Exploring Generalization Capacity of Artificial Neural Network for Myelin Water Imaging

Jieun Lee¹, Joon Yul Choi¹,², Dongmyung Shin¹, Eung Yeop Kim³, Se–Hong Oh⁴, Jongho Lee¹

¹Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea
²Epilepsy Center, Neurological Institute, Cleveland Clinic, OH, USA
³Department of Radiology, Gil Medical Center, Gachon University College of Medicine, Incheon, Korea
⁴Division of Biomedical Engineering, Hankuk University of Foreign Studies, Gyeonggi-do, Korea

Purpose: To understand the effects of datasets with various parameters on pre-trained network performance, the generalization capacity of the artificial neural network for myelin water imaging (ANN-MWI) is explored by testing datasets with various scan protocols (i.e., resolution and refocusing RF pulse shape) and types of disorders (i.e., neuromyelitis optica and edema).

Materials and Methods: ANN-MWI was trained to generate a T₂ distribution, from which the myelin water fraction value was measured. The training and test datasets were acquired from healthy controls and multiple sclerosis patients using a multi-echo gradient and spin-echo sequence with the same scan protocols. To test the generalization capacity of ANN-MWI, datasets with different settings were utilized. The datasets were acquired or generated with different resolutions, refocusing pulse shape, and types of disorders. For all datasets, the evaluation was performed in a white matter mask by calculating the normalized root-mean-squared error (NRMSE) between the results from the conventional method and ANN-MWI. Additionally, for the patient datasets, the NRMSE was calculated in each lesion mask.

Results: The results of ANN-MWI showed high reliability in generating myelin water fraction maps from the datasets with different resolutions. However, the increased errors were reported for the datasets with different refocusing pulse shapes and disorder types. Specifically, the region of lesions in edema patients reported high NRMSEs. These increased errors indicate the dependency of ANN-MWI on refocusing pulse flip angles and T₂ characteristics.

Conclusion: This study proposes information about the generalization accuracy of a trained network when applying deep learning to processing myelin water imaging.

Keywords: Artificial neural network; Myelin water imaging; T₂ relaxation; Deep learning; Generalization capacity

INTRODUCTION

Applying deep learning to MR data post-processing has been demonstrated to improve performance when compared to conventional processing methods (1–4).
For example, Yoon et al. (1) reconstructed a quantitative susceptibility map in high quality from a single head orientation using a deep neural network. Cohen et al. (2) showed a fast and noise-robust deep learning approach to reconstruct MR fingerprinting data. Kwon et al. (3) accelerated a reconstruction process of under-sampled MR images by applying a deep learning method.

Even though these deep learning approaches have shown exciting results with much improved data processing efficiency, the utility of the methods in practical applications is still limited because the data with various conditions (e.g., different scan parameters or different patient types) do not guarantee the performance. Recent studies have suggested that the performance of a neural network can highly depend on the characteristics of the training dataset (5-9). For example, a trained network tends to underestimate quantitative values in lesions with high magnetic susceptibility (7) or degrade the reconstruction quality when the images have different signal-to-noise ratios (8). Since the parameters such as scan protocols, image quality measurements, and disease types have effects on MR images (10, 11), it is necessary to explore the reliability of a trained model in various conditions.

Recently, we proposed an artificial neural network-based T₂ myelin water imaging (ANN-MWI) reconstruction method and demonstrated real-time processing of MWI (12). ANN-MWI was designed to generate a T₂ distribution, from which myelin water fraction (MWF) is calculated (13, 14). In the paper, the test dataset had the same scan protocol, noise level, and type of disease as the training dataset. Since the ANN-MWI utilized the same settings between the training and test datasets, it might generate high-quality results for the test dataset. When the datasets with different TEs and noise levels were tested, the network showed dependency on TEs, whereas not on noise levels. This examination represents that the performance of the pre-trained network can be affected by other varying parameters, which should be considered when applying deep learning to MWI processing. Therefore, further investigation is necessary to test the generalization performance of ANN-MWI for datasets of various conditions. In this study, we examine the reliability of ANN-MWI using the datasets of varying scan protocols (resolution and refocusing RF pulse shape) and types of diseases (neuromyelitis optica [NMO] and edema).

**MATERIALS AND METHODS**

**MRI Data**

For MWI, a 3D multi-echo gradient and spin-echo (GRASE) sequence (15) was implemented at a 3T MRI scanner ( Trio Siemens, Erlangen, Germany) with a 32-channel phased-array head coil. The default scan parameters were set as follows: repetition time (TR) = 1000 ms (or longer to reduce a specific absorption rate); echo time (TE) = 10, 20, ..., 320 ms (32 echoes with 10 ms echo spacing); flip angle = 90°.

As an original dataset for ANN-MWI, GRASE MWI datasets (field of view [FOV] = 240 × 180 × 112 mm³; resolution = 1.5 × 1.5 × 4.0 mm³; sinc RF shape) of 20 subjects from the previously published studies (16, 17) were utilized. Among the datasets, twelve datasets (6 healthy controls [HC] and 6 multiple sclerosis [MS] patients) were used for training, whereas eight datasets (3 HC and 5 MS) were used as a test set. Other scan parameters of the original dataset were as follows: number of echoes, 32; flip angle, 90°; EPI factor, 3; the first echo time, 10 ms; echo spacing, 10 ms; TR, 1000 ms and for a few subjects, longer TRs were used due to a specific absorption rate limit.

In order to confirm the effect of acquisition parameters (i.e., spatial resolution and refocusing RF shape), and disease type, the datasets with different scan protocols and types of disorders were generated or acquired. Detailed scan information is described below:

1) Spatial resolution: two different resolutions, 2.0 × 2.0 × 4.0 mm³ and 2.5 × 2.5 × 4.0 mm³, were generated by truncating the k-space data from the eight datasets that were used as the test set (3 HC and 5 MS). The resolution of the original data was 1.5 × 1.5 × 4.0 mm³.
2) RF shape: the Shinnar-Le Roux (SLR) designed refocusing pulse (time-bandwidth product = 2) (18) data were acquired in 14 HC (FOV = 240 × 168 × 150 mm³; resolution = 1.5 × 1.5 × 5.0 mm³). The original dataset was acquired with a sinc-shaped refocusing RF (time-bandwidth product = 4).
3) Disease types: different types of disorders were tested by utilizing the datasets of twenty-three NMO patients (19) and seven edema patients (FOV = 280 × 168 × 130 mm³; resolution = 1.7 × 1.7 × 5.0 mm³) in order to check the influence of disease types on the performance. The other parameters were the same as the original dataset used for the network training.

Except for the datasets with the SLR-designed RF pulse and edema patients, the following conventional MR images...
were utilized to identify lesions: 2D spin-echo T₁-weighted images (resolution, 0.8 × 0.7 mm²; slice thickness, 3.2 mm; number of slices, 32; TR/TE, 550/9.2 ms; flip angle, 70°), 2D fast spin-echo T₁-weighted images (resolution, 0.5 × 0.5 mm²; slice thickness, 3.2 mm; number of slices, 32; TR/TE, 8750/90 ms; number of echoes, 21), and 2D fluid-attenuated inversion recovery (FLAIR) images (resolution, 0.7 × 0.7 mm²; slice thickness, 3.2 mm; number of slices, 32; TR/TE, 9000/87 ms; inversion time, 2500 ms; number of echoes, 16).

ANN-MWI

To explore the reliability of ANN-MWI, all datasets were processed using both conventional and ANN-MWI methods (Fig. 1).

Using the conventional method as a reference, the multi-echo GRASE data were processed voxel-wise to generate a stimulated echo-corrected T₂ distribution (14) with commonly used parameters (20-23) as follows: the number of T₂ time points, 120; T₂ values, logarithmically spaced from 0.015 to 2 s; and a chi-square regularizer with a target misfit of 1.02. From this T₂ distribution, MWI was generated using the area of myelin water (from 15 to 40 ms) over the total distribution area (from 15 to 2000 ms).

On the other hand, for ANN-MWI, the previously trained network (12) was utilized as data inference. ANN-MWI generates a T₂ distribution of 120-time points when the 32-echo data are fed as an input. The network structure was composed of an input layer (32 neurons) and an output layer (120 neurons) with 7 fully-connected hidden layers (160, 240, 320, 360, 480, 520, and 600 neurons each). The training parameters were set as follows: activation function = a leaky rectified linear unit (24) with 0.2 alpha; optimizer = Adam (25); learning rate = starting from 0.001 and decreased by one-tenth at epochs of 900, 1200, 1500, 1800 (26); batch size = 2, 3, 4... 2,002 for each epoch (27); loss function = mean squared error between labels and outputs. The source code for the network is available at GitHub (https://github.com/snu-list/ANN-MWI).

Data Analysis

To test the performance of ANN-MWI with different datasets, a white matter mask was applied. The white matter mask was generated with a method proposed by Choi et al. (16). After generating a white matter mask using T₁ over T₂-weighted images, voxels with unrealistic MWF values higher than 30% and zero values (28, 29) were excluded from the datasets (i.e., different resolution and NMO patient) that have the T₁- and T₂-weighted images. On the other hand, for the other datasets (i.e., SLR-designed RF pulse and edema patients), the mask was generated by extracting the brain using the brain mask (30), and then excluding unrealistic MWF voxels (i.e., MWF ≥ 30% and MWF = 0). To detect the region of lesions in the patient data, a threshold was applied to the FLAIR image from the MS and NMO patients. For the edema patients, the lesion mask was produced by applying a threshold to the T₂-weighted image and then, overlaying manually segmented

![Fig. 1. Exploring the generalization capacity of the artificial neural network for myelin water imaging. The network was trained with 6 MS and 6 HC data that were acquired using a resolution of 1.5 × 1.5 × 4.0 mm³ and sinc shape RF. Two different types of datasets, different in scan protocols (resolutions = 2.0 × 2.0 × 4.0 and 2.5 × 2.5 × 4.0 mm³; RF shape = SLR shape) and different disease types (NMO and edema patients), were utilized for generalization. HC = healthy control; MS = multiple sclerosis; NMO = neuromyelitis optica; SLR = Shinnar-Le Roux](https://github.com/snu-list/ANN-MWI)
lesions. These lesion masks were superimposed onto the white matter mask to include the voxels of the lesions that were extracted.

The MWIs from the conventional method and ANN-MWI were compared to check the visual difference. To analyze the results from ANN-MWI, the normalized root-mean-square error (NRMSE) of MWF was evaluated in the white matter mask with the conventional MWI results as a reference. Additionally, for the MS, NMO, and edema patients, the NRMSE was analyzed in the lesion masks.

**RESULTS**

For the test set that has the same parameters as those of the training data, the NRMSE of ANN-MWI was 2.26 ± 0.20%. The results of MWI for the different scan protocols are summarized in Figure 2 for the different resolutions and in Figure 3 for the different RF shapes. The mean NRMSE of the low-resolution datasets was 2.19 ± 0.24% for 2.0 × 2.0 × 4.0 mm$^3$ and 2.21 ± 0.25% for 2.5 × 2.5 × 4.0 mm$^3$ (Fig. 2). Since these errors are similar to that of the original resolution (2.26 ± 0.20%), the results suggest that the resolution does not affect the performance of ANN-MWI. When the results of the different RF shape data were examined, they reported a significantly higher NRMSE of 4.81 ± 0.79% (Fig. 3b), indicating the importance of the RF shape.

In Figure 4, the MWF maps of the four different subject types (HC, MS, NMO, and edema) are displayed with the error maps. The error maps were generated by multiplying the absolute difference between the maps from the conventional method and ANN-MWI by 10 in order to visualize different regions. The average NRMSEs were 2.33 ± 0.06% in 3 HC, 2.21 ± 0.24% in 5 MS, 2.21 ± 0.32% in
23 NMO, and 3.84 ± 0.22% in 7 edema patients. Although the MWF maps from both methods look almost identical, the difference map of the edema data reveals significantly larger errors, particularly in the edema lesion. When the lesion was analyzed in each lesion mask, the mean NRMSEs were 4.59 ± 0.87% in 5 MS, 3.27 ± 1.12% in 23 NMO, and 9.59 ± 1.88% in 7 edema patients, demonstrating increased NRMSEs in lesions.

**DISCUSSION**

In this study, we explored the generalization capacity of a recently proposed neural network, ANN-MWI, for MWI. Our results demonstrate the feasibility of the network on datasets with different resolutions.

The increased errors were reported for the datasets with the different refocusing RF shape, suggesting the dependency of ANN-MWI on the refocusing RF pulse profile. This outcome of the different refocusing RF shape...
may be explained by the Extended Phase Graph algorithm (31) process for different RF pulse flip angles (32). These errors might be reduced by fine-tuning the pre-trained network with a few data using transfer learning (8, 33).

In terms of disease types, ANN-MWI showed similar results in the NMO data to the conventional process but increased errors in the edema data. This outcome can be explained by the NRMSEs in the lesion mask. When compared to the NRMSE in the white matter mask, larger errors were reported in the lesion masks for all three patient datasets. The degraded performance in MS and NMO lesions may be caused by a relatively small number of lesion voxels (approximately 0.3% of the total data number) in the training data (34). However, the errors in the edema lesion mask were large (9.59 ± 1.88%) when compared to those in the MS (4.59 ± 0.87%) and NMO (3.27 ± 1.12%) lesion masks. The increased errors in the edema lesion might be caused by different $T_2$ characteristics of the edema lesion that were not included in the training set. To decrease these errors, the data augmentation method (7, 35) could be utilized, which may expand the generalization capacity of ANN-MWI.

The resolution of edema patient data (1.7 × 1.7 × 5.0 mm$^3$) was different from that of the training data (1.5 × 1.5 × 4.0 mm$^3$). Since changing the resolution was demonstrated to have few effects on the performance, the difference in resolution may not have any influence on the increased errors in the edema patient data.

All the data used in this work were obtained by the same MR scanners manufactured by the same vendor (Siemens). Since the MWF values reconstructed by the conventional method were demonstrated to be reproducible in multiple vendors (36), ANN-MWI can be applied to the data from other manufacturers. However, as the scan protocols can be varied with different scanners, further tests may be required to ensure the reliability of ANN-MWI on the data from different scanners.

In conclusion, the generalization capacity of ANN-MWI was explored by testing datasets with different scan protocols and types of disorders. Our results have demonstrated the feasibility of the ANN-MWI on datasets with different resolutions. On the other hand, the ANN-MWI may have a dependency on different RF shapes and disorder types. This study provided insight towards considerations for applying a deep learning approach to myelin water imaging.

Acknowledgments

This work was supported by the Creative-Pioneering Researchers Program and the Institute of Engineering Research in Seoul National University, the Hankuk University of Foreign Studies Research Fund, Advanced MR Study Group of 2019’ Korean Society of Magnetic Resonance in Medicine, and the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science ICT & Future Planning (NRF-2015M3C7A1031969).

REFERENCES

1. Yoon J, Gong E, Chatnuntawech I, et al. Quantitative susceptibility mapping using deep neural network: QSMnet. Neuroimage 2018;179:199–206
2. Cohen O, Zhu B, Rosen MS. MR fingerprinting Deep ReConstruciton NEtwork (DRONE). Magn Reson Med 2018;80:885–894
3. Kwon K, Kim D, Park H. A parallel MR imaging method using multilayer perceptron. Med Phys 2017;44:6209–6224
4. Lee D, Lee J, Ko J, Yoon J, Ryu K, Nam Y. Deep learning in MR image processing. Investig Magn Reson Imaging 2019;23:81–99
5. Ben-David S, Blitzer J, Crammer K, Pereira F. Analysis of representations for domain adaptation. Advances in Neural Information Processing Systems (NIPS) 2006:137–144
6. Mårtensson G, Ferreira D, Granberg T, et al. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. Med Image Anal 2020;66:101714
7. Jung W, Yoon J, Ji S, et al. Exploring linearity of deep neural network trained QSM: QSMnet†. Neuroimage 2020;211:116619
8. Knoll F, Hammernik K, Kobler E, Pock T, Recht MP, Sodickson DK. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. Magn Reson Med 2019;81:116–128
9. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530, 2016
10. Youn SW, Kwon OD, Hwang MJ. Multi-parametric quantitative MRI for measuring myelin loss in hyperglycemia-induced hemichorea. Investig Magn Reson Imaging 2019;23:148–156
11. Seo JP, Kwon YH, Jang SH. Mini-review of studies reporting the repeatability and reproducibility of diffusion tensor imaging. Investig Magn Reson Imaging 2019;23:26–33
12. Lee J, Lee D, Choi JY, Shin D, Shin HG, Lee J. Artificial neural network for myelin water imaging. Magn Reson Med 2020;83:1875–1883
13. MacKay A, Whittall K, Adler J, Li D, Paty D, Graeb D. In
vivo visualization of myelin water in brain by magnetic resonance. Magn Reson Med 1994;31:673-677
14. Prasloski T, Madler B, Xiang QS, MacKay A, Jones C. Applications of stimulated echo correction to multicomponent T2 analysis. Magn Reson Med 2012;67:1803-1814
15. Prasloski T, Rauscher A, MacKay AL, et al. Rapid whole cerebral myelin water imaging using a 3D GRASE sequence. Neuroimage 2012;63:533-539
16. Choi JY, Jeong IH, Oh SH, et al. Evaluation of normal-appearing white matter in multiple sclerosis using direct visualization of short transverse relaxation time component (ViSTA) myelin water imaging and gradient echo and spin echo (GRASE) myelin water imaging. J Magn Reson Imaging 2019;49:1091-1098
17. Jeong IH, Choi JY, Kim SH, et al. Comparison of myelin water fraction values in periventricular white matter lesions between multiple sclerosis and neuromyelitis optica spectrum disorder. Mult Scler 2016;22:1616-1620
18. Pauly J, Le Roux P, Nishimura D, Macovski A. Parameter relations for the Shinnar-Le Roux selective excitation pulse design algorithm [NMR imaging]. IEEE Trans Med Imaging 1991;10:53-65
19. Jeong IH, Choi JY, Kim SH, et al. Normal-appearing white matter demyelination in neuromyelitis optica spectrum disorder. Eur J Neurol 2017;24:1652-1658
20. Borich MR, Mackay AL, Vavasour IM, Rauscher A, Boyd LA. Evaluation of white matter myelin water fraction in chronic stroke. Neuroimage Clin 2013;2:569-580
21. Dvorak AV, Ljungberg E, Vavasour IM, et al. Rapid myelin water imaging for the assessment of cervical spinal cord myelin damage. Neuroimage Clin 2019;23:101896
22. Baumeister TR, Kim JL, Zhu M, McKeown MJ. White matter myelin profiles linked to clinical subtypes of Parkinson’s disease. J Magn Reson Imaging 2019;50:164-174
23. Lakhani B, Hayward KS, Boyd LA. Hemispheric asymmetry in myelin after stroke is related to motor impairment and function. Neuroimage Clin 2017;14:344-353
24. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In Proc of the 30th Int Conf Mach Learn (ICML), Atlanta, GA, USA, 2013:3
25. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014
26. Wilson DR, Martinez TR. The need for small learning rates on large problems. In Proc 2001 Int Joint Conf Neural Netw (IJCNN), Washington DC, USA, 2001:115-119
27. Smith SL, Kindermans P-J, Ying C, Le QV. Don’t decay the learning rate, increase the batch size. arXiv preprint arXiv:1711.00489, 2017
28. Du YP, Chu R, Hwang D, et al. Fast multislice mapping of the myelin water fraction using multicompartiment analysis of T2* decay at 3T: a preliminary postmortem study. Magn Reson Med 2007;58:865-870
29. Meyers SM, Laule C, Vavasour IM, et al. Reproducibility of myelin water fraction analysis: a comparison of region of interest and voxel-based analysis methods. Magn Reson Imaging 2009;27:1096-1103
30. Smith SM. Fast robust automated brain extraction. Hum Brain Mapp 2002;17:143-155
31. Hennig J. Multiecho imaging sequences with low refocusing flip angles. J Magn Reson 1988;78:397-407
32. Wang J, Mao W, Qiu M, Smith MB, Constable RT. Factors influencing flip angle mapping in MRI: RF pulse shape, slice-select gradients, off-resonance excitation, and B0 inhomogeneities. Magn Reson Med 2006;56:463-468
33. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2009;22:1345-1359
34. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng 2009;21:1263-1284
35. Krawczyk B. Learning from imbalanced data: open challenges and future directions. Artificial Intelligence 2016;5:221-232
36. Lee LE, Ljungberg E, Shin D, et al. Inter-vendor reproducibility of myelin water imaging using a 3D gradient and spin echo sequence. Front Neurosci 2018;12:854