Dataset variability leverages white-matter lesion segmentation performance with convolutional neural network

Domen Ravnik, Tim Jerman, Franjo Pernuš, Boštjan Likar, and Žiga Špiclin

Faculty of Electrical Engineering, University of Ljubljana
Tržaška 25, SI1000, Ljubljana, Slovenia

ABSTRACT

Performance of a convolutional neural network (CNN) based white-matter lesion segmentation in magnetic resonance (MR) brain images was evaluated under various conditions involving different levels of image preprocessing and augmentation applied and different compositions of the training dataset. On images of sixty multiple sclerosis patients, half acquired on one and half on another scanner of different vendor, we first created a highly accurate multi-rater consensus based lesion segmentations, which were used in several experiments to evaluate the CNN segmentation result. First, the CNN was trained and tested without preprocessing the images and by using various combinations of preprocessing techniques, namely histogram-based intensity standardization, normalization by whitening, and train dataset augmentation by flipping the images across the midsagittal plane. Then, the CNN was trained and tested on images of the same, different or interleaved scanner datasets using a cross-validation approach. The results indicate that image preprocessing has little impact on performance in a same-scanner situation, while between-scanner performance benefits most from intensity standardization and normalization, but also further by incorporating heterogeneous multi-scanner datasets in the training phase. Under such conditions the between-scanner performance of the CNN approaches that of the ideal situation, when the CNN is trained and tested on the same scanner dataset.

Keywords: magnetic resonance, brain, lesion, segmentation, machine learning, inter-scanner variability, normalization, standardization

1. INTRODUCTION

Neuroimaging biomarkers of brain lesions based on quantification of magnetic resonance (MR) images represent important surrogates of clinical signs in a number of neurological and cerebrovascular diseases, and mental disorders. In multiple sclerosis (MS) patients, for instance, inflammatory activity in brain parenchyma and spinal cord is visible as hyperintense lesions in T2-weighted and fluid attenuated inversion recovery (FLAIR) MR modalities, which typically occur very early in the disease onset. For this reason, monitoring of disease development and treatment response[1] and outcome prediction[2] are increasingly relying on biomarkers like lesion volume and count. To obtain such biomarkers, accurate segmentation of lesions in the MR images is required.

Lesion segmentation from brain MR images is challenging because of complex brain anatomy and high variability of possible locations, shapes and appearances of lesions, but also due to the inter-site and inter-scanner variabilities of image quality. Since manual lesion segmentation is not practical for routine use and also prone to intra- and inter-rater variabilities[3] several automated methods were developed.[4]

In recent years automated medical image segmentation is almost exclusively in the domain of machine learning techniques, like convolutional neural networks (CNNs). Some CNN based methods were also applied for white-matter lesion segmentation in MR images of patients with MS. For instance, Brosch *et al.*[5] proposed a lesion segmentation approach based on deep 3D convolutional encoder networks with shortcut connections similar to an U-net,[6] where two pathways, one convolutional and one deconvolutional, were interconnected so as to act as a feature extraction and prediction at various scales. To incorporate both local intensity and contextual information Kamnitsas *et al.*[7] instead proposed a multi-scale 3D CNN, which simultaneously processed two differently sized 3D image patches regions around a voxel of interest. The obtained segmentations were further

Z.S.: E-mail: ziga.spiclin@fe.uni-lj.si

post processed to remove false positives using a 3D fully connected conditional random field. Ghafoorian *et al.*[8] used a multi-scale network similar to the one used by Kamnitsas *et al.* , but incorporated additional eight spatial features and showed that these improve the accuracy of lesion segmentation. Valverde *et al.*[9] used a cascade of two 3D patch-wise CNNs, where the first network was trained to segment the input images into background and candidate lesion voxels, while the second network was trained to prune the misclassified voxels of the first network. Conversely to the aforementioned CNN architectures for lesion segmentation, Havaei *et al.*[10] focused on increasing the robustness of segmentations to an incomplete or missing MR multi-modal dataset. A CNN was separately trained for each input modality and the resulting trained layers were combined into a joint model that was shown to successfully deal with missing input modalities.

Brosch *et al.*[5] and Valverde *et al.*[9] evaluated their corresponding network architectures on the publicly available MICCAI 2008 MS lesion database,[11] where they performed similarly or better compared to more traditional segmentation methods, thus, showing high suitability of CNNs for the task of MR lesion segmentation.

While the use of CNN based methods demonstrated good performance when applied to *homogeneous* datasets (same site, same scanner), the large heterogeneity of the visual appearance and location of lesions and use across different site and different scanner vendors may substantially deteriorate their performance. Furthermore, their performance is heavily determined by the quality and size of datasets used in training of the model parameters, and by ingenuity and craftsmanship in model tuning. Herein, we focus on the use of CNNs in the context of white-matter lesion segmentation from brain MR images of 60 MS patients and aim to objectively evaluate i) the impact of image preprocessing like intensity standardization and normalization, and train dataset augmentation on the segmentation performance; and ii) the impact of training database construction on the segmentation performance across different MR scanners.

2. MATERIALS AND METHODS

2.1 Image Acquisition

Brain MR scans were acquired for 60 MS patients (47 females/13 males) in the age-span from 25 to 64 years. The MS phenotype distribution of the patients were the following: 54 relapsing-remitting, 2 secondary progressive, 1 primary progressive, 2 clinically isolated syndrome, and 1 unspecified. A total of 60 MR brain scans were acquired, 30 on a 3T Siemens Magnetom Trio MR system and the other 30 on a 3T Philips Achieva scanner, both at the University Medical Center Ljubljana (UMCL). Each scan included a T1-weighted and fluid-attenuated inversion recovery (FLAIR) images. Acquisition parameters are reported in Table 1.

All 60 subjects have given written informed consent at the time of enrollment for imaging and the UMCL approved the use of MRI data for this study. The authors confirm that the data were anonymized prior to analysis.

Table 1: Acquisition parameters of the 3T Siemens Magnetom Trio and 3T Philips Achieva scanners.

	T1-weighted		FLAIR	
Parameters	Siemens	Philips	Siemens	Philips
Sampling [pix]	$408 \times 512 \times 36$	$352 \times 165 \times 352$	$192 \times 512 \times 512$	$240 \times 321 \times 240$
Spacing [mm]	$0.42 \times 0.42 \times 3.30$	$0.66 \times 1.00 \times 0.66$	$0.47 \times 0.47 \times 0.80$	$0.97 \times 0.64 \times 0.97$
Echo time [ms]	20	4.3	392	276
Repetition time [ms]	2000	9.2	5000	4800
Inversion time [ms]	800	-	1800	1650
Flip angle [°]	120	8	120	90

2.2 Image Preprocessing

Prior to performing lesion segmentation, each subject's T1-weighted and FLAIR images were preprocessed. First, images were oriented to RAI patient coordinate space by permuting the image axes accordingly. Next,

both the T1-weighted and FLAIR images were denoised using curvature anisotropic diffusion filtering and then the denoised T1-weighted image was resampled to isotropic resolution $1.00 \times 1.00 \times 1.00$ using cubic interpolation. Brain region was extracted from the T1w image using Robex,[12] followed by mutual-information based registration of the FLAIR onto the T1-weighted image using affine transformation.[13] The FLAIR image was then affine-transformed and resampled into the space of T1-weighted image space using cubic interpolation. Finally, intensity inhomogeneity correction[14] was performed on each of the brain masked T1w and FLAIR images. The voxels lying within the brain mask were considered in lesion segmentation.

For the purpose of CNN model training, and for evaluating the impact on the CNN lesion segmentation performance, the preprocessed T1-weighted and FLAIR images were further enhanced using various combinations of three following preprocessing techniques: i) intensity normalization[15] based on subtracting the mean image intensity value and dividing by the standard deviation, ii) an intensity standardization by matching intensity histogram quartiles[16] of training and test images, iii) augmentation of training images by sagittal mirroring of the T1-weighted and FLAIR images, and reference lesion segmentations (cf. Section 2.3).

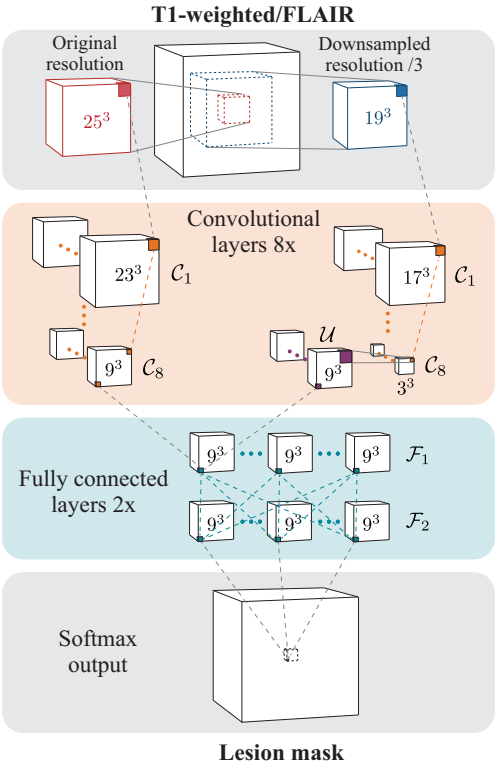


Figure 1: CNN model for lesion segmentation based on the DeepMedic framework.[7]

2.3 Lesion segmentation

The CNN model for lesion segmentation as shown on Fig. 1 was based on the DeepMedic framework,[7] which consists of two branches that are mutually independent until the final fully connected layer. First branch processes the input image patches of size $25 \times 25 \times 25$, while the second branch processes larger input image patches of $57 \times 57 \times 57$, but downsampled to $19 \times 19 \times 19$. Each branch consists of eight convolutional layers using $3 \times 3 \times 3$ kernels. Before the last layer, resolution of the second branch output is matched to the output of the first branch. Residual connections between layers were not used, while a default number of feature maps (i.e. [30, 30, 40, 40, 40, 40, 50, 50]) were used in each of the eight convolutional layers. The DeepMedic was based on Theano backend that utilized an NVidia GTX 970 GPU for computational speedup during training and test phases. The default optimizer settings were used in the training phase. The only exception was the number of epochs that was set to 5 for computational purposes, since several CNN models were trained and

the model parameters generally converged after 5 epochs as can be verified in Fig. 2. Image patches from both T1-weighted and FLAIR MR sequences were input into the CNN model, which was trained and validated with respect to the reference white-matter lesion segmentations.

The prediction obtained by the CNN model were thresholded at 0.5 to obtain candidate lesion segmentations. These segmentations were postprocessed by removing all candidate lesions outside the brain mask and those with volume equal or smaller than 5 mm^3 (empirically determined).

For evaluation purposes reference lesion segmentation were created on all 60 image datasets using a manual delineation protocol and tools as described in Lesjak *et al.*[3] In short, the reference lesion segmentation was based on a pair of T1-weighted and FLAIR sequences, on which three expert raters individually segmented the white-matter lesions, using in-house developed semi-automated lesion contouring tools. The raters then revised the segmentations in several joint sessions to reach a consensus segmentation, which was later used as a reference for evaluating performance of automated methods.

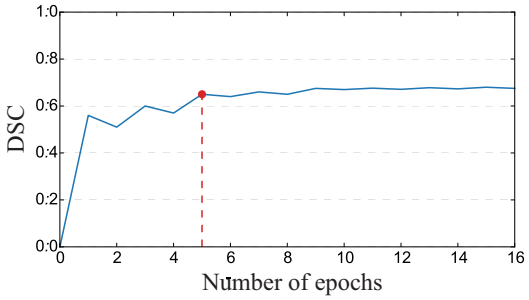


Figure 2: Validation of CNN model during training with respect to the number of epochs.

3. EXPERIMENTS

3.1 Setup

The performance of the CNN based lesion segmentation was evaluated in three experiments. In first, the Siemens datasets were used to train and test the CNN lesion segmentation without any preprocessing and by various combinations of the three preprocessing techniques, namely the histogram based intensity standardization, train dataset augmentation by sagittally flipping images and reference masks and intensity normalization. The purpose was to identify best combination of preprocessing steps. In second, the datasets acquired on either Siemens or Philips, or both scanner were used for training, while testing was done on the remaining images. In this way we could observe the impact of scanner and other image characteristics on the CNN lesion segmentation performance. All experiments were performed in a two-fold cross-validation manner, using half of the datasets for training and remaining for testing, and repeating the process by reversing the use of the datasets. For instance, to evaluate intra-scanner lesion segmentation performance, we used 15 datasets for training and the other 15 from the same scanner for testing and vice versa. In order to evaluate inter-scanner lesion segmentation performance, we used 30 datasets from one scanner for training and 30 datasets from the other scanner for testing. Finally, training datasets were also composed from the two scanners, i.e. 15 from each scanner for training and the remaining from both scanner for testing and vice versa.

The obtained lesion segmentations were compared to the reference segmentations using five performance metrics: 1) Dice similarity coefficient (DSC), 2) true positive rate (TPR), 3) positive predictive value (PPV), 4) lesion-wise TPR (LTPR), and 5) lesion-wise false positive rate (LFPR).

3.2 Results

In the same-scanner train and test set situation, no particular combination of the preprocessing techniques applied was found clearly superior to others, also in contrast to using the images without any preprocessing, as seen from the results in Table 2 and Fig. 3. Solely augmenting the training datasets using sagittally flipped images yielded worst segmentation performance, despite increased training dataset. In general, best results were

Table 2: Segmentation performance for same-scanner (Siemens) train and test images with different combinations of preprocessing and train set augmentation: quartile-based intensity histogram standardization (H), training set augmentation by sagittally flipped images (F) and whitening-like intensity normalization (N). Best values are in **bold**.

H	F	N	DSC / %	TPR / %	PPV / %	LTPR / %	LFPR / %
			63.15 ± 19.88	65.69 ± 25.18	64.95 ± 13.26	55.10 ± 11.15	51.83 ± 21.40
✓			62.66 ± 20.96	64.90 ± 25.23	65.74 ± 16.17	55.56 ± 11.57	50.49 ± 19.68
	✓		61.95 ± 24.41	63.99 ± 28.84	66.92 ± 12.84	53.00 ± 11.07	51.41 ± 22.59
✓	✓		**64.89 ± 19.35**	**66.74 ± 23.11**	65.79 ± 16.43	54.86 ± 12.84	47.46 ± 18.05
		✓	62.44 ± 19.95	63.17 ± 26.02	66.21 ± 14.39	53.22 ± 10.27	50.95 ± 21.10
✓		✓	62.44 ± 19.95	63.17 ± 26.02	66.21 ± 14.39	53.22 ± 10.27	50.95 ± 21.10
	✓	✓	63.08 ± 19.49	61.92 ± 23.55	**70.56 ± 15.78**	**55.91 ± 10.39**	50.43 ± 21.30
✓	✓	✓	64.72 ± 18.97	63.65 ± 22.33	68.82 ± 16.48	55.59 ± 10.97	**46.99 ± 18.47**

Table 3: Segmentation performance for mixed-scanner training set (Siemens+Philips) and per scanner test set, shown with respect to (w.r.t.) different combinations of applied train data preprocessing and augmentation (cf. Table 2).

Test	H	F	N	DSC / %	TPR / %	PPV / %	LTPR / %	LFPR / %
Siemens				54.44 ± 27.38	57.00 ± 31.92	64.24 ± 15.40	47.83 ± 11.62	64.42 ± 19.59
		✓	✓	62.36 ± 20.02	62.68 ± 25.45	67.90 ± 15.04	51.00 ± 9.77	52.42 ± 20.59
	✓	✓	✓	61.84 ± 20.43	61.71 ± 24.59	66.10 ± 16.05	51.41 ± 11.61	55.92 ± 17.44
Philips				73.29 ± 10.76	**78.65 ± 15.27**	72.88 ± 15.44	59.20 ± 11.67	51.32 ± 17.17
		✓	✓	**79.20 ± 8.26**	78.07 ± 10.55	83.01 ± 12.83	66.73 ± 10.40	**36.92 ± 12.89**
	✓	✓	✓	78.48 ± 8.82	73.14 ± 13.13	**86.62 ± 6.10**	**71.56 ± 10.45**	48.18 ± 14.82

obtained using combinations of augmentation and intensity normalization or standardization, or both. The difference between performance metrics among these combinations, however, were not statistically significant (Wilcoxon signed rank, $p < 0.05$) compared to other tested combinations.

When using mixed-scanner training set, the use of preprocessing turned out as a must, where the combination of augmentation and intensity normalization proved best, followed by the combination of all three considered preprocessing techniques (Table 3). In the latter two cases, the differences compared to the results obtained without using any preprocessing were statistically significant (Wilcoxon signed rank, $p < 0.05$).

The choice of combination of training and test datasets, e.g. same or different scanner, or mixed, did have a substantial impact on the lesion segmentation performance, as can be seen from Fig. 4 and Table 4. In general, best results were obtained in situations when the training and test datasets were from the same scanner and the worst results in situations when they were different. In these two scenarios, the values of performance metrics obtained during testing were statistically significantly different (Wilcoxon signed rank, $p < 0.05$). Constructing a mixed training dataset by employing MR datasets across the two scanners yielded lesion segmentation performance comparable to using solely the dataset from the same scanner (Wilcoxon signed rank, $p > 0.05$).

There is, however, an important difference between the lesion segmentation performance obtained on dataset

Table 4: Segmentation performance w.r.t. different combinations (same, different, mixed) of train and test data.

Test	Train	DSC / %	TPR / %	PPV / %	LTPR / %	LFPR / %
Siemens	Siemens	63.08 ± 19.49	61.92 ± 23.55	70.56 ± 15.78	55.91 ± 10.39	50.43 ± 21.30
	Philips	46.25 ± 24.92	51.48 ± 32.83	55.15 ± 12.44	46.32 ± 10.05	77.71 ± 16.41
	Siemens+Philips	62.36 ± 20.02	62.68 ± 25.45	67.90 ± 15.04	51.00 ± 9.77	52.42 ± 20.59
Philips	Philips	**81.49 ± 6.60**	**80.15 ± 9.86**	**84.55 ± 9.54**	68.61 ± 11.07	41.38 ± 13.11
	Siemens	73.00 ± 12.25	71.59 ± 14.36	78.76 ± 14.84	**69.95 ± 10.67**	38.70 ± 11.24
	Siemens+Philips	79.20 ± 8.26	78.07 ± 10.55	83.01 ± 12.83	66.73 ± 10.40	**36.92 ± 12.89**

from the two scanners, e.g. the best DSC on the Siemens dataset was about 62%, while on Philips it was about 79%. This difference may be attributed to a lower resolution T1-weighted image on the Siemens versus the Philips dataset ($0.42 \times 0.42 \times 3.00$ mm vs. $0.67 \times 1.00 \times 0.67$ mm) and a larger variety and larger number of smaller white-matter lesions on patients from the Siemens datasets. Example segmentations on cases from Siemens and Philips datasets are shown in Figure 5. The segmentations in the Siemens case contain a much higher number of false positives (Figs. 4c and 5), mainly on FLAIR hyperintense regions around the ventricles and in the medial temporal lobe; the latter is an artifact characteristic of Siemens scanners. These hyperintensities are less pronounced on the FLAIR images of the Philips scanner. Hence, this indicates that scanner characteristics, settings and/or artifacts can importantly impact segmentation performance.

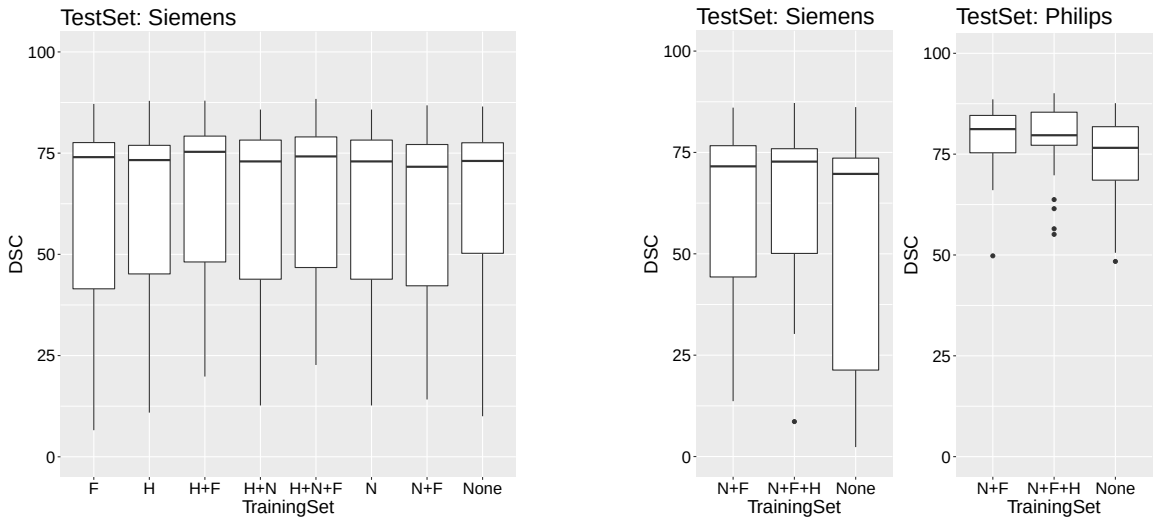


Figure 3: Segmentation performance for (a) same-scanner (Siemens) training and test set and (b) mixed-scanner (Siemens+Philips) training set and per scanner test set. Shown w.r.t. different combinations of applied train data preprocessing and augmentation (cf. Table 2).

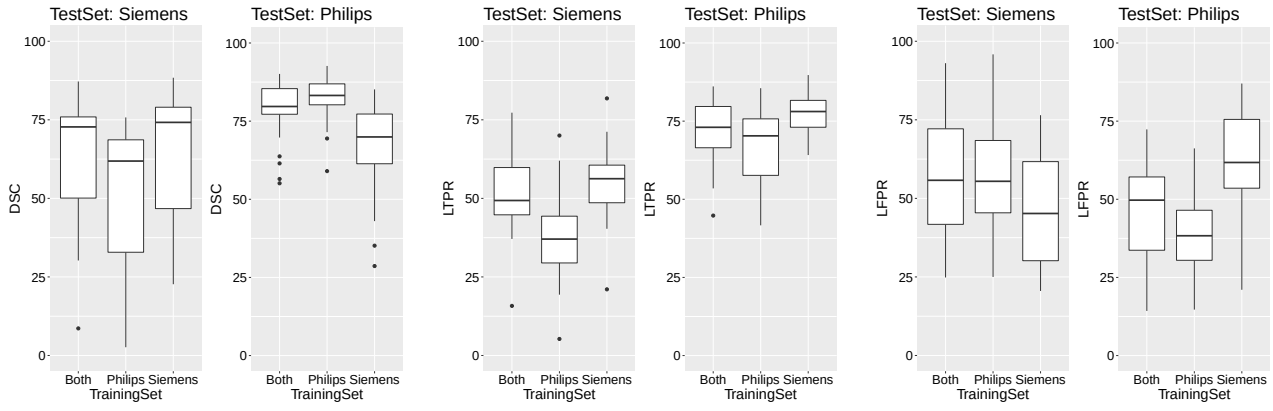
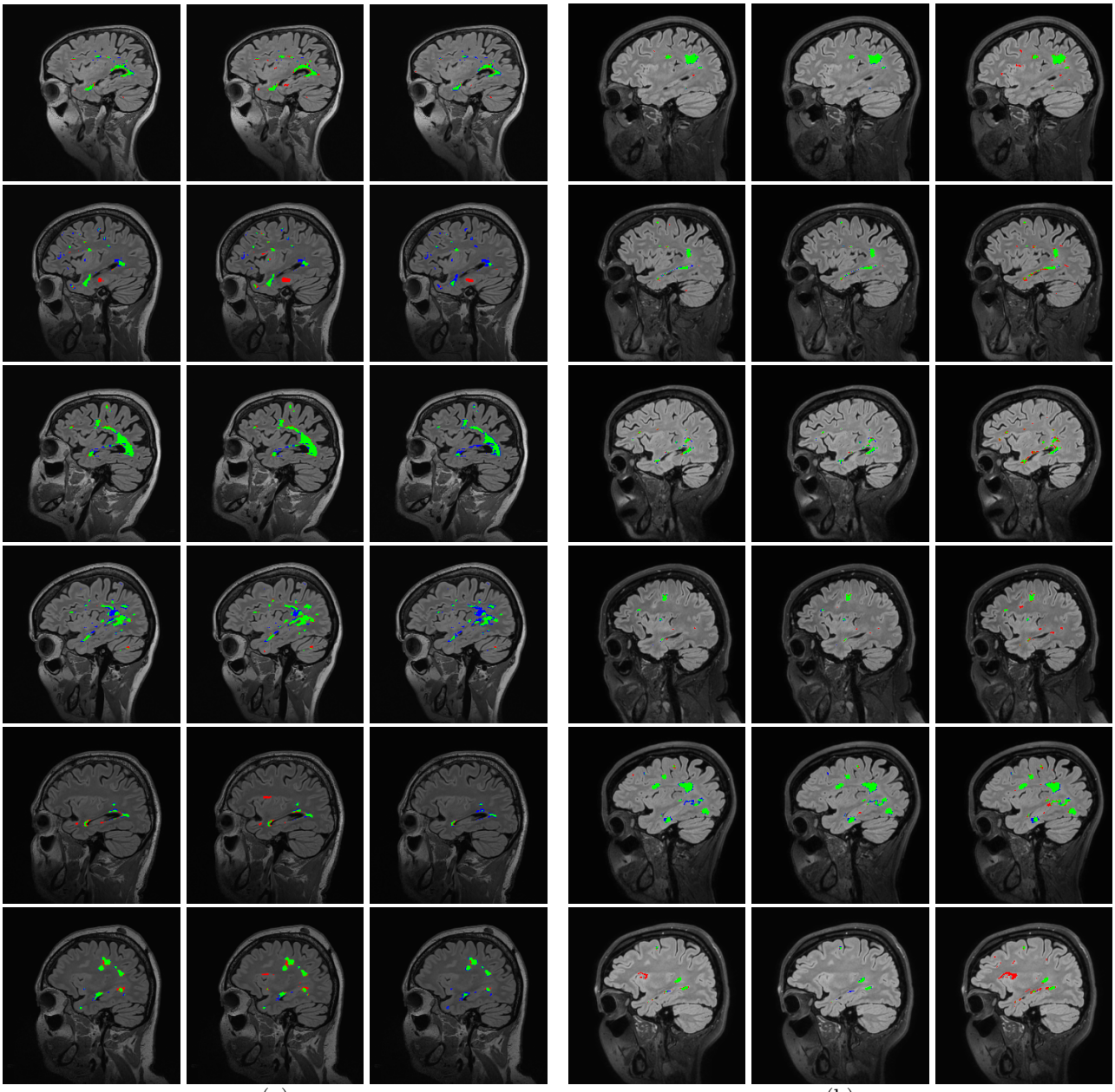


Figure 4: Segmentation performance: (a) DSC (b) LTPR, and (c) LFPR for mixed-scanner (Siemens+Philips) training set and per scanner test set. All three preprocessing techniques were applied (cf. Section 2.2).

4. CONCLUSION

Performance of a CNN based white-matter lesion segmentation in MR brain images was evaluated under various conditions involving different levels of image preprocessing and augmentation applied and different compositions



(a) (b)

Figure 5: Segmented FLAIR images in the sagittal cross-sections for six patients from the (a) Siemens and (b) Philips datasets. The three columns *from left to right* show segmentations w.r.t. training dataset composed of Siemens+Philips, Philips, and Siemens. True positives, false positives and false negatives are shown in *green, red and blue* colors, respectively.

of the training dataset. On images of sixty multiple sclerosis patients, half acquired on one and half on another scanner of different vendor, the results indicate that image dataset variability increased by incorporating heterogeneous multi-scanner datasets in the training phase leverages the performance of white-matter lesion segmentation with CNNs. Furthermore, it renders the CNN model applicable across datasets from different scanners with performance level comparable to the ideal situation, when the CNN is trained and tested on the same dataset.

Image intensity standardization and normalization and dataset augmentation did not improve the performance of CNNs in the context of same-scanner lesion segmentation performance. On the contrary, performance

was sometimes even worse as in case of using no preprocessing. Next, inter-scanner variability had an adverse effect on performance when datasets from different scanners were used in training and testing phases. Using mixed-scanner datasets for training has managed to overcome this issue. Hence, incorporating as much variability in the training datasets by using heterogeneous multi-scanner datasets in the training phase seems to enable the CNN to extract features more resilient to within- and across-scanner intensity variability.

5. ACKNOWLEDGMENTS

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0232, research grant funding Nos. J7-6781, J2-7211, J2-7118, and J2-8173).

REFERENCES

[1] Stangel, M., Penner, I. K., Kallmann, B. A., Lukas, C., and Kieseier, B. C., "Towards the implementation of 'no evidence of disease activity' in multiple sclerosis treatment: the multiple sclerosis decision model," *Therapeutic Advances in Neurological Disorders* **8**(1), 3–13 (2015).

[2] Uher, T., Vaneckova, M., Sobisek, L., Tyblova, M., Seidl, Z., Krasensky, J., Ramasamy, D., Zivadinov, R., Havrdova, E., Kalincik, T., and Horakova, D., "Combining clinical and magnetic resonance imaging markers enhances prediction of 12-year disability in multiple sclerosis," *Multiple Sclerosis* **23**, 51–61 (Jan. 2017).

[3] Lesjak, Z., Galimzianova, A., Koren, A., Lukin, M., Pernuš, F., Likar, B., and Špiclin, Z., "A Novel Public MR Image Dataset of Multiple Sclerosis Patients With Lesion Segmentations Based on Multi-rater Consensus," *Neuroinformatics* **In Press** (2017).

[4] García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L., "Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging," *Med. Image Anal.* **17**(1), 1–18 (2013).

[5] Brosch, T., Tang, L. Y. W., Yoo, Y., Li, D. K. B., Traboulsee, A., and Tam, R., "Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *IEEE Transactions on Medical Imaging* **35**(5), 1229–1239 (2016).

[6] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*], Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., eds., 234–241, Springer International Publishing, Cham (2015).

[7] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., and Glocker, B., "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical Image Analysis* **36**, 61 – 78 (2017).

[8] Ghafoorian, M., Karssemeijer, N., Heskes, T., Van Uden, I., I. Sanchez, C., Litjens, G., de Leeuw, F.-E., Ginneken, B., Marchiori, E., and Platel, B., "Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities," *Scientific Reports* **7**(5110), 2045–2322 (2017).

[9] Valverde, S., Cabezas, M., Roura, E., Gonzalez-Villa, S., Pareto, D., Vilanova, J. C., Ramio-Torrenta, L., Rovira, A., Oliver, A., and Llado, X., "Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach," *NeuroImage* **155**, 159–168 (2017).

[10] Havaei, M., Guizard, N., Chapados, N., and Bengio, Y., "Hemis: Hetero-modal image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*], Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G., and Wells, W., eds., 469–477, Springer International Publishing, Cham (2016).

[11] Styner, M., Lee, J., Chin, B., Chin, M. S., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., and Warfield, S., "3d Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation," *MIDAS Journal* **17**, 463–468 (2008).

[12] Iglesias, J., Liu, C.-Y., Thompson, P., and Tu, Z., "Robust Brain Extraction Across Datasets and Comparison With Publicly Available Methods," *IEEE Transactions on Medical Imaging* **30**(9), 1617 –1634 (2011).

[13] Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. P. W., "elastix: A Toolbox for Intensity-Based Medical Image Registration," *IEEE Transactions on Medical Imaging* **29**(1), 196–205 (2010).

[14] Tustison, N., Avants, B., Cook, P., Zheng, Y., Egan, A., Yushkevich, P., and Gee, J., "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imaging* **29**(6), 1310–1320 (2010).

[15] Shinohara, R. T., Sweeney, E. M., Goldsmith, J., Shiee, N., Mateen, F. J., Calabresi, P. A., Jarso, S., Pham, D. L., Reich, D. S., Crainiceanu, C. M., Australian Imaging Biomarkers Lifestyle Flagship Study of Ageing, and Alzheimer's Disease Neuroimaging Initiative, "Statistical normalization techniques for magnetic resonance imaging," *NeuroImage Clinical* **6**, 9–19 (2014).

[16] Zhuge, Y. and Udupa, J. K., "Intensity Standardization Simplifies Brain MR Image Segmentation," *Computer vision and image understanding* **113**(10), 1095–1103 (2009).