franz Schilling^{2,3*} 

Abstract

Background: Hyperpolarization enhances the sensitivity of nuclear magnetic resonance experiments by between four and five orders of magnitude. Several hyperpolarized sensor molecules have been introduced that enable high sensitivity detection of metabolism and physiological parameters. However, hyperpolarized magnetic resonance spectroscopy imaging (MRSI) often suffers from poor signal-to-noise ratio and spectral analysis is complicated by peak overlap. Here, we study measurements of extracellular pH (pH_e) by hyperpolarized zymonic acid, where multiple pH_e compartments, such as those observed in healthy kidney or other heterogeneous tissue, result in a cluster of spectrally overlapping peaks, which is hard to resolve with conventional spectroscopy analysis routines.

Methods: We investigate whether deep learning methods can yield improved pH_e prediction in hyperpolarized zymonic acid spectra of multiple pH_e compartments compared to conventional line fitting. As hyperpolarized ^{13}C -MRSI data sets are often small, a convolutional neural network (CNN) and a multilayer perceptron (MLP) were trained with either a synthetic or a mixed (synthetic and augmented) data set of acquisitions from the kidneys of healthy mice.

Results: Comparing the networks' performances compartment-wise on a synthetic test data set and eight real kidney data shows superior performance of CNN compared to MLP and equal or superior performance compared to conventional line fitting. For correct prediction of real kidney pH_e values, training with a mixed data set containing only 0.5% real data shows a large improvement compared to training with synthetic data only. Using a manual segmentation approach, pH maps of kidney compartments can be improved by neural network predictions for voxels including three pH compartments.

Conclusion: The results of this study indicate that CNNs offer a reliable, accurate, fast and non-interactive method for analysis of hyperpolarized ^{13}C MRS and MRSI data, where low amounts of acquired data can be complemented to achieve suitable network training.

Keywords: Deep learning, Convolutional neural network, pH, Hyperpolarized ^{13}C MRSI

Introduction

In living species, extracellular pH (pH_e) is an important physiological parameter that is tightly regulated by intrinsic buffer systems. Locally, deviations from the systemic pH are often caused by pathologies, such as cancer, inflammation, infection, ischemia, renal failure

*Correspondence: schilling@tum.de

[†]Wai-Yan Ryana Fok and Martin Grashei contributed equally to this work

² Department of Nuclear Medicine, TUM School of Medicine, Klinikum Rechts der Isar, Technical University of Munich, 81675 Munich, Germany
Full list of author information is available at the end of the article

or pulmonary disease [1–3]. Since pH_e can play a critical role in disease progression [4] and can influence therapeutic success [5], many efforts have been undertaken to develop a quantitative non-invasive pH imaging technique [3, 4, 6]. However, there is no clinical routine method available for spatial quantification of pH_e , rendering it still an important target in biomedical imaging.

Magnetic resonance-based pH imaging methods offer high spatial resolution without limitations on the penetration depth and without involving ionizing radiation. In addition, conventional ^1H MRI offers high anatomical soft tissue contrast that can be overlaid on top of pH images. MRI-based pH_e imaging techniques that have been applied in vivo require the use of exogenous molecules and rely either on their pH-dependent chemical exchange saturation transfer (CEST) or on their pH dependence of chemical shifts [1]. Utilizing endogenous molecules, the intracellular pH (pH_i) can be measured by pH_i -dependent proton exchange from amide groups of intracellular proteins [4].

Magnetic resonance-based detection of biochemical and physicochemical quantities by exogenous molecules was revolutionized by dissolution dynamic nuclear polarization (DNP) which lifts nuclear spin polarization to a so-called hyperpolarized state leading to a sensitivity gain of more than four orders of magnitude [7]. Hyperpolarized [$1\text{-}^{13}\text{C}$]pyruvic acid is currently being used in clinical studies to examine its use for metabolic imaging of cancer, as well as in the brain and the heart [8–10]. Several pH-sensitive molecules have been hyperpolarized and been used for in vitro pH mapping including ^{13}C , ^{15}N , ^{31}P , ^{89}Y and ^{129}Xe spin-1/2 nuclei [11]. Only two of those have so far been applied for pH imaging in vivo: hyperpolarized ^{13}C -labelled bicarbonate [3, 12] and hyperpolarized [$1,5\text{-}^{13}\text{C}_2$]zymonic acid (ZA) [13] as well as its deuterated variant [$1,5\text{-}^{13}\text{C}_2,3,6,6,6\text{-D}_4$]zymonic acid (ZA_d) [14].

With hyperpolarized bicarbonate, pH_e is being determined by the signal intensity ratio of the CO_2 and HCO_3^- peaks, while the pH_e determination with ZA works via spectral analysis of the peak position, i.e. the chemical shifts. Chemical shift-based pH_e detection offers the unique advantage compared to intensity-based pH detection that multiple pH_e compartments within one imaging voxel can be resolved if their spectral peaks are separable, e.g. for resolving different pH_e compartments in the kidney [13]. For intensity-based pH_e detection, on the other hand, multiple pH compartments within one imaging voxel result in one signal intensity ratio, allowing only the determination of an average voxel pH. The concept of chemical shift-based detection of quantitative physiological measures

using hyperpolarized magnetic resonance sensors has, besides for detection of pH_e , also been used to quantify zinc [15], calcium/magnesium/iron ions [16], temperature [17], or ligand-receptor interactions [18].

Quantification of these measurements with hyperpolarized NMR sensors is done via analysis of the peak positions of the respective molecular sensors. Typically, the NMR spectra and all respective peaks are fitted via an optimization procedure giving the peak positions and amplitudes. However, such line-fitting procedures are error-prone in cases of low signal-to-noise ratio (SNR) and peak overlap, e.g. for multiple pH_e compartments within the kidneys [13]. In recent years, deep learning has shown its potential for magnetic resonance spectroscopy (MRS) and magnetic resonance spectroscopic imaging (MRSI) data in several applications to improve analysis of noisy data with interfering signals [19, 20]. Among these, artificial neural networks (ANN) demonstrated their value for spectroscopy analysis in medicine by classifying lung cancer tissue based on ^1H MRS [21] or denoising of brain ^1H MRS [22]. Furthermore, it was shown that convolutional neural networks (CNN) and multilayer perceptrons (MLP) can be trained to classify specific chemical compounds in various spectroscopy data sets [23, 24]. Nevertheless, we hypothesize that there is an advantage in applying a CNN for spectral analysis, as this class of network is invariant under frequency shifts of the entire spectrum which can be caused by B_0 inhomogeneities.

We also hypothesize that transfer learning with real mice kidney data could improve the performance for our deep learning model. Transfer learning and domain adaptation have been used to adapt the model trained by one data distribution to the target data domain [25, 26], especially when the target domain data is limited [27]. Our target domain data, ^{13}C -labelled zymonic acid kidney spectra, are by definition of the animal study and experimental efforts limited in size. Only one or two PRESS spectra or one CSI data set, still containing only a few single voxel spectra from kidneys, can be obtained from a single imaging experiment.

In this work, we investigate whether deep learning can improve the prediction of multiple pH_e compartments from magnetic resonance data using hyperpolarized ZA. For this task, we evaluate the performance of a CNN compared to a MLP as well as to conventional line fitting on both a single type of data (synthetic) and real data adaptation (a mix of real and synthetic data). For deep learning evaluation, both real data using line fitting as a gold-standard for evaluation of pH_e compartments as well as synthetic data with known pH_e compartments are used.

Methods

Neural network architecture

We implemented a multi-output regression convolutional neural network (CNN) and a multilayer perceptron (MLP), as shown in Fig. 1. The neural networks learned to map nuclear magnetic resonance spectra to specific pH_e values of a specific number of distinct pH_e compartments, of which were three for our specific case of in vivo kidney data of healthy mice.

The proposed CNN (Fig. 1a) consists of 4 sequential hidden layers for feature extraction. Each layer consists of a 1D convolutional kernel, a rectified linear unit (ReLU) activation function, and max pooling. The input into each layer was first convolved with the sliding kernel with stride of 1 and with length of 300, 150, 50 and 20. The initial weight (kernel) was a random value drawn from a truncated normal distribution. The convolutional kernel length decreased along the layers for extracting sub-regional features [28] as the input was downsampled from 1024 to 512, 256, 128 and 64 due to max pooling. The number of filters for feature extraction increased from 4, 4, to 8 and 8. ReLU activation functions were used in each layer to provide sparsity and thus robustness to small changes in input such as noise [29], as noise is inevitably present within the acquired spectra. The pooling layer was used to reduce the tensor size which could potentially merge semantically similar features [28].

Dropout was applied in the last feature extraction layer as a regularization to prevent over-fitting [30], for which it randomly dropped out 10% of the weights during training. Before the output layer, these feature maps were then flattened and passed to a dense layer with hyperbolic tangent activation function [31], and the outputs were then mapped into the target pH_e range 6.32 to 7.44.

The proposed MLP (Fig. 1b) consists of 4 fully connected dense layers, followed by max pooling. Each dense layer has a filter size 16, 16, 32 and 32, for which the output of each node is connected to all of the input nodes in the next layer. To compare their performance in mapping pH_e compartments, the architecture of CNN and MLP were designed such that the number of weights in both neural networks were similar, at approximately 8000.

Modelling of NMR pH_e spectra

Relative to the urea peak position, the chemical shifts ZA_5 and ZA_1 can be described as a function of pH_e by the following scaled logistic function [13]:

$$\text{ZA}_i(\text{pH}_e) = \text{ZA}_{i,\text{min}} + \frac{\delta_i}{1 + 10^{(\text{pK}_a - \text{pH}_e)}} \quad (1)$$

where $\text{ZA}_{5,\text{min}} = 12.57$ ppm, $\text{ZA}_{1,\text{min}} = 8.52$ ppm, $\delta_1 = 2.57$ ppm, $\delta_5 = 5.13$ ppm and $\text{pK}_a = 6.90$ [13]. Using a Lorentzian peak model, the spectrum can be described by the following equation:

$$x(f) = \sum_i^N \frac{a_i \cdot w_i^2}{w_i^2 + (x + \text{ZA}_i(\text{pH}_e))^2} \quad (2)$$

where $x(f)$ is the spectrum, N is the number of peaks (in our case $n=7$; 6 zymonic acid peaks and one urea peak), a_i is the peak amplitude, w_i is the full width at half maximum, and $\text{ZA}_i(\text{pH}_e)$ is the corresponding chemical shifts of ZA_5 and ZA_1 peaks found by Eq. 1. For the ^{13}C -urea peak, $\text{ZA}_i(\text{pH}_e)$ is set to zero.

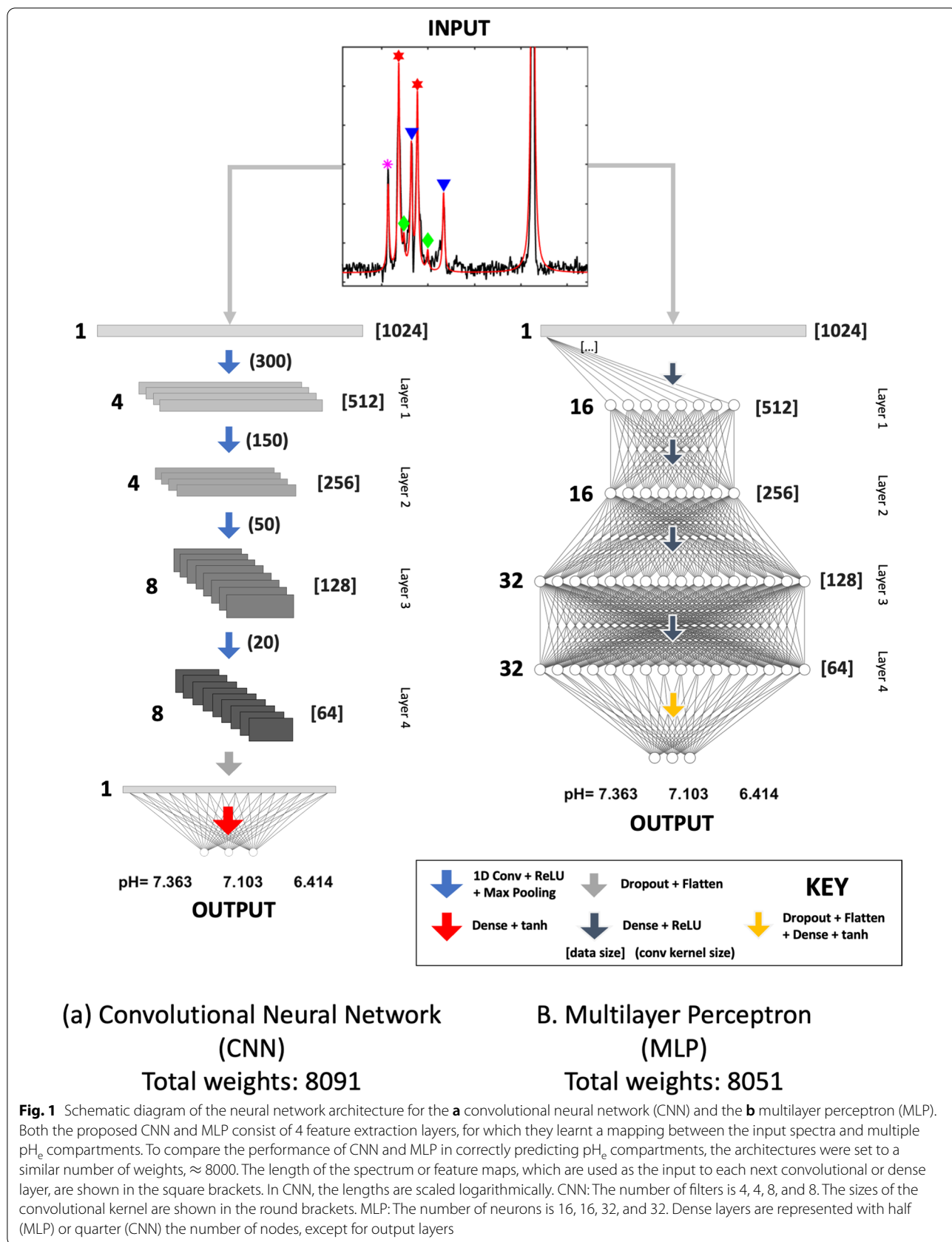
Hyperpolarized ^{13}C magnetic resonance spectroscopy

Hyperpolarization

27 mg [$1,5\text{-}^{13}\text{C}_2,3,6,6,6\text{-D}_4$]zymonic acid [14], 1.7 mg Ox063 trityl radical (GE Healthcare, Chicago, IL, USA) and 24 μL DMSO were vortexed for 35 min. The mixture was added to a DNP sample cup and frozen in liquid nitrogen. 25 μL of a sample, containing 10 M ^{13}C -urea in glycerol, 30 mM Ox063 and 1.5 mM DOTAREM (Guerbet, Villepinte, France) was subsequently added on top of the frozen layer and also frozen in liquid nitrogen. The sample cup was then polarized at 1.2 K for three hours by irradiation with a microwave frequency of 94.169 GHz and a power of 100 mW using a HyperSense[®] DNP Polarizer (Oxford Instruments, Abingdon, UK). Dissolution was performed with 2.99 ml D_2O containing 80 mM TRIS, 0.3 mM EDTA and 50 mM NaOH, resulting in solutions containing 50 mM hyperpolarized zymonic acid and 100 mM urea with a pH 6.7 ± 0.4 .

Hyperpolarized ^{13}C -magnetic resonance spectroscopy

All experiments were performed on a horizontal bore small animal 7 T magnet MRI scanner (Agilent/GE) MR901 with Bruker AVANCE III HD electronics and a 31 mm $^1\text{H}/^{13}\text{C}$ volume resonator (RAPID Biomedical, Rimpar, Germany). Experiments were performed on seven healthy C57BL/6 mice (Charles River, Wilmington, MA, USA) in accordance with pertinent laws and regulations and approved by an ethical review board (Regierung von Oberbayern, Munich, Germany, Approval Number ROB-55.2-2532.Vet_02-17-177). Prior to imaging, animals were anaesthetized with 1.5–2.5% Isoflurane (v/v) in oxygen as a carrier gas, a tail vein catheter was placed and animals were positioned together with a 4.4 M [$1\text{-}^{13}\text{C}$] lactate phantom for B_1 calibration inside the magnet. Breathing rate ($40 \pm 7 \text{ min}^{-1}$) and animal temperature ($37.0 \pm 0.6 \text{ }^\circ\text{C}$) were monitored with an ECG trigger unit (RAPID Biomedical) and an MR-compatible temperature monitoring system Model 1030 (SA Instruments Inc,



Stony Brook, New York, NY, USA) respectively. Kidneys were located using ^1H RARE with FOV $32 \times 32 \text{ mm}^2$, slice thickness 1 mm, matrix size 128×128 , repetition time 4000 ms, effective echo time 48 ms, RARE factor 12, 10 averages. Following manual B_1 calibration using ^{13}C -FID acquisitions with non-selective excitation by a 1 ms block pulse of varying RF power and fitting of the resulting signal vs. excitation power curves, the hyperpolarized solution was injected and ^{13}C acquisitions started 5 s after end of injection. Single voxel ^{13}C spectroscopy used PRESS on single kidneys with typical parameters: Total echo time 13.9 ms, total scan time 531 ms, excitation pulse flip angle 90° , refocusing pulse flip angle 180° , receive bandwidth 2000 Hz, 1024 points, voxel size $5 \times 5 \times 7 \text{ mm}$. Hyperpolarized ^{13}C -MRSI was performed using FIDCSI with typical parameters: total scan time 14 s, FOV $28 \times 24 \text{ mm}^2$, slice thickness 5 mm, matrix size 14×12 , repetition time 83.1 ms, flip angle 15° , receive bandwidth 3200 Hz, 256 spectral points. Overall, nine CSI data sets and six PRESS data sets were acquired and used for training and testing.

Data analysis and conventional line fitting

All data processing was performed in MatLab (The Mathworks Inc., Natick, MA, USA). For PRESS acquisitions, spectra were line-broadened by 5 Hz and phased manually. For CSI acquisitions, no line-broadening was applied, and magnitude spectra were averaged across both kidneys. C_1 - and C_5 -peaks of zymonic acid and of urea were identified by a standard automatic peak picking algorithm in MatLab for each pH_e compartment and selection was inspected manually. Peaks were fitted according to the model described in Eq. 2 where peak height, position relative to urea, and a uniform peak width for all compartments were fitted as free parameters using a built-in non-linear least squares algorithm. Second, the corresponding pH_e value was fitted from the relative zymonic acid peak distance to urea according to Eq. 1. Pairs of zymonic acid peaks were grouped into pH compartments and for each detected pH_e compartment, a mean pH_e was calculated which was weighted by the signal intensities from both the C_1 - and the C_5 -peaks. Common values for peak linewidths in Hz and ratios of signal amplitudes of the different kidney compartments were also extracted for generation of synthetic spectra for the training of the neural networks.

Evaluation and training of the neural networks

Data sets

Spectra synthesization Due both to the fact that in vivo experiments are necessarily small in sample size for ethical reasons, and that in vivo hyperpolarized ^{13}C experiments are labour intensive, synthetic data was generated

for the purpose of training the neural networks. To avoid over-fitting, noise was included in the spectral synthesization and was performed based on the following model:

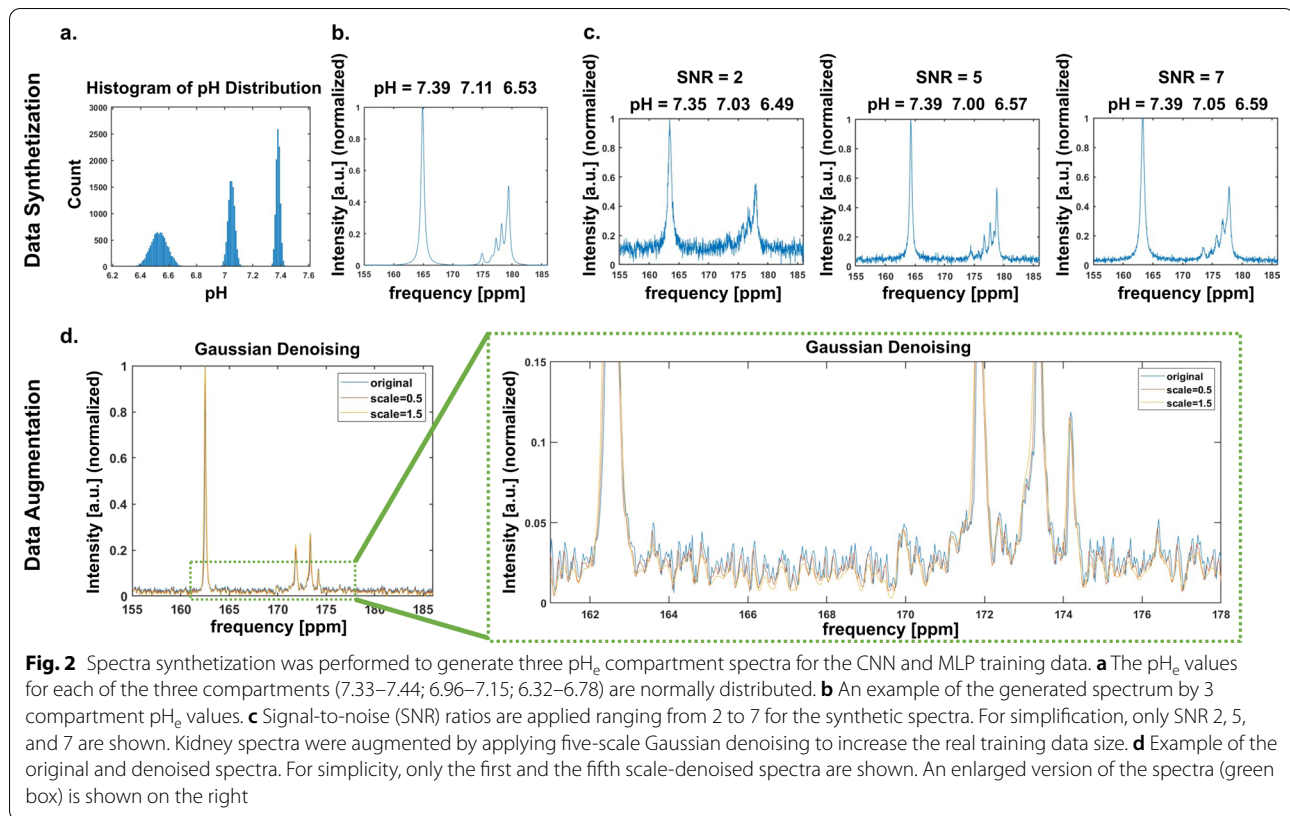
$$X(f) = b_0 + \varepsilon + S \cdot x(f) \quad (3)$$

where $X(f)$ is the synthetic spectrum, b_0 is a constant baseline, ε is the additive noise, S is the SNR scale factor, $x(f)$ is the set of Lorentzian peaks for the 3 metabolite peaks: urea, ZA_5 and ZA_1 . Figure 2a shows the distribution of the 3 pH_e compartment values of the synthetic spectra, which were initiated from a normal distribution respectively in the ranges: 7.33–7.44; 6.96–7.15; 6.32–6.78, which were found from the measurements carried out according to “Hyperpolarized ^{13}C -magnetic resonance spectroscopy” section. The urea peak positions and widths were initiated from a normal distribution with a standard deviation of 0.580 ppm starting from 163 ppm, to represent potential B_0 inhomogeneities and the peak widths were initiated between 30 and 70 Hz (0.397–0.927 ppm) to account for variations in shim quality. The corresponding chemical shifts ZA_5 and ZA_1 peaks were then found by Eq. 2 using the Lorentzian peak model, and a set of basis peaks for the 3 metabolites peaks urea, ZA_5 and ZA_1 was then generated (Fig. 2b). The ratios for the urea, ZA_5 and ZA_1 peak amplitudes were set to be 4:1:2, which represents the in vivo signal ratios of zymonic acid peaks detected in the different pH_e compartments compared to urea. Gaussian noise and baseline were added to the set of combined basis peaks to take the noise from MRS acquisition into account. The range of SNR scale factor was 2 to 7 and baseline was set from -0.2 to 0.2 , both drawn from the normal distribution (Fig. 2c). 10,020 spectra were synthesized, 10,000 spectra were used for training and 20 spectra were used for testing.

Kidney data and augmentation To adapt the neural networks to our target domain of mice kidney spectra, we generated 40 augmented spectra as training data based on the eight acquired mice kidney spectra by applying a five-scale Gaussian denoising (scale factors 1.5, 1.2, 1, 0.8, 0.5) (Fig. 2d). The original eight spectra were used for testing.

Experiments

We set up four experiments to evaluate the performance of CNN and MLP on mixed training data: CNN_{mix} , MLP_{mix} , and single type of training data: CNN_{syn} , and MLP_{syn} . A total of 10,000 spectra were used (Training: 8500, Validation: 1500). CNN_{mix} and MLP_{mix} consisted of 9960 synthetic spectra, 40 augmented kidney data from PRESS and ROI-averaged CSI acquisitions, whereas CNN_{syn} and MLP_{syn} consisted of 10,000 synthetic datasets. All augmented spectra were



set to size of 1024, the magnitude of all spectra was normalized between 0 and 1 for training and testing.

A total of 28 spectra were used for testing, which included 20 augmented spectra and eight real mice kidney spectra. Generation of augmented spectra and pre-processing of in vivo kidney spectra was implemented in MatLab.

Network training

Both the CNN and the MLP were trained with a batch size of 200 spectra and 400 epochs. Training progress was achieved by minimizing the sum of mean-square error loss of three pH_e compartments: $L = L_{\text{cortex}} + L_{\text{medulla}} + L_{\text{ureter}}$, where L_i is the L2 norm of the difference between the predicted pH_e compartment from the input spectra, and the ground truth pH_e compartment. The loss was then back-propagated for updating the weight kernel for each layer using NADAM (Nesterov-accelerated Adaptive Moment Estimation) optimizer [32]. Both networks were implemented in Keras using TensorFlow as the backend [33]. The training time for both neural networks was approximately five minutes, both training and testing were performed on a NVIDIA Tesla P100 GPU.

Synthesis of line fitted pH values and neural network predicted pH values into a combined pH map

Based on the network performance results from tests on synthetic and real kidney data, the best performing network is chosen for neural-network-assisted improvement of pH mapping in healthy mice kidneys. For this purpose, supervised line fitting was performed voxel-wise on seven CSI data sets from four mice for which the correct number of fitted compartments and fit quality was assessed. Spectra which were fitted with three pH compartments were extracted. For each image, a segmentation mask was created, indicating voxels either containing three pH compartments (“1”) or less than three compartments (“0”; corresponding to 0, 1 or 2 pH compartments), and fed voxel-wise into the best-performing neural network. pH maps for each CSI data set were then generated where the segmentation mask-positive area pH values were replaced by the predicted pH values from the network for the respective compartment, resulting in hybrid pH maps which are composed of pH values either based on line fitting or neural network predictions. The corresponding mean pH maps were calculated by averaging all compartmental pH values.

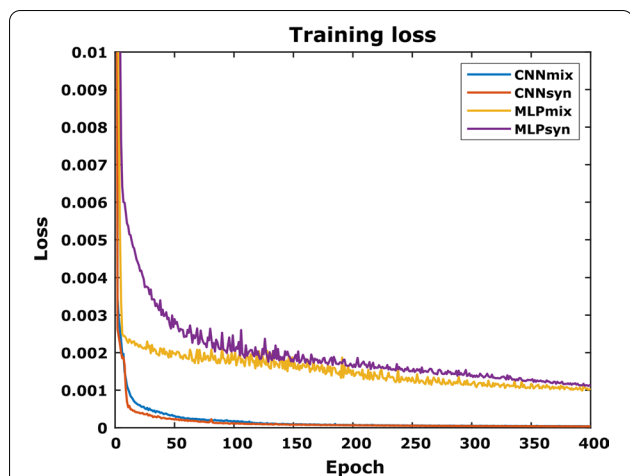


Fig. 3 The training losses for four models: CNN_{mix} , CNN_{syn} , MLP_{mix} and MLP_{syn} stably converge to a minimal loss. While the CNNs achieve absolute minimum loss after 400 epochs, the MLPs remain at a considerably higher loss having not yet reached their absolute minimum at this training stage

Results

^{13}C spectroscopic data

Spectra from ^{13}C -acquisitions of healthy kidney (top “Input” in Fig. 1) show the urea peak (164 ppm), six zymonic acid peaks (173–178 ppm) and the C_5 -peak of parapyruvate-hydrate (179 ppm), a decay product of zymonic acid. For zymonic acid, three pairs of C_1 and C_5 peaks can be grouped unambiguously to a single pH_e compartment, and each compartment corresponds

respectively to the three anatomical regions of the kidney, namely the cortex (red stars), the medulla (green diamonds) and the ureter (blue triangles). Multiple measurements on three mice return consistent pH_e values for the cortex ($pH_e = 7.38 \pm 0.03$, $n = 13$), the medulla ($pH_e = 7.06 \pm 0.06$, $n = 11$) and the ureter ($pH_e = 6.53 \pm 0.16$, $n = 9$).

Network training

The training losses of CNN_{mix} , CNN_{syn} , MLP_{mix} and MLP_{syn} over 400 epochs are shown in Fig. 3. While the CNNs rapidly converge to their respective limit, the MLPs’ minimal loss after 400 epochs remains higher compared to the CNNs, having not yet reached a converging limit. Interestingly, for both networks, this behaviour is independent of the training data set.

Comparison of performance of neural networks on augmented test data and conventional line fitting

To validate the network outputs and perform a comparison of the trained networks and conventional line fitting performed by an expert MRS scientist, 20 synthetically generated spectra with known ground truth pH_e compartments were blind-fitted by the conventional line fitting and pH analysis routine. In cases of sufficient pH difference between compartments together with sufficient SNR, all peaks can be detected reliably with a high accuracy of the predicted pH_e values (Fig. 4a).

In cases of low SNR and low intensity for one or multiple compartments relative to a third one, the conventional line fitting only partially allows for detection of

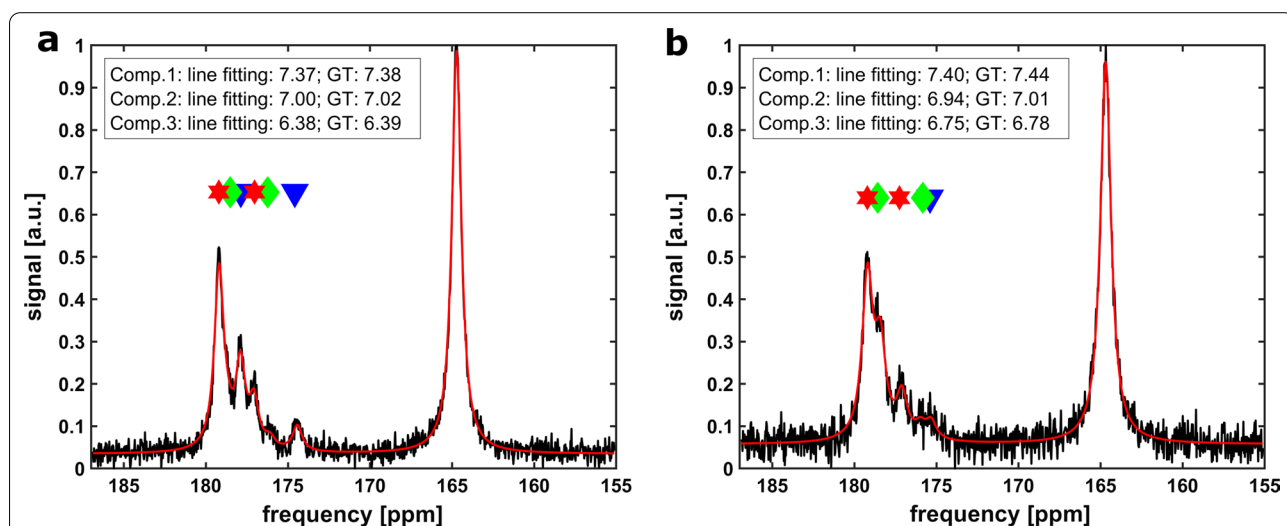


Fig. 4 **a** Conventional line fitting of a synthetic kidney spectrum with three fitted pH_e compartments indicated by coloured markers (compartment 1: red stars, compartment 2: green diamonds, compartment 3: blue triangles) and comparison of calculated pH_e values and ground truth data is shown as inset resulting in good agreement. **b** Conventional line fitting on a noisy spectrum and low compartment intensities of the medulla and the ureter. Three compartments can only be partially detected and agreement with ground truth is rather poor

pH_e compartments with less accuracy when comparing to the ground truth pH values (Fig. 4b). In the next step, the same twenty spectra were analyzed with the two CNN models CNN_{mix} and CNN_{syn} and the two MLP models MLP_{mix} and MLP_{syn} which were all previously trained either with a mixed (augmented + synthetic spectra) or with synthetic spectra only. The performance of the conventional line fitting and the four network models on the synthetic spectra test set relative to each other was compared by linear regression of the predicted and the ground truth pH value for each compartment. The adjusted coefficient of determination R^2 and slopes of the linear regression β to compare statistical and systematic uncertainties respectively are listed in Table 1. The dominant pH_e compartment of the cortex is best predicted by the CNN_{mix} ($\beta=1.01$), however with greater uncertainty ($R^2=0.78$) compared to conventional line fitting ($R^2=0.85$) or CNN_{syn} ($R^2=0.90$). The pH_e compartment of the medulla is best and equally well predicted by CNN_{mix} and CNN_{syn} compared to all other methods, however with greater systematic uncertainty ($\beta_{\text{CNNmix}}=1.30$, $\beta_{\text{CNNsyn}}=1.26$) and lower scattering ($R_{\text{CNNmix}}^2=0.91$, $R_{\text{CNNsyn}}^2=0.92$) compared to the cortex. Interestingly, for the pH_e compartment of the ureter, line fitting achieves equivalent statistical uncertainty ($R^2=0.99$) compared to the CNNs ($R_{\text{CNNmix}}^2=0.98$, $R_{\text{CNNsyn}}^2=0.99$) while outperforming them regarding systematic error ($\beta_{\text{Fit}}=1.02$). The performance of the MLPs is very poor for the cortex and the medulla compartment and only modest for the ureter compartment. This observation of poor MLP performance is also visualized in the modified Bland–Altman plots (Fig. 5), indicating that the MLPs have deviations of more than 0.2 pH units for the ureter compartments for some spectra. Furthermore, the good performance of both trained CNNs for all compartments is verified while for the conventional line fitting

some systematic underestimation of the cortex pH_e can be observed.

Comparison of performance of neural networks on real kidney data

To evaluate the suitability for routine spectroscopic data analysis, all four networks were tested on eight real mice kidney spectra. A comparison to the values obtained by conventional line fitting as a pseudo ground truth is visualized with Bland–Altman plots (Fig. 6). Following the trend observed for the synthetic test data, the CNN_{mix} network outperforms all other networks, showing the predicted pH_e values to deviate less than 0.1 pH units from the fitted data for all compartments. In contrast to testing on synthetic spectra, a CNN network being trained only with synthetic data CNN_{syn}, shows only poor performance when tested on real data with predicted pH_e values deviating up to 0.3 pH units from conventional fit values and decreasing performance from compartment 1 (cortex) to compartment 3 (ureter). For compartment 2 (medulla) and 3 (ureter), MLP_{mix} also achieves better agreement with fitted pH_e values. Analogous to testing on synthetic test spectra, MLP_{syn} shows the worst agreement with conventional line fitting, exceeding 0.1 pH units mean difference for compartment 3 (ureter).

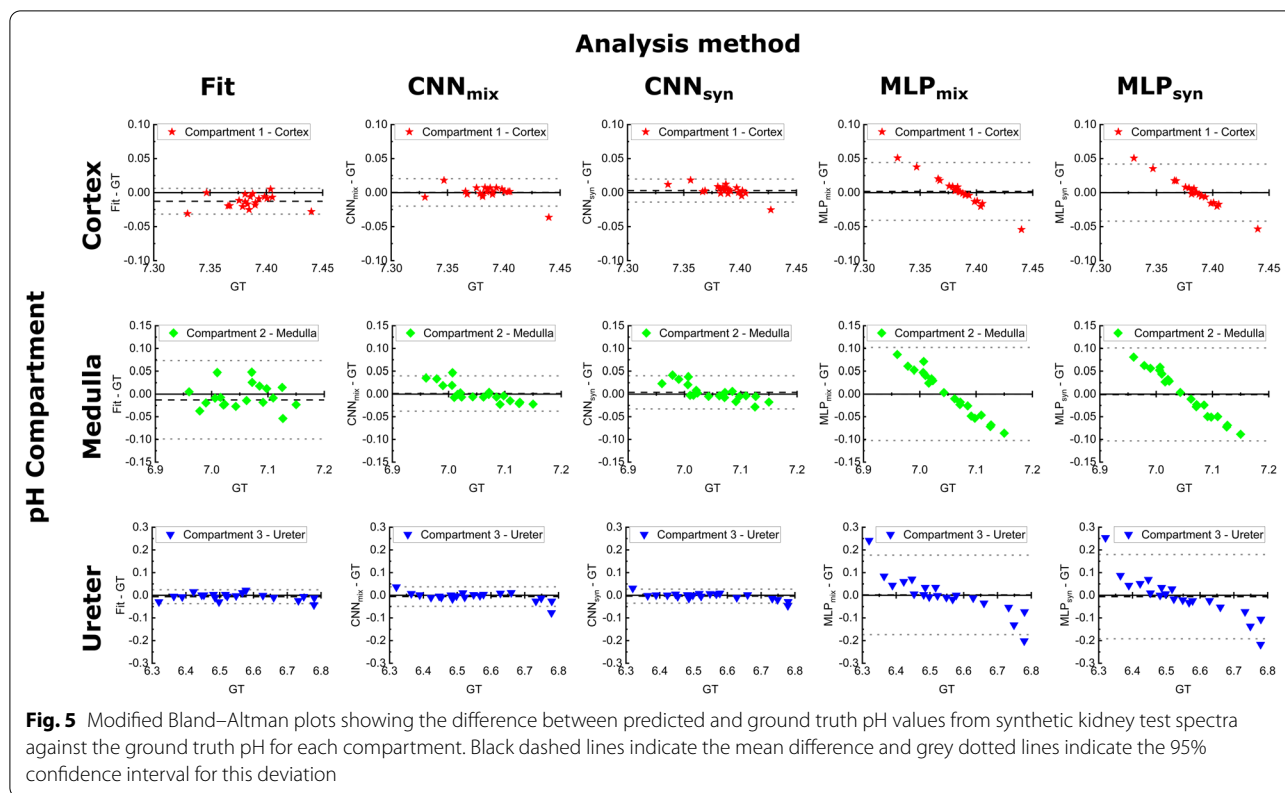
Hybrid pH mapping by voxel-wise combination of pH values reconstructed from CNN_{mix} and line fitting

Based on the performance measurements for synthetic and real kidney spectra, CNN_{mix} was chosen for application in pH mapping of healthy mice kidney (Fig. 7a). An exemplary segmentation mask for CSI data matching the anatomy in Fig. 7a based on supervised line fitting to distinguish voxels with three pH compartments and voxels with less than three pH compartments is shown in Fig. 7b

Table 1 Top: Evaluation of the prediction accuracy by compartment-wise comparison of the adjusted coefficient of determination R^2 derived from a linear regression of ground truth and pH values predicted by conventional line fitting (model “Modelling of NMR pH_e spectra” and “Data analysis and conventional line fitting” sections) and the neural networks after application to 20 synthetic test spectra; Bottom: Linear slope coefficients β derived from linear regressions to evaluate prediction bias

| R^2 | Line fitting | CNN _{mix} | CNN _{syn} | MLP _{mix} | MLP _{syn} |
|-----------------------|--------------|--------------------|--------------------|--------------------|--------------------|
| Compartment 1—Cortex | 0.85 | 0.78 | 0.90 | 0.11 | 0.22 |
| Compartment 2—Medulla | 0.65 | 0.91 | 0.92 | 0.05 | 0.08 |
| Compartment 3—Ureter | 0.99 | 0.98 | 0.99 | 0.65 | 0.59 |
| β | Line Fitting | CNN _{mix} | CNN _{syn} | MLP _{mix} | MLP _{syn} |
| Compartment 1—Cortex | 0.81 | 1.01 | 1.30 | 3.83 | 4.79 |
| Compartment 2—Medulla | 0.59 | 1.30 | 1.26 | 1.87 | 2.64 |
| Compartment 3—Ureter | 1.02 | 1.09 | 1.08 | 1.52 | 1.62 |

Both parameters show poor accuracy and strong prediction bias for the medulla for line fitting and MLP networks potentially due to low SNR



(white: three pH compartments, black: less than three pH compartments).

pH mapping based on supervised line fitting (Fig. 7c, top row) reveals a globally present physiological pH compartment (top left), a heterogeneous, slightly acidic second pH compartment which can be attributed to the medulla (top middle-left) and a third pH compartment corresponding to the ureter (top middle-right). Voxel-wise compartment-averaging generates kidney-specific pH contrast. Substitution of line fitted pH values by values predicted by CNN_{mix} shows slightly more basic pH values predicted by the neural network compared to the line fitting. For the medulla compartment, network predictions appear to be more homogeneous compared to line fitted maps. For the ureter compartment, line fitted as well as neural network predicted maps agree well with each other. The mean pH map based on these hybrid compartment pH maps shows good inter- and intra-kidney homogeneity in mean pH values compared to line fitted maps.

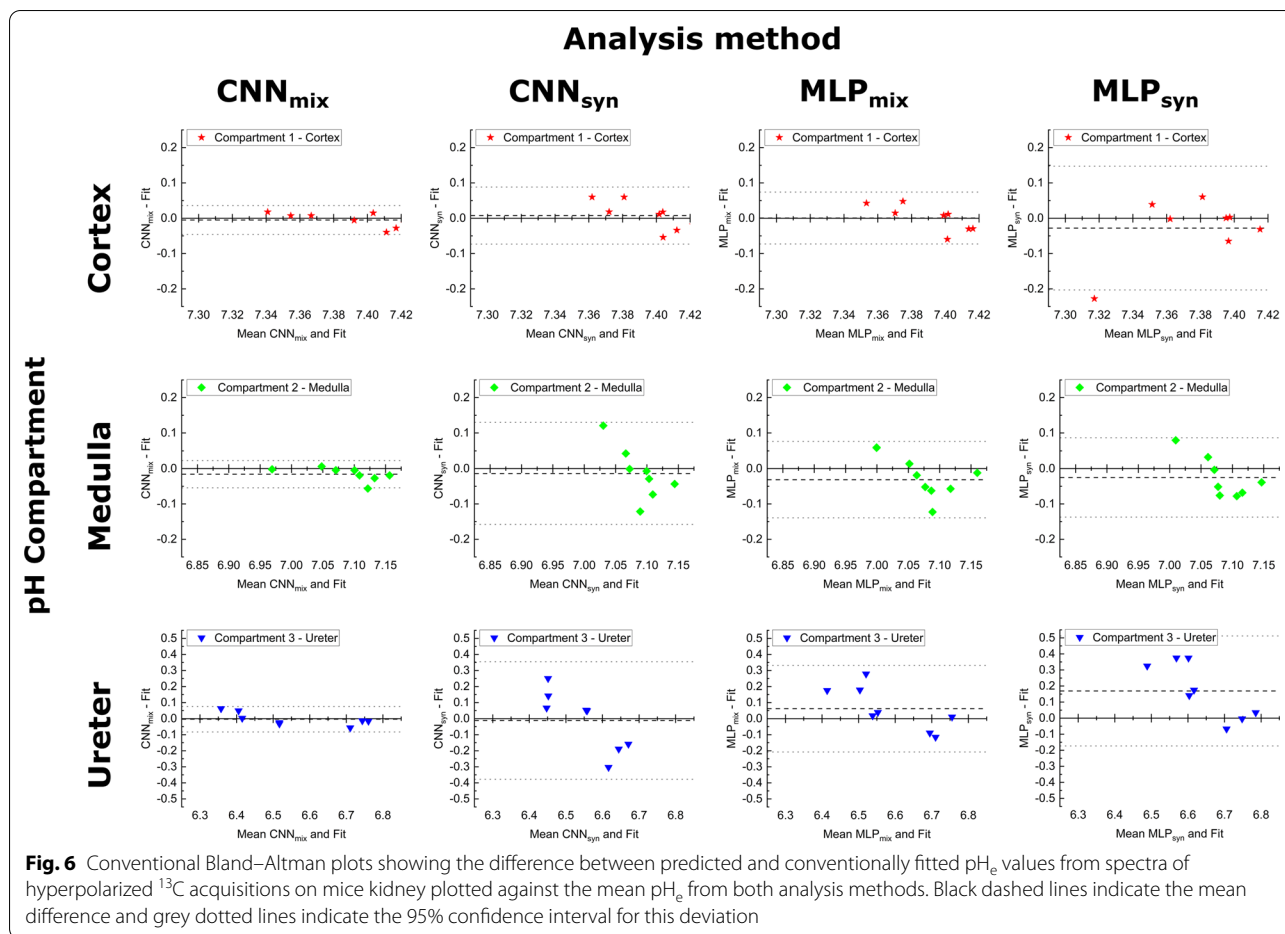
For quantitative comparison, pH compartments derived from line fitting and neural networks were both averaged across individual kidneys for multiple acquisitions on different animals (Fig. 8). pH values derived from line fitting show lowest inter- and intra-subject variation for the cortex ($pH_{cortex,fit} = 7.41 \pm 0.02, n = 14$) while pH

values for the medulla ($pH_{medulla,fit} = 7.09 \pm 0.10, n = 14$) and the ureter ($pH_{ureter,fit} = 6.70 \pm 0.13, n = 14$) are distributed across larger pH ranges while all compartments can be well separated from each other based on pH. pH compartments predicted by the CNN_{mix} agree well with line-fitted compartments, despite the cortex ($pH_{cortex,CNN} = 7.43 \pm 0.01, n = 14$) and the medulla ($pH_{medulla,CNN} = 7.13 \pm 0.04, n = 14$) exhibiting overall slightly more basic pH values compared to the line-fitted ones. For the ureter, no relevant difference can be observed ($pH_{ureter,CNN} = 6.72 \pm 0.04, n = 14$). In agreement with lower intra-subject variations as seen in compartment maps in Fig. 7, the inter- and intra-subject variations of compartment pH values are lower for the neural-network-predicted pH values, while the values for each subject are in good agreement relative to the compartment-specific standard deviation (black crosses are corresponding to the same kidney in the same subject in Fig. 8).

Discussion

Modelling and fitting of hyperpolarized ^{13}C spectroscopic data

Analogous to published data on spectra of zymonic acid in kidney of healthy rats [13], several pH compartments can be detected in hyperpolarized ^{13}C acquisitions of

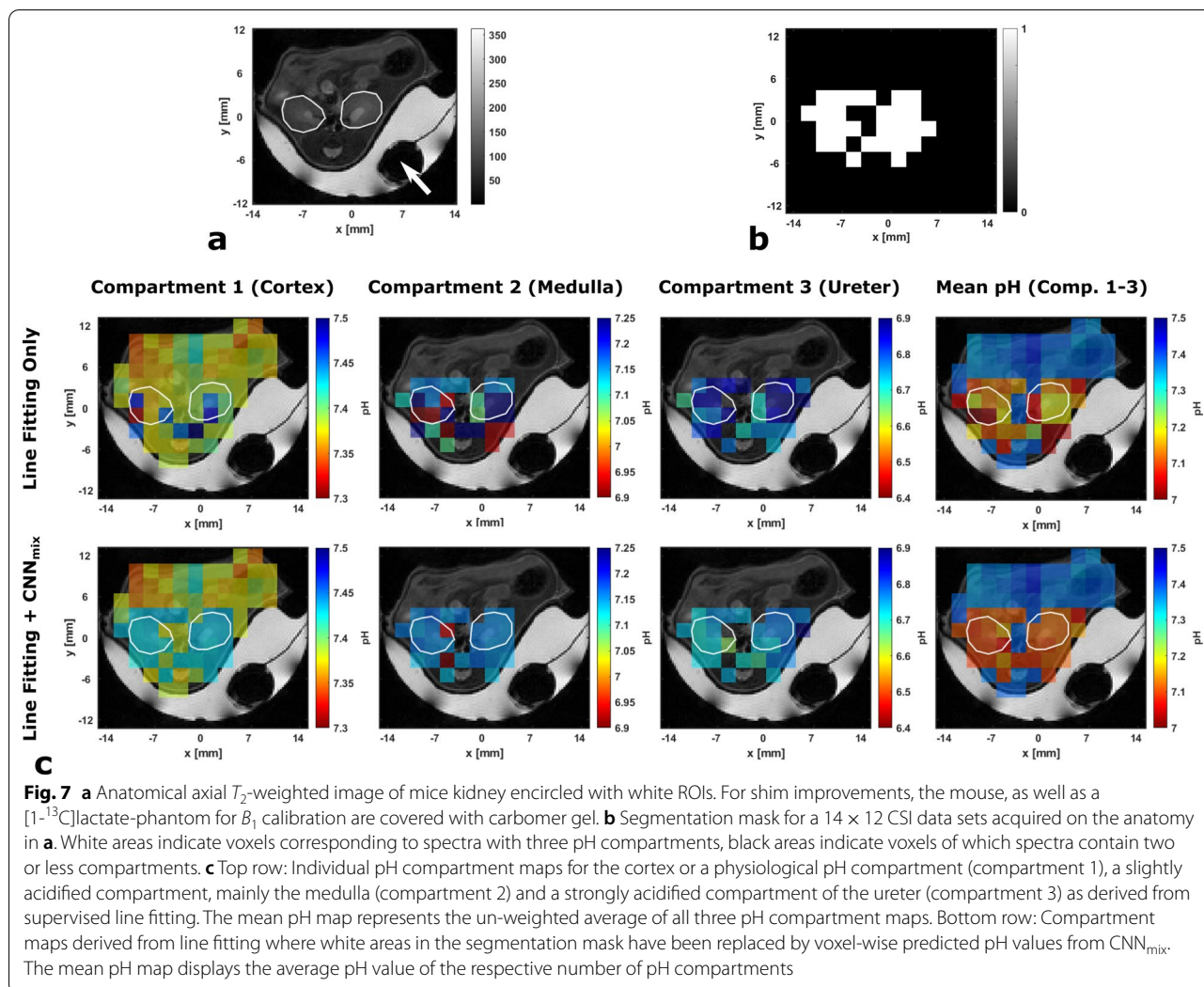


zymonic acid in healthy kidney of mice. However, out of three detected pH_e compartments which could be assigned to cortex, medulla and ureter, the latter two often suffer from lower compartment signal due to limited perfusion of the agent within the short acquisition time [13] or enhanced signal decay in case of injury or pathological alterations [34–36]. Consequently, line fitting with conventional methods becomes challenging when trying to resolve the pH_e compartments of the kidney. For these reasons, manual spectral pre-processing together with fitting of a linear combination of individual Lorentzian peaks while minimizing the free parameters to amplitude, peak position, and peak width (FWHM) was chosen as an appropriate fitting routine despite inherent SNR limitations. For these reasons, manual spectral pre-processing together with fitting of a linear combination of individual Lorentzian peaks while minimizing the free parameters to amplitude, peak position, and peak width (FWHM) was chosen as an appropriate fitting routine despite inherent SNR limitations. Additionally, other standard fitting routines for MR spectra such as LCModel [37] or AMARES [38], were either

unsuitable, unstable, or of no significant benefit in this work.

For instance, for AMARES it was already shown that this algorithm suffers from unstable fitting when the peak frequencies are allowed as free parameters [39]. In addition, zymonic acid spectra on kidney are less sparse than the previously fitted pyruvate- and lactate-containing spectra.

LCModel is a standard fitting routine for magnetic resonance spectra which allows excellent peak quantification for 1H metabolites. Nevertheless, this method also has several limitations regarding the application to the data of this work. LCModel predominantly aims to quantify spectral peaks which requires the input of a set of basis spectra of high spectral quality and good SNR. In such cases, the peak positions are fixed, and only minor peak shifts due to eddy currents and magnetic field inhomogeneities are tolerated. This contrasts with hyperpolarized ^{13}C acquisitions using zymonic acid in which SNR is typically modest, spectral resolution limited and quantification not necessary. As zymonic acid peaks strongly shift with pH_e , a suitable set of basis spectra would

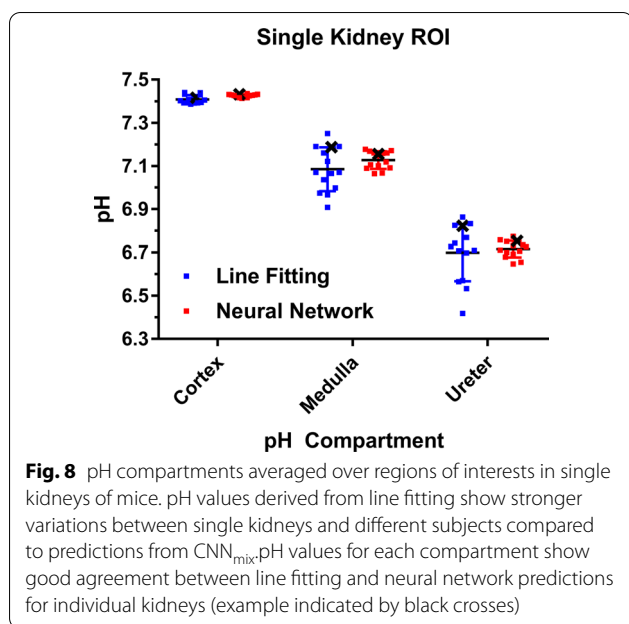


require multiple zymonic acid spectra at different pH_e values which either requires a high amount of basis components or an inherent limitation in the measured pH accuracy limited by a small basis set. In addition, since C_1 - and C_5 -peak intensities vary relative to each other in different acquisitions, combined modelling as one basis spectrum for a fixed pH is difficult. Furthermore, LCMoDel requires well separated peaks for proper differentiation which is not the case for the densely packed pH compartment peaks as seen in Fig. 1.

CNN and MLP performance

In our study, both CNN_{mix} and CNN_{syn} outperform MLP_{mix} and MLP_{syn} in predicting all three pH compartments in synthetic test data. Here, the CNNs have better accuracy and less uncertainty as shown in the regression analysis. Interestingly, for the cortex and medulla compartments, CNN_{mix} was giving a better

prediction accuracy as compared to the conventional line fitting method. While the CNN and the MLP have a similar number of weights (≈ 8000), the CNN used kernels in convolutional layers to perform elementwise multiplications to inputs while the MLP used densely connected neurons. When applied systematically across the entire input spectra, these convolutional kernels could extract spectral features such as the metabolite peaks distances, as the kernels account for the values on neighboring pixels. Because of the weight sharing that occurs when the convolutional kernels slide across the spectrum [40, 41], the CNN becomes less susceptible than the MLP to spectral variance or drifts in spectral peak positions which can be caused by B_0 inhomogeneities. However, the choice of specific neural network depends on the type of learning tasks and features to be extracted, as previous studies showed MLP performed well in classification [42, 43], while CNN



also demonstrated good performance in image segmentation [44] or classifications [45]. Some even explore the synergies of MLP and CNN networks [46, 47].

Performance of neural networks trained with mixed data

We showed that it is possible to train models on a limited amount of real data by transfer learning, whilst most of the training data were synthetically generated based on a spectral model for $[1,5-^{13}C_2]$ zymonic acid and ^{13}C -urea. When tested with synthetic data, CNN_{syn} performed better than CNN_{mix} in predicting medulla and ureter pH as shown in the linear regression analysis and modified Bland–Altman plots. Especially for the medulla compartment, both line fitting and MLP_{syn} and MLP_{mix} show poor performance what might be due to higher sensitivity to low SNR. As there is no absolute ground truth for the kidney in healthy mice, we compared the neural network predictions with the results in line fitting as a pseudo ground truth. We found CNN_{mix} had the most consistent and comparable results to line fitting, as it has the smallest difference compared to other models, and its 95% confidence level is also smaller than CNN_{syn} in cortex and ureter (see Fig. 5). Moreover, MLP_{mix} are more comparable to line fitting than CNN_{syn} , and MLP_{mix} has generally smaller mean difference.

Hybrid pH map generation by combination of line fitting and CNN_{mix} predictions

Combination of line fitting and neural networks appears to improve pH mapping in kidneys of healthy mice. Based on compartmental pH maps, line fitting appears

to be only quantitatively robust for the cortex while the medulla and the ureter show considerable inter- and intra-kidney pH variability, the latter being physiologically rather unreasonable. We assume that this high variability stems from the low SNR of zymonic acid peaks corresponding to these two compartments. Substitution of voxels in pH maps corresponding to spectra containing three pH compartments by neural network predictions results in more homogeneous compartment maps while quantitatively still agreeing with line-fitted compartments. This suggests a superior performance of the neural network compared to the line fitting approach for low SNR compartments. Furthermore, as the network is predicting pH compartments voxel-wise based on individual spectra, it has to be pointed out that the predictions of neighboring voxels are independent from each other and therefore the observed spatial homogeneity of compartments therefore indicates a good robustness of the pH predictions. In addition, high quantitative prediction accuracy is suggested by the observation that cortex compartments are systematically predicted with a higher pH value compared to the line fitting method, which agrees with the observation that line fitting systematically underestimates the cortex pH when evaluating the method performance for artificial spectra of known pH.

Data size

These observations suggest that the real and augmented data might consist of spatially independent features, such as the noise during the spectra acquisition, which is crucial to train a more accurate model. Neural networks usually require a large amount of training data, and the number of training data depends on the complexity of the tasks and features to extract. However, generating a larger data set is challenging for hyperpolarized ^{13}C MRSI. In vivo spectra obtained by preclinical studies are limited in size for ethical reasons: the number of animals should be as low as possible. Additionally, the experimental efforts are rather large. Also, with regard to application for hyperpolarized ^{13}C acquisitions in humans, data set size is critical as clinical trials currently performed with this imaging technique are typically limited to 5–100 patients [48]. Efforts to obtain larger amounts of data might involve the generation of databases but, especially for imaging using hyperpolarized ^{13}C -labelled zymonic acid, this is at an early stage. Nevertheless, in our study, we showed improved network performance by including less than 0.5% real augmented data (40 augmented spectra out of 10,000 training spectra), an amount that can be realistically generated from single preclinical studies.

Future works

In this study, we found that the convolutional layers enable the network to better extract spectral features in the spectra. Future works could extend the application of convolutional layers to denoise the spectra or automating peak picking. Moreover, the neural networks here only predict the chemical shift of the spectra—they do not yet consider the signal intensity as in the conventional line fitting method. An extension the current approach might also predict signal intensity, which could then allow a more direct calculation of weighted-average pH maps. In addition, for imaging of cancer or unknown tissue, networks could be trained to predict the correct number of pH compartments and using this information to selectively pass the spectra through other networks which predict the correct pH values.

Conclusion

Two different types of neural networks trained once with a fully synthetic data set and once with a mixed data set, containing real and synthetic data, were each evaluated for prediction of pH compartments from hyperpolarized ^{13}C acquisitions of zymonic acid on kidney in healthy mice. CNNs trained with a mixed set of augmented and synthetic spectra show the ability to accurately predict multiple pH compartments in hyperpolarized ^{13}C spectra. This network achieves the best results out of all tested networks and its performance competes with or outperforms conventional line fitting being supervised by humans. The trained network can be used to improve pH mapping by segmentation-based substitution of line fitted pH values by neural network predictions. Therefore, small amounts of experimental data and appropriate neural network and training method choice allows fast, accurate, and reliable evaluation of hyperpolarized ^{13}C magnetic resonance spectroscopic acquisitions for pH measurements in kidney. Using appropriate training data sets and slightly modified output layers of the networks to account for different amounts of detected pH compartments, the presented concept could potentially be applied to other organs or tumours.

Acknowledgements

We acknowledge support from Dr. Geoffrey J. Topping for help with setting up hyperpolarized ^{13}C acquisition protocols. Further, we acknowledge support from Dr. Christian Hundshammer for help with zymonic acid polarization and synthesis and help from Sandra Sühnel with animal experiments.

Authors' contributions

WYRF and JGS designed and implemented the software code of the neural networks. MG acquired MRSI data. WYRF and MG interpreted and compared data from MRSI and neural network data. MG performed supervised line fitting and generated segmentation masks for hybrid pH maps. All authors analyzed the data in this study, BHM and FS designed the study, WYRF, MG, JGS and FS drafted the manuscript. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, Sonderforschungsbereich (SFB) 824, subprojects A7 and Z3, Grant Number 391523415), the Young Academy of the Bavarian Academy of Sciences and Humanities and the European Union's Horizon 2020 research and innovation program under Grant Agreement No 820374.

Availability of data and materials

The data sets used and analyzed during the current study are available at <https://github.com/ryanayf/KidNeYronal>

Declarations

Ethics approval and consent to participate

Experiments were performed in accordance with pertinent laws and regulations and approved by an ethical review board (Regierung von Oberbayern, Munich, Germany, Approval Number ROB-55.2-2532.Vet_02-17-177).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Informatics, Technical University of Munich, 85748 Garching, Germany. ²Department of Nuclear Medicine, TUM School of Medicine, Klinikum Rechts der Isar, Technical University of Munich, 81675 Munich, Germany. ³Munich Institute of Biomedical Engineering, Technical University of Munich, 85748 Garching, Germany.

Received: 12 January 2022 Accepted: 5 April 2022

Published online: 23 April 2022

References

- Gillies RJ, Raghunand N, Garcia-Martin ML, Gatenby RA. pH imaging. A review of pH measurement methods and applications in cancers. *IEEE Eng Med Biol Mag.* 2004;23:57–64. <https://doi.org/10.1109/EMEM.2004.1360409>.
- Adrogue HJ, Madias NE. Management of life-threatening acid-base disorders. Second of two parts. *N Engl J Med.* 1998;338:107–11. <https://doi.org/10.1056/NEJM199801083380207>.
- Gallagher FA, Kettunen MI, Day SE, Hu DE, Ardenkjaer-Larsen JH, Zandt R, Jensen PR, Karlsson M, Golman K, Lerche MH, et al. Magnetic resonance imaging of pH *in vivo* using hyperpolarized ^{13}C -labelled bicarbonate. *Nature.* 2008;453:940–3. <https://doi.org/10.1038/nature07017>.
- Hashim AI, Zhang X, Wojtkowiak JW, Martinez GV, Gillies RJ. Imaging pH and metastasis. *NMR Biomed.* 2011;24:582–91. <https://doi.org/10.1002/nbm.1644>.
- Parks SK, Chiche J, Pouyssegur J. Disrupting proton dynamics and energy metabolism for cancer therapy. *Nat Rev Cancer.* 2013;13:611–23. <https://doi.org/10.1038/nrc3579>.
- Aime S, Castelli DD, Terreno E. Novel pH-reporter MRI contrast agents. *Angew Chem.* 2002;41:4334–6. [https://doi.org/10.1002/1521-3773\(2002115\)41:22%3C4334::AID-ANIE4334%3E3.0.CO;2-1](https://doi.org/10.1002/1521-3773(2002115)41:22%3C4334::AID-ANIE4334%3E3.0.CO;2-1).
- Ardenkjaer-Larsen JH, Fridlund B, Gram A, Hansson G, Hansson L, Lerche MH, Servin R, Thaning M, Golman K. Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. *Proc Natl Acad Sci USA.* 2003;100:10158–63. <https://doi.org/10.1073/pnas.1733835100>.
- Kurhanewicz J, Vigneron DB, Ardenkjaer-Larsen JH, Bankson JA, Brindle K, Cunningham CH, Gallagher FA, Keshari KR, Kjaer A, Laustsen C, et al. Hyperpolarized ^{13}C MRI: path to clinical translation in oncology. *Neoplasia.* 2019;21:1–16. <https://doi.org/10.1016/j.neo.2018.09.006>.
- Cunningham CH, Lau JY, Chen AP, Geraghty BJ, Perks WJ, Roifman I, Wright GA, Connelly KA. Hyperpolarized ^{13}C metabolic MRI of the human heart: initial experience. *Circ Res.* 2016;119:1177–82. <https://doi.org/10.1161/CIRCRESAHA.116.309769>.

10. Grist JT, Miller JJ, Zaccagna F, McLean MA, Riemer F, Matys T, Tyler DJ, Laustsen C, Coles AJ, Gallagher FA. Hyperpolarized ^{13}C MRI: a novel approach for probing cerebral metabolism in health and neurological disease. *J Cereb Blood Flow Metab.* 2020;40:1137–47. <https://doi.org/10.1177/0271678X20909045>.
11. Hundshammer C, Düwel S, Schilling F. Imaging of extracellular pH using hyperpolarized molecules. *Isr J Chem.* 2017;57:788–99. <https://doi.org/10.1002/ijch.201700017>.
12. Korenchan DE, Flavell RR, Baligand C, Sriram R, Neumann K, Sukumar S, VanBrocklin H, Vigneron DB, Wilson DM, Kurhanewicz J. Dynamic nuclear polarization of biocompatible ^{13}C -enriched carbonates for in vivo pH imaging. *Chem Commun (Camb).* 2016;52:3030–3. <https://doi.org/10.1039/c5cc09724j>.
13. Düwel S, Hundshammer C, Gersch M, Feurecker B, Steiger K, Buck A, Walch A, Haase A, Glaser SJ, Schwaiger M, et al. Imaging of pH in vivo using hyperpolarized ^{13}C -labelled zymonic acid. *Nat Commun.* 2017;8:15126. <https://doi.org/10.1038/ncomms15126>.
14. Hundshammer C, Duwel S, Kocher SS, Gersch M, Feurecker B, Scheurer C, Haase A, Glaser SJ, Schwaiger M, Schilling F. Deuteration of hyperpolarized ^{13}C -labeled zymonic acid enables sensitivity-enhanced dynamic MRI of pH. *ChemPhysChem.* 2017;18:2422–5. <https://doi.org/10.1002/cphc.201700779>.
15. Suh EH, Park JM, Lumata L, Sherry AD, Kovacs Z. Hyperpolarized ^{15}N -labeled, deuterated tris(2-pyridylmethyl)amine as an MRI sensor of freely available Zn^{2+} . *Commun Chem.* 2020. <https://doi.org/10.1038/s42004-020-00426-6>.
16. Mishra A, Pariani G, Oerther T, Schwaiger M, Westmeyer GG. Near-infrared photoacoustic imaging probe responsive to calcium. *Anal Chem.* 2016;88:10790–4. <https://doi.org/10.1021/acs.analchem.6b03546>.
17. Schilling F, Schroder L, Palaniappan KK, Zapf S, Wemmer DE, Pines A. MRI thermometry based on encapsulated hyperpolarized Xenon. *ChemPhysChem.* 2010;11:3529–33. <https://doi.org/10.1002/cphc.201000507>.
18. Keshari KR, Kurhanewicz J, Macdonald JM, Wilson DM. Generating contrast in hyperpolarized ^{13}C MRI using ligand-receptor interactions. *Analyst.* 2012;137:3427–9. <https://doi.org/10.1039/c2an35406c>.
19. Chen D, Wang Z, Guo D, Orekhov V, Qu X. Review and prospect: deep learning in nuclear magnetic resonance spectroscopy. *Chemistry.* 2020;26:10391–401. <https://doi.org/10.1002/chem.202000246>.
20. Das D, Coello E, Schulte RF, Menze BH. Quantification of Metabolites in Magnetic Resonance Spectroscopic Imaging Using Machine Learning. In: *Medical image computing and computer assisted intervention—MICCAI 2017.* Springer Singapore, 2017; p 462–70.
21. Brougham DF, Ivanova G, Gottschalk M, Collins DM, Eustace AJ, O'Connor R, Havel J. Artificial neural networks for classification in metabolomic studies of whole cells using ^1H nuclear magnetic resonance. *J Biomed Biotechnol.* 2011;2011: 158094. <https://doi.org/10.1155/2011/158094>.
22. Lee HH, Kim H. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. *Magn Reson Med.* 2019;82:33–48. <https://doi.org/10.1002/mrm.27727>.
23. Chatzidakis M, Botton GA. Towards calibration-invariant spectroscopy using deep learning. *Sci Rep.* 2019;9:2126. <https://doi.org/10.1038/s41598-019-38482-1>.
24. Ghosh K, Stuke A, Todorovic M, Jorgensen PB, Schmidt MN, Vehtari A, Rinke P. Deep learning spectroscopy: neural networks for molecular excitation spectra. *Adv Sci (Weinh).* 2019;6:1801367. <https://doi.org/10.1002/advs.201801367>.
25. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng.* 2010;22:1345–59. <https://doi.org/10.1109/tkde.2009.191>.
26. Guan H, Mingxia L. Domain adaptation for medical image analysis. a survey. 2021; [arXiv:2102.09508](https://arxiv.org/abs/2102.09508).
27. Inoue T, Chaudhury S, De Magistris G, Dasgupta S. Transfer Learning from Synthetic to real images using variational Autoencoders for precise position detection. In: *Proceedings of 25th IEEE international conference on image processing (ICIP)*, pp 2725–29.
28. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436–44. <https://doi.org/10.1038/nature14539>.
29. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: *Proceedings of fourteenth international conference on artificial intelligence and statistics*, p 315–23.
30. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Drop-out: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929–58.
31. LeCun Y, Bottou L, Orr GB, Müller KR. Efficient BackProp. In: *Neural networks: tricks of the trade.* Berlin: Springer; 1998. p. 9–51.
32. Dozat T. Incorporating Nesterov momentum into Adam. In: *Proceedings of national conference on learning representations 2016—workshop track.*
33. Chollet F, et al. e. Keras. GitHub. <https://github.com/fchollet/keras>. Accessed 11 June 2021.
34. Reed GD, von Morze C, Verkman AS, Koelsch BL, Chaumeil MM, Lustig M, Ronen SM, Bok RA, Sands JM, Larson PE, et al. Imaging renal urea handling in rats at millimeter resolution using hyperpolarized magnetic resonance relaxometry. *Tomography.* 2016;2:125–35. <https://doi.org/10.18383/jtom.2016.00127>.
35. Grist JT, Mariager CO, Qi H, Nielsen PM, Laustsen C. Detection of acute kidney injury with hyperpolarized [^{13}C , ^{15}N]Urea and multiexponential relaxation modeling. *Magn Reson Med.* 2020;84:943–9. <https://doi.org/10.1002/mrm.28134>.
36. Laustsen C, Stokholm Norlinger T, Christoffer Hansen D, Qi H, Mose Nielsen P, Bonde Bertelsen L, Henrik Ardenkjaer-Larsen J, Stodkilde Jorgensen H. Hyperpolarized ^{13}C urea relaxation mechanism reveals renal changes in diabetic nephropathy. *Magn Reson Med.* 2016;75:515–8. <https://doi.org/10.1002/mrm.26036>.
37. Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn Reson Med.* 1993;30:672–9. <https://doi.org/10.1002/mrm.1910300604>.
38. Vanhamme L, van den Boogaart A, Van Huffel S. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. *J Magn Reson.* 1997;129:35–43. <https://doi.org/10.1006/jmre.1997.1244>.
39. Topping GJ, Heid I, Trajkovic-Arsic M, Kritzer L, Grashei M, Hundshammer C, Aigner M, Skinner JG, Braren R, Schilling F. Hyperpolarized ^{13}C spectroscopy with simple slice-and-frequency-selective excitation. *Biomedicines.* 2021. <https://doi.org/10.3390/biomedicines902021>.
40. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE.* 1998;86:2278–324. <https://doi.org/10.1109/5.726791>.
41. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging.* 2018;9:611–29. <https://doi.org/10.1007/s13244-018-0639-9>.
42. Yun J, Park JE, Lee H, Ham S, Kim N, Kim HS. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci Rep.* 2019;9:5746. <https://doi.org/10.1038/s41598-019-42276-w>.
43. Naraei P, Abhari A, Sadeghian A. Application of multilayer perceptron neural networks and support vector machines in classification of health-care data. In: *Proceedings of 2016 future technologies conference (FTC).*
44. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention—MICCAI 2015.* vol 9351. Cham: Springer; 2015. p 234–41.
45. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM.* 2017;60:84–90. <https://doi.org/10.1145/3065386>.
46. Lai Z, Deng H. Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron(). *Comput Intell Neurosci.* 2018;2018:2061516. <https://doi.org/10.1155/2018/2061516>.
47. Ahsan MM, Alam TE, Trafalis T, Huebner P. Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and non-COVID-19 patients. *Symmetry.* 2020. <https://doi.org/10.3390/sym12091526>.
48. National Library of Medicine. *ClinicalTrials.gov.* 2021.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.