23Na MRI in ischemic stroke: Acquisition time reduction using postprocessing with convolutional neural networks

Anne Adlung1 | Nadia K. Paschke1 | Alena-Kathrin Golla1,2 | Dominik Bauer1,2 | Sherif A. Mohamed3 | Melina Samartzi4 | Marc Fatar4 | Eva Neumaier-Probst3 | Frank G. Zöllner1,2 | Lothar R. Schad1

1Computer Assisted Clinical Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
2Mannheim Institute for Intelligent Systems in Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
3Department of Neuroradiology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
4Department of Neurology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

Correspondence
Anne Adlung, Computer Assisted Clinical Medicine, Heidelberg University, University Hospital Mannheim, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany.
Email: anne.adlung@medma.uni-heidelberg.de

Funding information
Dietmar Hopp Stiftung

Abstract
Quantitative 23Na magnetic resonance imaging (MRI) provides tissue sodium concentration (TSC), which is connected to cell viability and vitality. Long acquisition times are one of the most challenging aspects for its clinical establishment. K-space undersampling is an approach for acquisition time reduction, but generates noise and artifacts. The use of convolutional neural networks (CNNs) is increasing in medical imaging and they are a useful tool for MRI postprocessing. The aim of this study is 23Na MRI acquisition time reduction by k-space undersampling. CNNs were applied to reduce the resulting noise and artifacts. A retrospective analysis from a prospective study was conducted including image datasets from 46 patients (aged 72 ± 13 years; 25 women, 21 men) with ischemic stroke; the 23Na MRI acquisition time was 10 min. The reconstructions were performed with full dataset (FI) and with a simulated dataset an image that was acquired in 2.5 min (RI). Eight different CNNs with either U-Net–based or ResNet-based architectures were implemented with RI as input and FI as label, using batch normalization and the number of filters as varying parameters. Training was performed with 9500 samples and testing included 400 samples. CNN outputs were evaluated based on signal-to-noise ratio (SNR) and structural similarity (SSIM). After quantification, TSC error was calculated. The image quality was subjectively rated by three neuroradiologists. Statistical significance was evaluated by Student’s t-test. The average SNR was 21.72 ± 2.75 (FI) and 10.16 ± 0.96 (RI). U-Nets increased the SNR of RI to 43.99 and therefore performed better than ResNet. SSIM of RI to FI was improved by three CNNs to 0.91 ± 0.03. CNNs reduced TSC error by up to 15%. The subjective rating of CNN-generated images showed significantly better results than the subjective image rating of RI. The

Abbreviations used: CNN, convolutional neural network; FI, full image (23Na MRI, reconstructed from full k-space data); GM, gray matter; MD, Medical Doctor; ResNet, residual neural network; RI, reduced image (23Na MRI, reconstructed from 25% of k-space data); SNR, signal-to-noise ratio; SSIM, structural similarity index; TSC, tissue sodium concentration; WM, white matter.
acquisition time of $^{23}$Na MRI can be reduced by 75% due to postprocessing with a CNN on highly undersampled data.

**KEYWORDS**

$^{23}$Na MRI, acquisition time reduction, CNN, k-space undersampling, quantitative MRI

---

1 | INTRODUCTION

Sodium magnetic resonance imaging ($^{23}$Na MRI) provides information about physiological and biochemical processes related to cell vitality and viability that cannot be achieved by $^1$H MRI alone. $^{23}$Na MRI can enable the quantification of the tissue sodium concentration (TSC) of the human brain and has been proposed previously for the diagnosis of pathologies defined by a change in cell behavior, vitality or density in pathologies, such as stroke, malignant tumors or multiple sclerosis. Furthermore, additional applications of $^{23}$Na MRI have emerged, for example, quantitative relaxometry and distinguishing between intracellular and extracellular space.

However, $^{23}$Na MRI faces technical challenges that have prevented its clinical establishment to date. The most prominent hurdle is the long acquisition times due to low signal-to-noise ratio (SNR) and biexponential signal decay. $^{23}$Na MRI signal in the human brain is approximately 12,000 times lower than the $^1$H signal. The reason lies within characteristics of the element $^{23}$Na, like its concentration in the human body ($35 \text{ mM } ^{23}\text{Na}$ compared with 80 M $^1\text{H}$ in the brain) or its gyromagnetic ratio $\gamma = 11.42 \text{s}^{-1}\text{T}^{-1}$, which is almost lower by a factor of 4 than for $^1\text{H}$ ($\gamma_{^1\text{H}} = 42.58 \text{s}^{-1}\text{T}^{-1}$). Long acquisition times are particularly problematic for acute diseases like ischemic stroke (“time is brain”) that require fast diagnosis and treatment decisions.

K-space undersampling is one approach to reduce MRI acquisition time, but image quality suffers and artifacts appear. Previously, it has been applied to $^{23}$Na MRI in combination with compressed sensing, and results have been promising. However, some compressed sensing approaches were not applicable for quantifiable sodium data, and others generated a bias. Generally, high errors are produced around high image intensities, which are not favorable for the crucial TSC quantification. Furthermore, compressed sensing suffers from long reconstruction times.

Convolutional neural networks (CNNs) are specific networks in the area of machine learning that were inspired by the physiological process of neuron connectivity. Today, CNNs are emerging for medical imaging preprocessing and postprocessing (e.g. image segmentation and robust $^1$H MRI reconstruction). Furthermore, CNNs successfully improved image quality with the postprocessing of highly undersampled $^1$H MRI data and have been proposed for acquisition time reduction.

In this study, we implemented, tested and compared different CNNs that improve the image quality of highly undersampled $^{23}$Na MRI data from patients with ischemic stroke. In addition to previous publications related to $^1$H MRI, we aimed to maintain the accuracy in quantification of $^{23}$Na MRI. Moreover, the CNN’s ability to predict images with a high SNR, with a high similarity to the ground truth, and with a small TSC quantification error, was evaluated.

The evaluated CNNs are based on U-Net or residual neural network (ResNet) architectures. The U-Net architecture was initially developed by Ronneberger et al. for fast image segmentation. It consists of an encoding and a decoding path. The ResNet implementation is an auto-encoder version. In contrast to the U-Net, it does not use skip connections, but residual connections. U-Nets have previously provided promising results for medical imaging postprocessing and ResNets are supposed to be easier to optimize with smaller training errors.

Furthermore, we tested the effect of batch normalization with the purpose of reducing overfitting. The number of applied filters per block in each convolutional layer was varied for the purpose of optimizing feature extraction. The applied loss functions were the established $L_1$ (absolute difference) and $L_2$ (mean squared error) and an additional loss function called gradient difference loss (GDL), which is supposed to improve edge accuracy and image sharpness.

Ischemic strokes can affect gray matter (GM) and white matter (WM) tissue. Previously, studies have considered the cerebral spinal fluid (CSF) or the vitreous humor as reference for the quantification because it has a sodium concentration of around 145 mM, and therefore generates the highest signal within the cerebral cortex. Inaccuracies in these high-signal areas were one of the major drawbacks in previous studies on k-space undersampling. Therefore, quantification error was evaluated for all three tissue types (CSF, WM and GM) individually for the three best performing networks.

Finally, a crucial aspect for clinical establishment of a diagnostic technique is the acceptance from the performing Medical Doctors (MD). Therefore, subjective evaluation of the image quality by neuroradiologists was considered.
2 | MATERIALS AND METHODS

This study comprised a retrospective evaluation of data from a prospective study that was performed from November 2016 to February 2019. The study included datasets of 46 patients with an acute ischemic stroke (aged 72 ± 13 years), of whom 25 were women and 21 were men. The data were acquired within 48 h after onset of symptoms. The study was approved by our local ethical review board (Med. Ethik-Kommission II) and written informed consent was obtained from all patients or an immediate family member (approval number: 2010-328 N-MA).

2.1 | Image acquisition and processing

$^{23}$Na MR images were acquired at 3 T (Magnetom Trio, Siemens Healthineers, Erlangen, Germany) in addition to a standard $^1$H MRI stroke protocol comprising a FLAIR, diffusion-weighted imaging and dynamic susceptibility contrast perfusion imaging. The entire protocol had a measurement time of 26 min 56 s, of which 10 min were for the $^{23}$Na MRI sequence.

A dual-tuned $^{23}$Na/$^1$H head coil (Rapid Biomedical, Rimpar, Germany) was used for $^{23}$Na image acquisition with a three-dimensional (3D) radial density-adapted sequence,45 reconstruction was performed offline in MATLAB 2015a (MathWorks Inc., Natick, MA, USA). The sequence used a TR of 100 ms, a TE of 0.2 ms, and 6000 equidistant spokes were acquired within a 10 min measurement time resulting in a 3D dataset with a nominal resolution of 4 mm$^3$, calculated based on the full width at half maximum of the point-spread function, the sequence details are provided in Table 1. The sequence is programmed for equidistant projections without golden angle configuration. The reconstruction uses a Hanning filter in k-space and the data was grided using a Kaiser-Bessel window (width = 4 k-space points) prior to the inverse Fourier transformation. A zero-filling factor of 2 was used to achieve an image with an apparent resolution of 2 mm$^3$ and a field of view of 241 mm$^3$ for every patient/dataset.

The Nyquist criterion requires the number of spokes ($N_N$) to be chosen depending on the number of pixels ($P_x$) of the image.

$$N_N = \left(\frac{P_x}{2}\right)^2 \frac{4\pi}{4\pi} = 11,310.$$  

Therefore, the undersampling factor ($R$) of the fully reconstructed images (FI) is:

$$R_{FI} = \frac{N_N}{N_{FI}} = \frac{11,310}{6000} = 1.885.$$  

For the neural network training, images were simulated that could have been acquired within 25% of the measurement time (ie, 2.5 instead of 10 min).

The reduced $^{23}$Na image (RI) of each patient was reconstructed using only 25% of the acquired k-space data. $N_N = 1500$ projections of every dataset were chosen equidistant to each other, which corresponds to image information that could have been acquired with an acquisition time reduced by 75%. Image reconstruction was performed equal to the FI's reconstruction. The undersampling factor ($R_{RI}$) of RI is therefore 4-fold higher compared with FI.

$$R_{RI} = \frac{N_N}{N_{RI}} = \frac{11,310}{1500} = 7.540.$$  

| Parameter                  | 3D radial density-adapted |
|---------------------------|---------------------------|
| Sequence type             | 3D radial density-adapted |
| Repetition time            | 100 ms                    |
| Echo time                 | 0.2 ms                    |
| Number of spokes           | 6000                      |
| Samples per spoke          | 384                       |
| Gradient amplitude         | 4.6 mT/m                  |
| Resolution                | (4.01 mm)$^3$             |
| Measurement time           | 10 min                    |
| Zero-filling               | 2                         |
| Field of view              | (241 mm)$^3$              |
| Matrix                    | (120 voxel)$^3$           |
2.2 CNN implementation and training

A total of eight networks with varying parameters and architectures were implemented for comparison. All the networks consisted of four encoding and four decoding stages, each stage having two to three convolutional layers. The batch size was set to eight and the Adam optimizer was used. The training ran 20 epochs with a learning rate of 0.001. The varying parameters were as follows:

1. Architecture: U-Net and ResNet.
2. Usage of batch normalization: False and True.
3. Number of filters: Small (16, 32, 64, 128, 256, 512, 256, 128, 64, 32) and Large (32, 64, 128, 256, 512, 256, 128, 64, 32). The U-Net architecture had additional residual connections and skip connections that were concatenated, whereas the ResNet architecture did not have such skip connections and concatenations. Both architectures are displayed in Figure 1.

The number of filters represents the applied filters per block in each convolutional layer. As loss functions, we initially considered \( L_1 \) (absolute difference) and \( L_2 \) (mean squared error). After initial evaluation, the three best-performing networks were tested with an additional loss function called gradient difference loss (GDL),

\[
L_{GDL}(\hat{y}, y) = \sum_{i,j} \left| y_{ij} - \hat{y}_{ij} \right|^2 + \sum_{i,j} \left| y_{ij} - \hat{y}_{ij-1} \right|^2,
\]

with \( y \) being the network’s label and \( \hat{y} \) being the network’s output.

The GDL was implemented with the weighting factor \( \lambda = 0.5 \), in addition to \( L_1 \) or \( L_2 \), which resulted in the combined loss function:

\[
L_A = L_1 + \lambda L_{GDL}, \quad \text{and} \quad L_B = L_2 + \lambda L_{GDL}.
\]

The CNNs were implemented in Python 3.5 using Tensorflow 1.10.

Our dataset of 46 patients was split into 38 training and eight test cases; no validation cases were used. The networks were trained and tested in image space after traditional reconstruction. For the network’s training, the RI were used as input, while the FI served as label and ground truth. The networks’ outputs were the artificially upsampled \(^{23}\)Na MR images.

**FIGURE 1** Graphical illustration of the implemented U-Net architecture with residual and skip connections and the ResNet architecture. The main differences between the two architectures are the U-Net’s skip connections with concatenation, which are not present in the ResNet architecture. Each 3 x 3 convolution is illustrated as a horizontal red arrow to the right, and 1 x 1 convolutions are represented by horizontal gray arrows. The downsampling via 2 x 2 Max-Pooling is indicated by a vertical arrow down, and the upsampling via 3 x 3 deconvolutions is shown by a vertical upwards arrow.
The network was trained with two-dimensional 64 x 64 voxel patches. Those patches were from connected transversal slices which are examined by radiologists. We dynamically sampled five random patches per image slice. Fifty central slices were considered per patient, which resulted in 250 samples per patient and therefore a total of 9500 training samples (with 64 x 64 voxels) per training epoch.

The network was tested with 50 central full slices (120 x 120 voxels) per test dataset resulting in 400 test samples.

2.3 Evaluation based on SNR and structural similarity index

The $^{23}$Na MR images were evaluated by calculating the structural similarity index (SSIM) and the SNR. Results were compared between FL (which served as ground truth), RI, and between the CNNs.

SNR calculation was based on masks ($M$) segmenting the dataset into the brain ($M_\text{B}$) and the background ($M_\text{BG}$) regions. The respective masks were calculated by thresholding. The full brain as well as the full background masks were used. SNR of the image ($I$) was defined as the division of mean ($\mu$) signal intensity ($S$) in the brain by the standard deviation ($\sigma$) in the background region:

$$\text{SNR}_I = \frac{\mu(IM_\text{B})}{\sigma(IM_\text{BG})}$$

SSIM calculations were performed with the MATLAB function ssim. The parameter compares image degradation, luminance and contrast between two images. The value can range between 0 and 1, where 1 represents an ideal equivalency between two images and 0 represents no similarities.

2.4 Evaluation of the $^{23}$Na quantification accuracy

For the three best performing CNNs, FL and RI, the $^{23}$Na MR images were quantified and coregistered to each patient’s FLAIR image, which was acquired using the same dual-tuned head coil prior to the $^{23}$Na MRI and during the same measurement session. The FLAIR image served as the basis for image segmentation into the CSF, WM and GM. Segmentation and coregistration were performed with the statistical parametric mapping software SPM12 (Wellcome Centre for Human Neuroimaging, UCL, London, UK); the software generates probability maps for different tissues types. The probabilities were given as values between 0 and 255. All probabilities above the value of 127 (50%) were considered as the evaluated tissue.

TSC quantification was performed based on two reference phantoms that were attached to the patient’s head during data acquisition. The phantoms (cylindrical tubes of 14 ml) contained pure water with 50 mM NaCl (Phantom 1) and 100 mM NaCl (Phantom 2) and 2% agar each. The TSC was calculated by obtaining the average signal intensities in manually defined 3D ROIs for both phantoms, which served as references to perform a three-point linear fit (signal intensity of both phantoms, and zero) that was applied to the image. The ROIs were exclusively defined within the homogenous central region of the phantoms to minimize potential partial volume effects.

Quantification was evaluated by calculating the absolute TSC error (considering FL as ground truth), which was compared with the RI TSC error, indicating relative improvement or deterioration. CSF, WM and GM tissue were considered individually.

Statistical significance for SNR, SSIM and TSC error was evaluated using paired Student’s t-test as the data followed a normal distribution, which was evaluated with the Kolmogorov-Smirnov test. A p-value of below 0.05 was considered significant. Statistical analysis was performed using MATLAB 2018a (MathWorks).

2.5 Subjective image quality rating

Furthermore, three neuroradiologists with differing years of experience (1, 3 and >30 years) evaluated and rated the image quality of the $^{23}$Na MRI versions for all eight test datasets. The evaluation was blinded by randomizing image order and the rating was between 1 (worst) and 5 (best), based on the Likert scale. Statistical significance was evaluated using Friedman’s test. A p-value of less than 0.05 was considered significant.

3 RESULTS

All the CNNs improved the images’ SNR. The U-Net–based networks were also able to improve the SSIM and to reduce the TSC quantification error. Below, a detailed analysis is presented.
3.1 | Evaluation of SNR and SSIM

The traditional image reconstruction including loading of the raw data into MATLAB and the inverse Fourier transformation took approximately 33 s for FI as well as for RI with a GPU NVIDIA GeForce GT630. The network training took 1.5–4.5 h for every CNN configuration. After the networks were trained, the additional CNN postprocessing took approximately 20 s for the full dataset of one patient, which is in addition to the time of the traditional image reconstruction. Table 2 summarizes the performance of the different network architectures to process $^{23}$Na images from highly undersampled data (RI) in terms of SNR and SSIM to FI. SNR and SSIM are given as mean values and standard deviation across all eight test datasets. The average SNR of FI was $21.72 \pm 2.75$, while the average SNR of RI was $10.16 \pm 0.96$, which is less than half of FI’s SNR and, therefore, significantly lower with a $p$ of less than 0.001.

### 3.1.1 | $L_1$ and $L_2$

Considering only the CNNs with loss functions $L_1$ and $L_2$, the SNR of the U-Net–based architectures was between $35.24 \pm 7.08$ and $59.12 \pm 10.99$, with a mean of 43.99. For ResNet- based architectures, SNR was between $15.48 \pm 2.23$ and $69.88 \pm 12.11$, with a mean of 35.31, which was significantly lower with a $p$ of less than 0.001.

Figure 2 shows one slice for each test dataset from FI, RI, and the network architectures CNN 2 ($L_1, L_2, L_A, L_B$) and CNN 6 ($L_1, L_2$). The figure shows the results of artificial upsampling for the different networks and loss functions.

The SSIM between RI and FI is $0.87 \pm 0.03$, whereas it is between $0.69 \pm 0.05$ (CNN 8 $L_1$) and $0.90 \pm 0.03$ (CNN 1 $L_2$, CNN 2 $L_2$) for the CNN output images. Improvement in SSIM could only be observed for CNNs with U-Net architectures. The network configurations with a large number of filters (independent of batch normalization), and those with a small number of filters without batch normalization, improve the SSIM.

### Table 2

List of FI, RI and all CNNs with their generated signal-to-noise ratio (SNR, mean ± SD) and structural similarity index (SSIM) to FI. Networks performing significantly better ($p < 0.05$) than RI are marked in bold.

| CNN | I Architecture | II Number of filters | III Batch Norm. | IV Loss | SNR ($\mu \pm \sigma$) | SSIM ($\mu \pm \sigma$) |
|-----|----------------|---------------------|-----------------|--------|------------------------|------------------------|
| FI  | U-Net Large    | False               | $L_1$           | 21.72 ± 2.75 | 1.00                |
| RI  | Large          | True                | $L_1$           | 10.16 ± 0.96 | 0.87 ± 0.03          |

| 1   | U-Net Large    | False               | $L_1$           | 35.24 ± 07.08 | 0.89 ± 0.03          |
| 2   | U-Net Large    | True                | $L_1$           | 36.67 ± 07.42 | 0.90 ± 0.03          |
| 3   | U-Net Large    | True                | $L_1$           | 35.33 ± 05.54 | 0.88 ± 0.03          |
| 4   | U-Net Small    | False               | $L_1$           | 45.02 ± 07.75 | 0.89 ± 0.03          |
| 5   | U-Net Small    | True                | $L_1$           | 41.31 ± 07.91 | 0.89 ± 0.03          |
| 6   | U-Net Large    | False               | $L_1$           | 49.26 ± 08.80 | 0.85 ± 0.04          |
| 7   | ResNet Large   | False               | $L_1$           | 49.97 ± 07.19 | 0.86 ± 0.04          |
| 8   | ResNet Large   | True                | $L_1$           | 29.11 ± 05.24 | 0.73 ± 0.03          |
| 9   | ResNet Large   | True                | $L_1$           | 25.63 ± 04.63 | 0.73 ± 0.03          |
| 10  | ResNet Large   | True                | $L_1$           | 52.98 ± 10.49 | 0.74 ± 0.04          |
| 11  | ResNet Large   | True                | $L_1$           | 32.51 ± 04.20 | 0.76 ± 0.04          |
| 12  | ResNet Small   | False               | $L_1$           | 30.80 ± 06.39 | 0.73 ± 0.04          |
| 13  | ResNet Small   | True                | $L_1$           | 20.14 ± 02.43 | 0.70 ± 0.04          |
| 14  | ResNet Small   | True                | $L_1$           | 75.51 ± 14.36 | 0.69 ± 0.05          |
| 15  | ResNet Small   | True                | $L_1$           | 15.77 ± 02.60 | 0.74 ± 0.03          |
| 16  | U-Net Large    | False               | $L_A$           | 32.19 ± 05.98 | 0.90 ± 0.03          |
| 17  | U-Net Large    | False               | $L_B$           | 41.32 ± 07.60 | 0.89 ± 0.03          |
| 18  | U-Net Large    | True                | $L_A$           | 34.10 ± 04.72 | 0.91 ± 0.02          |
| 19  | U-Net Large    | True                | $L_B$           | 36.64 ± 06.33 | 0.89 ± 0.03          |
| 20  | U-Net Small    | False               | $L_A$           | 23.55 ± 03.13 | 0.91 ± 0.02          |
| 21  | U-Net Small    | False               | $L_B$           | 40.55 ± 06.74 | 0.89 ± 0.03          |
The improvement was significant with a \( p \) of less than 0.05 for CNN 1 \( L_1 \) and \( L_2 \), CNN 2 \( L_2 \) and CNN 3 \( L_1 \) and \( L_2 \), but not for CNN 2 \( L_1 \) (\( p = 0.0946 \)). Therefore, CNN 1–3 were further evaluated and tested with the additional GDL loss function. The other network configurations deteriorated the structural similarity to FI. CNNs with ResNet-based architectures caused a loss of accuracy in anatomical structures and checkerboard artifacts occurred. Figure 2 illustrates these for the example of CNN 6.

### 3.1.2 Gradient difference loss

The additional application of the GDL loss function deteriorated the mean SNR for the U-Net–based networks to 34.73. At the same time, \( L_{GDL} \) improved the image similarity to FI when added to \( L_1 \), but not when added to \( L_2 \). Maximal SSIM of 0.91 was achieved by adding \( L_{GDL} \) to \( L_1 \) (\( L_A \)) in CNN 2 (U-Net, large number of filters, batch normalization) and CNN 3 (U-Net, small number of filters, no batch normalization). The improvements compared with SSIM from RI were significant (\( p < 0.01 \)).

### 3.2 Evaluation of \( ^{23}\text{Na} \) quantification accuracy

Quantification of the output from the three best performing networks showed that additional CNN postprocessing after traditional reconstruction was able to decrease the TSC error in CSF, GM and WM compared with RI (traditional reconstruction only).

Table 3 displays the absolute TSC error for RI and the output from the different networks, considering FI as ground truth. The \( p \)-value is given for networks that improved TSC accuracy. Overall, the addition of \( L_{GDL} \) to \( L_1 \) or \( L_2 \) had no significant impact on TSC error reduction.

In WM, TSC error was reduced by all CNNs except for CNN 1 \( L_1 \). In GM, the TSC error was also decreased by all CNNs except for CNN 1 \( L_1 \) (deterioration in GM and WM) and CNN 3 \( L_1 \) (improvement in WM only). The TSC quantification error in CSF was decreased with \( L_2 \) for CNN 1, 2 and 3, but was increased for all CNNs with \( L_1 \). The additional \( L_{GDL} \) decreased the TSC error in CSF for CNN 2 \( L_1 \) and \( L_2 \) and for CNN 3 \( L_2 \).

### FIGURE 2

Every row shows one exemplary slice of one \( ^{23}\text{Na} \) MRI test dataset (1–8) in versions of the fully sampled original image (FI), the undersampled reduced image (RI), and the artificially upsampled images CNN2 (\( L_1 \), \( L_2 \), \( L_A \) and \( L_B \)) and CNN6 (\( L_1 \) and \( L_2 \)).
The error reduction in WM was significant for CNN 3 with $L^2$ ($p = 0.0058$) and $LB$ ($p = 0.441$). In GM, the improvement was significant for CNN 3 $L^2$ ($p = 0.0045$) and CNN 2 $L_1$ ($p = 0.0336$). The improvements in the CSF were nonsignificant with a $p$ of less than 0.05. Therefore, CNN 3 ($U$-Net, small number of filters, no batch normalization) with $L^2$ is the only network that decreased TSC quantification error significantly for both tissue types (WM and GM). It also decreased the TSC error in CSF from a mean of 4.21 to 3.94 mM, but the improvement was nonsignificant with a $p$ of 0.0782. The TSC error reduction relative to the TSC error in RI was 15.27% in WM, 14.89% in GM and 6.41% in CSF. CNN 3 $L^2$ plus $LGDL$ ($LB$) also decreased TSC error in WM, GM and CSF. For this network, error reduction was significant in WM (9.51%, $p = 0.0441$), and nonsignificant in GM (8.44%, $p = 0.0882$) and CSF (0.05%, $p = 0.9030$). Its training took 1 h 55 min.

Figure 3 displays one representative image slice from one test dataset. The figure displays the FLAIR image and the quantitative and coregistered $^{23}$Na MRI FI, RI and CNN 3 ($L^2$ and $LB$) output, together with their WM and GM TSC error maps. The CNN-generated images display a lower TSC error compared with RI. The error reduction is especially prevalent in WM.

### 3.3 Subjective image quality rating

The subjective ratings from the three neuroradiologists considered image quality and impression. The ratings for the three different image types (FI, RI, CNN) were averaged for all eight test datasets. Figure 4 illustrates the results as boxplots. RI (1.88 ± 0.78) was rated lower than FI (4.04 ± 0.79) and CNN (3.92 ± 0.81). Friedman’s test showed no significant difference between the ratings of FI and of CNN output ($p = 0.7055$), but there was a significant difference between the ratings of FI and RI and between RI and the CNN-generated images (both $p = 0.0047$).

### 4 DISCUSSION

The aim of the current study was to reduce the $^{23}$Na MRI acquisition time for patients with ischemic stroke. Therefore, we implemented eight different CNNs with different loss functions. The CNNs were trained with a total of 9500 samples (64 x 64 voxels) from 38 patients with ischemic stroke while 400 samples (120 x 120 voxels) from eight patients served as test cases. The networks were implemented with the objective of achieving a 4-fold measurement time reduction without increasing the image’s noise while maintaining TSC quantification accuracy.

### Table 3

| CNN | I Architecture | II Number of filters | III Batch Norm. | IV Loss | WM $\Delta$TSC [mM] | p-value | GM $\Delta$TSC [mM] | p-value | CSF $\Delta$TSC [mM] | p-value |
|-----|----------------|----------------------|-----------------|--------|---------------------|---------|---------------------|---------|---------------------|---------|
| 1   | U-Net Large    | False               | $L_1$           | 4.82   | 4.92                | 6.28    | 4.09                | 0.1383  | 4.08                | 0.2487  | 4.05                | 0.4602  |
| 2   | U-Net Large    | True                | $L_1$           | 4.11   | 0.1512              | 4.01    | 4.01                | 0.0638  | 4.82                |         |                    |         |
| 3   | U-Net Large    | True                | $L_2$           | 4.48   | 0.9055              | 4.11    | 4.11                | 0.1705  | 4.00                | 0.2112  |                    |         |
| 2   | U-Net Large    | False               | $L_A$           | 4.10   | 0.1512              | 4.19    | 4.19                | 0.2533  | 4.00                | 0.1512  | 4.00                | 0.1512  |
| 2   | U-Net Large    | True                | $L_B$           | 3.98   | 0.0625              | 3.90    | 3.90                | 0.0336  | 3.88                | 0.1872  |                    |         |
| 3   | U-Net Large    | True                | $L_A$           | 4.10   | 0.1339              | 3.97    | 3.97                | 0.0511  | 3.82                | 0.0507  |                    |         |
| 3   | U-Net Large    | False               | $L_B$           | 4.25   | 0.3333              | 4.20    | 4.20                | 0.2718  | 4.41                |         |                    |         |

Abbreviations: CNN, convolutional neural network; CSF, cerebral spinal fluid; GM, gray matter; RI, reduced image ($^{23}$Na MRI, reconstructed from 25% of k-space data); TSC, tissue sodium concentration; WM, white matter.
CNNs with U-Net architecture performed better than the ResNet-based CNNs, while the other considered parameters (the number of filters, batch normalization and loss function) individually did not have a prominent impact on the network's performance, but were rather dependent on their deployment and combinations. The preliminary results indicated that artifacts from undersampling were reduced with the application of a properly adjusted U-Net.

A U-Net architecture with the number of filters (16, 32, 64, 128, 256, 128, 64, 32, 16), no batch normalization, and an $L^2$ loss function that could be complemented with a gradient loss function, displayed the best results.

4.1 | SNR and SSIM

All evaluated CNN increased the SNR of the highly undersampled image RI. The relative improvement was between 55% (CNN 8 $L^2$) and a more than 8-fold improvement (CNN 8 $L^1$).
The high improvement of SNR (even compared with FI) could be due to the networks' recognition and elimination of background noise. CNNs have already been used for image denoising, which has also been applied to different medical imaging modalities. According to the SSIM, only CNN 1, 2, and 3 increased the image's similarity to the fully reconstructed image (FI), whereas the index decreased for CNN 4–8. An increased SSIM should indicate an improvement in image quality in the relevant region of the image: the patient's brain. Therefore, the three networks CNN 1, 2, and 3 were chosen for an in-depth evaluation. Only U-Net–based networks increased the SSIM. Therefore, the additional skip connections in the U-Net architecture are considered essential for accurate maintenance of the location information.

The addition of L_GDL increased the SSIM, which depends on image degradation and contrast. Therefore, we hypothesize that L_GDL improves edge accuracy, which was observed previously and was intended.

4.2 Quantification

The accurate quantification of $^{23}$Na MRI is crucial as TSC provides valuable information about cell vitality and viability, which can be of high relevance for patients with ischemic stroke with prospective possible applications like onset time prediction for wake-up stroke patients. Considering FI as ground truth, all three evaluated networks were able to decrease quantification error compared with RI, which was reconstructed traditionally without application of a CNN. The regions of WM and GM can be affected by ischemic stroke. The CNNs decreased the quantification error in both tissues by up to 15%, below a mean TSC error of 4 mM. For patients with ischemic stroke, it has been suggested that tissue with an increased TSC of more than 50% could be irreversibly damaged and the literature suggests healthy tissue values of 20–60 mM for WM and 30–70 mM for GM. Those findings suggest that a 4 mM variation might not be of high relevance, but sufficiently accurate. The CSF, due to its high signal intensities, is sometimes used as a quantification reference. The TSC error in CSF was similar compared with WM and GM TSC error and was also reduced to less than 4 mM.

On one hand, compressed sensing has proven to be a robust technique without the risk of introducing false information, which has an advantage over CNN-based postprocessing techniques. On the other hand, the similarly high accuracy in high signal intensity tissue compared with lower signal intensity tissue as presented in this paper represents an important advantage compared with the previously investigated compressed sensing for k-space undersampling of $^{23}$Na MRI. The relatively fast reconstruction times represent an additional advantage compared with compressed sensing methods for k-space undersampling.

The addition of L_GDL enhanced quantification accuracy. Improvement was significant for WM, but not for GM. It can be hypothesized that the positive impact on edge accuracy from L_GDL might only improve contours, but has no to little impact on the overall signal intensities, and therefore the quantification. Further evaluation on whether different weighting factors $\lambda$ could improve the overall performance will follow as the results of the SSIM are promising, and the additional maintenance of edge accuracy can be of high value for $^{23}$Na MRI where resolution is generally low.

4.3 Subjective image quality rating

For the establishment of a new diagnostic tool in the clinical protocol, acceptance by MDs is a particularly relevant factor. Therefore, in addition to meeting the objective of quantifiable evaluation, subjective evaluation conducted by neuroradiologists was of considerable importance. Their ratings show that the image evaluation was lower for RI than for FI, as was expected. The 75% reduced acquisition time caused artifacts and noise. The application of a CNN was able to decrease the negative impact from undersampling on the image quality up to the point where there was no longer any significant difference in the subjective image quality rating between the image that was acquired within 10 min (and traditional reconstruction, FI) and the image that could have been acquired within 2.5 min with the postprocessing of a CNN, in addition to the traditional reconstruction.

4.4 Limitations

The current study's limitations include FI as the label and ground truth, and the size and homogeneity of the dataset. FI already has an undersampling factor of $R_S = 1.885$. Furthermore, the repetition time is set to 100 ms, which is below 5 x T1 (up to 55 ms), so that T1-weighting effects can be expected. The parameters were chosen as the acquisition time of 10 min for the $^{23}$Na MRI cannot be exceeded in our clinical stroke protocol. The network was exclusively trained and tested with datasets from 46 patients with acute ischemic stroke, who were aged 72 ± 13 years. To apply our method more generally, datasets from patients from a more heterogeneous age group and with
different pathologies are warranted and will be investigated in the future. The network was also trained and tested exclusively with images of a 4 x 4 x 4 mm³ resolution; no predictions of the network’s performance for images with higher resolutions can be made at this point.

Furthermore, the networks’ training and testing used DICOM data (image space) and not the complex MRI values from k-space, which implies that only part of the information was used. Therefore, further improvement of the performance could be expected by using the complex MRI data, as was already suggested and implemented previously.55,56

The image evaluation depends on the 23Na MRI coregistration to the FLAIR image and segmentation into GM, WM and CSF, both of which are performed with SPM 12. The software is well established and has displayed high accuracy.57,58

Another crucial drawback of all CNN-based postprocessing techniques is the potential introduction of false information into the image. Here, this issue was addressed by the quantified comparison of the CNN output images with those of the FI as well as the subjective image quality rating by the neuroradiologists.

4.5 | Conclusions

CNNs based on a U-Net architecture with residual connections proved to improve the image reconstruction of 4-fold undersampled 23Na MRI data resulting in an acquisition time of only 2.5 min and a total image reconstruction time of less than 1 min. Reconstructed images of the best performing CNN (CNN 3: U-Net, a small number of filters, no batch normalization) with L2 reduced mean TSC error below 4 mM compared with traditionally reconstructed images with a 4-fold longer acquisition time.

We showed that high undersampling of 23Na MRI is feasible for patients with acute ischemic stroke where acquisition time is crucial. We suggest using an appropriately designed and trained CNN to generate accurate and quantifiable images from the undersampled data.

The suggested imaging technique can be applied with any 3 T scanner, which is enabled for 23Na imaging and has dedicated coils available. It allows for further exploring 23Na MRI and for showing TSC to be a biomarker for tissue viability and cell vitality. Prospectively, this might have an impact on future diagnostics or even therapy decisions regarding ischemic stroke.

ACKNOWLEDGEMENTS

The authors would like to thank Victor Saase for participating in the subjective image quality rating. Shradha Jain, Dr Michaela A.U. Hoesl, and Valerie Klein carefully proofread the manuscript, which was much appreciated.

The study “Natrium (23Na) MRT zur Diagnostik von ischämischem aber noch vitalem Hirngewebe bei Schlaganfallpatienten” was funded by Dietmar-Hopp Stiftung.

Open access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

Research data are not shared.

FUNDING INFORMATION

The study “Natrium (23Na) MRT zur Diagnostik von ischämischem aber noch vitalem Hirngewebe bei Schlaganfallpatienten” was funded by Dietmar-Hopp Stiftung.

ORCID

Anne Adlung https://orcid.org/0000-0001-7032-9782

REFERENCES

1. Hilal SK, Maudsley AA, Ra JB, et al. In vivo NMR imaging of sodium-23 in the human head. J Comput Assist Tomogr. 1985;9(1):1-7.
2. Thulborn KR, Davis D, Adams H, Gindin T, Zhou J. Quantitative tissue sodium concentration mapping of the growth of focal cerebral tumors with sodium magnetic resonance imaging. Magn Res Med. 1999;41(2):351-359.
3. Ouwerkerk R, Bleich KB, Gillen JS, Pomper MG, Bottomley PA. Tissue sodium concentration in human brain tumors as measured with 23Na MR imaging. Radiology. 2003;227(2):529-537.
4. Lachner S, Ruck L, Niesporek SC, et al. Comparison of optimized intensity correction methods for 23Na MRI of the human brain using a 32-channel phased array coil at 7 Tesla. Z Med Phys. 2019;104-115.
5. Neumaier-Probst E, Konstandin S, Ssozi J, et al. A double-tuned 1H/23Na resonator allows 1H-guided 23Na-MRI in ischemic stroke patients in one session. Int J Stroke. 2015;10(SA100):56-61.
6. Wetterling F, Ansar S, Handwerker E. Sodium-23 magnetic resonance imaging during and after transient cerebral ischemia: multinuclear stroke protocols for double-tuned (23)Na/(1)H resonator systems. Phys Med Biol. 2012;57(21):6929-6946.
7. Boada F, LaVerde G, Jungreis C, Nemoto E, Tanase C. Triple/single quantum filtered sodium MRI of acute brain ischemia. Engine Med Biology Society, 2005 IEEE-EMBS 2005 27th Annual Intern Conf. 2006;731-734.
8. Haneder S, Giordano FA, Konstandin S, et al. (23)Na-MRI of recurrent glioblastoma multiforme after intraoperative radiotherapy: technical note. Neuroradiology. 2015;57(3):321-326.

9. Boada F, Tanase C, Davis D, et al. Non-invasive assessment of tumor proliferation using triple quantum filtered/sup 23/Na MRI: technical challenges and solutions. IEEE. 2004;2:5238-5241.

10. Biller A, Badde S, Nagel A, et al. Improved brain tumor classification by sodium MRI imaging: prediction of IDH mutation status and tumor progression. Am J Neuroradiol. 2016;37(1):66-73.

11. Worthoff WA, Shymanskaya A, Lindemeyer J, Langen KJ, Shah NJ. Relaxometry and quantification in sodium MRI of cerebral gliomas: A FET-PET and MRI small-scale study. NMR Biomed. 2020;33(10):e4361. https://onlinelibrary.wiley.com/doi/full/10.1002/nbm.4361

12. Donadie M, Le Fur Y, Maarouf A, et al. Metabolic counterparts of sodium accumulation in multiple sclerosis: A whole brain (23)Na-MRI and fast (1)H-MR study. Mult Scler. 2019;25(1):39-47.

13. Eiselle P, Konstandin S, Szabo K, et al. Temporal evolution of acute multiple sclerosis lesions on serial sodium (23Na) MRI. Mult Scler Relat Disord. 2019;29:48-54.

14. Madelin G, Oesingmann N, Nieelles-Vallespin S, Herbert J, Johnson G, Inglese M. 3T sodium MRI of patients with Multiple Sclerosis. Proc Intl Soc Mag Reson Med. 2008;16:3444.

15. Madelin G, Kline R, Walvick R, Regatte RR. A method for estimating intracellular sodium concentration and extracellular volume fraction in brain in vivo using sodium magnetic resonance imaging. Sci Rep. 2014;4(4763). https://www.nature.com/articles/srep04763

16. Worthoff WA, Shymanskaya A, Shah NJ. Relaxometry and quantification in simultaneously acquired single and triple quantum filtered sodium MRI. Magn Reson Med. 2018;81(1):303-315. https://onlinelibrary.wiley.com/doi/full/10.1002/mrm.27387

17. Thulborn KR. Quantitative sodium MR Imaging: A review of its evolving role in medicine. Neuroimage. 2018.168:250-268.

18. Hu R, Kleinmaier D, Malzacher M, Hoesl MA, Paschke NK, Schad LR. X-nuclei imaging: Current state, technical challenges, and future directions. J Magn Reson Imaging. 2020;51(2):355-376.

19. Ouwerverk R. Sodium magnetic resonance imaging: from research to clinical use. J Am Coll Radiol. 2007;4(10):739-741.

20. Feinberg DA, Crooks L, Kaufman L, et al. Magnetic resonance imaging performance: a comparison of sodium and hydrogen. Radiology. 1985;156(1):133-138.

21. Saver JL. Time is brain—quantified. Stroke. 2006;37(1):263-266.

22. Marler JR, Tilley B, Lu M, et al. Early stroke treatment associated with better outcome: the NINDS rt-PA stroke study. Neurology. 2000;55(11):1649-1655.

23. Kim BJ, Kang HG, Kim H-J, et al. Magnetic resonance imaging in acute ischemic stroke treatment. J Stroke. 2014;16(3):131-145.

24. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn Reson Med. 2007;57(6):1086-1098.

25. Blunck Y, Kolbe SC, Moffat BA, Ordidge RJ, Cleary JO, Johnston LA. Compressed sensing effects on quantitative analysis of undersampled human brain sodium MRI. Magn Reson Med. 2020;83(3):1025-1033.

26. Behl NG, Gnahm C, Bachert P, Ladd ME, Nagel AM. Three-dimensional dictionary-learning reconstruction of (23)Na MRI data. Magn Reson Med. 2016;75(4):1605-1616.

27. Madelin G, Chang G, Otazo R, Jerschow A, Regatte RR. Compressed sensing sodium MRI of cartilage at 77T: preliminary study. J Magn Reson. 2012;214(1):360-365.

28. Utschneider M, Behl NG, Lachner S, et al. Accelerated quantification of tissue sodium concentration in skeletal muscle tissue: quantitative capability of dictionary learning compressed sensing. Magnetic Reso Mater Phy, Bio Med. 2020;1-11.

29. Gnahm C, Bock M, Bachert P, Semmler W, Behl NG, Nagel AM. Iterative 3D projection reconstruction of 23Na data with an 1H MRI constraint. Magn Reson Med. 2014;71(5):1720-1732.

30. Schlemper J, Caballero J, Hajnal JV, Price A, Rueckert D. A deep cascade of convolutional neural networks for MR image reconstruction. International Conference on Information Processing in Medical Imaging. Springer, Cham. 2018;747-758. https://link.springer.com/chapter/10.1007/978-3-319-59050-9_51

31. Bernal J, Kushihar K, Asfaw DS, et al. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif Intell Med. 2019;95:64-81.

32. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys. 2019;29(2):102-127.

33. Eo T, Jun Y, Kim T, Jang J, Lee H, Hwang D. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magn Reson Med. 2018;80(5):2188-2201.

34. Plenge E, Poot DH, Bernsen M, et al. Super-resolution methods in MRI: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? Magn Reson Med. 2012;68(6):1983-1993.

35. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, Cham; 2015:234-241. https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28

36. Kaiming H, Xiangyu Z, Shaoqing R, Jian S. Deep residual learning for image recognition. arXiv Preprint arXiv:151203385. 2015.

37. Agrawal A, Mittal N. Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. Visual Computer. 2020;36(2):405-412.

38. Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error. arXiv Preprint arXiv:151105440; 2015.

39. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. arXiv Preprint arXiv:1502.03167. 2015.
44. Madelin G. Biomedical applications of sodium MRI in vivo. *J Magn Reson Imaging*. 2013;38(3):511-529.
45. Nagel AM, Laun FB, Weber MA, Matthies C, Semmler W, Schad LR. Sodium MRI using a density-adapted 3D radial acquisition technique. *Magn Reson Med*. 2009;62(6):1565-1573.
46. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv Preprint arXiv:1412.6980*. 2014. https://arxiv.org/abs/1412.6980
47. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612.
48. Niesporek SC, Hoffmann SH, Berger MC, et al. Partial volume correction for in vivo 23Na-MRI data of the human brain. *Neuroimage*. 2015;112:353-363.
49. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans Image Process*. 2017;26(7):3142-3155.
50. Koppers S, Coussoux E, Romanzetti S, Reetz K, Merhof D. Sodium image denoising based on a convolutional denoising autoencoder. *Bildverarbeitung für die Medizin*. Springer Vieweg; 2019:98-103.
51. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. *Jpn J Radiol*. 2019;37(1):73-80.
52. LaVerde GC, Jungreis CA, Nemoto E, Boada FE. Sodium time course using 23Na MRI in reversible focal brain ischemia in the monkey. *J Magn Reson Imaging*. 2009;30(1):219-223.
53. Madelin G, Regatte RR. Biomedical applications of sodium MRI in vivo. *J Magn Reson Imaging*. 2013;38(3):511-529.
54. Madelin G, Lee JS, Regatte RR, Jerschow A. Sodium MRI: methods and applications. *Prog Nucl Magn Reson Spectrosc*. 2014;79:14-47.
55. Wang S, Su Z, Ying L, et al. Accelerating magnetic resonance imaging via deep learning. *IEEE*. 2016;514-517.
56. Han Y, Sunwoo L, Ye JC. k-space deep learning for accelerated MRI. *IEEE Trans Med Imaging*. 2019;377-386.
57. Malone IB, Leung KK, Clegg S, et al. Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance. *Neuroimage*. 2015;104:366-372.
58. Fellhauer I, Zöllner FG, Schröder J, et al. Comparison of automated brain segmentation using a brain phantom and patients with early Alzheimer’s dementia or mild cognitive impairment. *Psychiatry Res Neurol*. 2015;233(3):299-305.

**How to cite this article:** Adlung A, Paschke NK, Golla A-K, et al. 23Na MRI in ischemic stroke: Acquisition time reduction using postprocessing with convolutional neural networks. *NMR in Biomedicine*. 2021;34:e4474. [https://doi.org/10.1002/nbm.4474](https://doi.org/10.1002/nbm.4474)