Published in final edited form as:

Magn Reson Imaging. 2020 January ; 65: 8–14. doi:10.1016/j.mri.2019.10.003.

Are multi-contrast magnetic resonance images necessary for segmenting multiple sclerosis brains? A large cohort study based on deep learning

Ponnada A. Narayana, PhD^{a,*}, Ivan Coronado, MS^a, Sheeba J. Sujit, PhD^a, Xiaojun Sun, MS^a, Jerry S. Wolinsky, MD^b, Refaat E. Gabr, PhD^a

^aDepartment of Diagnostic and Interventional Imaging, McGovern Medical School, University of Texas Health Science Center, Houston, TX 77030, United States of America

^bDepartment of Neurology, McGovern Medical School, University of Texas Health Science Center, Houston, TX 77030, United States of America

Abstract

Background: Magnetic resonance images with multiple contrasts or sequences are commonly used for segmenting brain tissues, including lesions, in multiple sclerosis (MS). However, acquisition of images with multiple contrasts increases the scan time and complexity of the analysis, possibly introducing factors that could compromise segmentation quality.

Objective: To investigate the effect of various combinations of multi-contrast images as input on the segmented volumes of gray (GM) and white matter (WM), cerebrospinal fluid (CSF), and lesions using a deep neural network.

Methods: U-net, a fully convolutional neural network was used to automatically segment GM, WM, CSF, and lesions in 1000 MS patients. The input to the network consisted of 15 combinations of FLAIR, T1-, T2-, and proton density-weighted images. The Dice similarity coefficient (DSC) was evaluated to assess the segmentation performance. For lesions, true positive rate (TPR) and false positive rate (FPR) were also evaluated. In addition, the effect of lesion size on lesion segmentation was investigated.

Results: Highest DSC was observed for all the tissue volumes, including lesions, when the input was combination of all four image contrasts. All other input combinations that included FLAIR also provided high DSC for all tissue classes. However, the quality of lesion segmentation showed strong dependence on the input images. The DSC and TPR values for inputs with the four contrast combination and FLAIR alone were very similar, but FLAIR showed a moderately higher FPR for lesion size < 100 μ l. For lesions smaller than 20 μ l all image combinations resulted in poor performance. The segmentation quality improved with lesion size.

Conclusions: Best performance for segmented tissue volumes was obtained with all four image contrasts as the input, and comparable performance was attainable with FLAIR only as the input, albeit with a moderate increase in FPR for small lesions. This implies that acquisition of only

*Corresponding author. Ponnada.a.narayana@uth.tmc.edu (P.A. Narayana).

FLAIR images provides satisfactory tissue segmentation. Lesion segmentation was poor for very small lesions and improved rapidly with lesion size.

Keywords

Deep learning; Magnetic resonance imaging; U-net; Segmentation; Multiple sclerosis; Dice similarity coefficient; False positive rate; False negative rate

1. Introduction

Multiple sclerosis (MS) is characterized by demyelinating lesions in the central nervous system (CNS). In addition, atrophy of gray matter (GM) and white matter (WM) is known to occur early on and continues throughout the disease course [1,2]. Magnetic resonance imaging (MRI) is the most common imaging modality for visualizing and quantifying lesion volumes and atrophy in MS [3]. Demyelinating lesions appear hyperintense on T2-weighted (T2) MRI. Both atrophy and total lesion volume may help in patient management and conducting efficient and effective clinical trials [2]. Routine clinical scans in MS include T1- (T1), T2- (T2), and proton density (PD)-weighted, fluid attenuation by inversion recovery (FLAIR), and post-contrast T1 images for evaluating the disease state.

Image segmentation or tissue classification is necessary for estimating lesion volume and atrophy. Manual, semi-automatic, and automatic segmentation techniques are used for segmenting brain MRI. Automatic techniques are preferable for bias free segmentation and analyzing large number of images [4,5]. Magnetic resonance images with multiple contrasts (or multi-contrast images) are commonly used for segmentation of brain images in MS [5–7]. However, acquisition of multi-contrast images prolongs the scan time and increases the amount of data processing. In addition, processing of multi-contrast MRI requires intra-patient image registration that may be affected in the presence of motion-induced artifacts in any of the image sets. Therefore, it is helpful to identify the minimum number of multi-contrast image volumes that still provides acceptable segmentation results.

A variety of automatic segmentation techniques have been proposed for segmenting various CNS tissues, including T2 lesions in MS [5]. However, there is a general consensus that these techniques lack generalizability for segmenting MRI data generated in a multi-center setting using different scanner platforms operating at different field strengths.

Deep learning (DL) with multiple layers of neural networks is increasingly used for automatic image segmentation [8,9]. DL automatically identifies the image features through hierarchical learning [10]. Fully convolutional neural networks (FCNNs) are commonly used for image segmentation. FCNNs consist of contracting (encoding) and expanding (decoding) paths, with interconnections between them. DL is also shown to handle heterogeneous data and is relatively insensitive to image noise and other artifacts [11–13] and is particularly well suited for analyzing multi-center data acquired with different quality control programs in place and different scanner platforms.

Large amount of annotated data is required for training DL models. Large annotated medical image data, for variety of reasons, is hard to come by. As part of phase 3 clinical trial,

CombiRx, large number of magnetic resonance image volumes were acquired on relapsing remitting MS (RRMS) and were annotated. These images were acquired on different scanners operating at 3 T and 1.5 T, the most common field strengths used clinically [14].

In this study we assessed the effect of images acquired with different contrasts and their combinations as the network input on segmentation accuracy in MS using ~1000 image volumes. The performance was evaluated based on the Dice similarity coefficient (DSC) for all tissue volumes, including lesions. For WM lesions, we also calculated true positive (TPR) and false positive (FPR) rates for different combinations of input images. Finally, we evaluated the effect of lesion size on lesion segmentation.

2. Methods

2.1. Ethics statement

Prospective analysis of these anonymized images acquired as a part of multi-center clinical trial was approved by our Institutional Review Board. Individual sites where the patients were recruited and scanned obtained IRB approvals and written consent was obtained from each patient. These studies are completely HIPAA compliant.

2.2. Image dataset

The brain images used in this study were acquired as part of a multi-center, double-blinded, randomized clinical trial, CombiRx [14] (Clinical trial identifier:). 1008 patients were recruited consecutively between 2005 and 2009. The main objective of this trial was to evaluate the relative efficacies of interferon β -1a (IFN) 30 μ g weekly and glatiramer acetate (GA) 20 mg daily or combination of these two drugs. Patients were recruited from 68 academic and stand-alone private clinics. MRI data were acquired on multiple platforms at 1.5 T and 3 T field strengths (Philips or GE or Siemens), with nearly 85% of the MRIs acquired at 1.5 T. 3D T1 images (0.94 mm $\times$ 0.94 mm $\times$ 1.5 mm voxel dimensions), 2D FLAIR and 2D dual echo turbo spin echo or fast spin echo (FSE) images (0.94 mm $\times$ 0.94 mm $\times$ 3 mm voxel dimensions) were acquired on this cohort. In addition, pre- and post-contrast T1 images with identical geometry as FLAIR and FSE images were acquired. As a part of this trial all acquired images were evaluated for quality [15].

2.3. Data preprocessing

Dual echo FSE (providing PD and T2 images), FLAIR, and pre-contrast T1 images were used for segmentation. All images were pre-processed and segmented using the magnetic resonance imaging automatic processing (MRIAP) pipeline [16,17]. The MRIAP pipeline includes several pre-processing and segmentation steps. These are described below briefly.

2.3.1. Pre-processing—Rigid-body registration was performed for intra-subject registration to align both FLAIR and T1 images to their corresponding T2 images. All non-neural tissues were removed (skull stripping), bias field correction was performed to improve the image homogeneity, image intensities were normalized [18], and images were filtered using anisotropic diffusion filter for noise reduction [19].

2.3.2. Segmentation—Following the preprocessing described above, initially lesions, CSF, and brain parenchyma were segmented using the FLAIR and T2-weighted images based on the Parzen window classifier [20]. False positives were minimized using the ratio of PD and T2 images (with same threshold for all subjects). Two-dimensional Fuzzy connectedness was used for lesion delineation. Then the brain parenchyma was further classified into GM and WM using the expectation maximization (EM) algorithm along with the hidden Markov random field model [21]. Original PD and T2 images (without bias field correction) were used for this purpose as bias field was corrected iteratively in the EM algorithm itself. These GM and WM segmented results were merged with the earlier CSF and lesion segmented results. Lesion boundaries tend to segment as GM, forming GM islands within WM. These false negatives were minimized by comparing lesions on the FLAIR images. Small errors in the image registration and partial volume effects tend to produce false lesion classifications around the brain edges. Careful observation on a large image database obtained from multiple patients indicated that most of these false positives occurred within 2 to 3 pixels from the brain surface. Using this criterion and employing the morphological erosion operation with 2D kernel of size 3×3 , all false lesions at the surface were eliminated.

The MRIAP segmentation results were further validated and edited, if necessary, by two neuroimaging experts with 25+ and 15+ years of experience in MRI of MS and other neurological disorders. The experts used FLAIR, T2, and PD images to validate the results. We developed software to display all these images simultaneously along with the necessary editing tools. This validated segmentation was used as the ground truth for training the network. In this study we focused only on the baseline CombiRx MRI data.

2.4. Network description

For this study we used the 2-dimensional U-net [22] which was shown to be successful in image segmentation [8,9]. The contracting and expanding paths in the U-net consist of blocks of convolutional layers (Fig. 1). These blocks are connected between the paths allowing the retention of high level features which are lost in the contracting path. Blocks in the encoding path are followed by a max pooling layer, while blocks in decoding path are followed by an up-sampling layer. Up-sampling layers are used to match learned feature dimensions with corresponding block's resolution before max pooling.

In the present study max pooling layers were used with a 2×2 kernel, effectively reducing resolution by a factor of 2 along each axis. Convolutional layers in each of the five convolutional blocks were assigned a number of filters, starting with 64 filters for the first convolutional block and the number of filters was doubled for all succeeding convolutional blocks in the encoding path. A similar approach was used in the decoding path but the number of filters was halved in successive convolutional blocks. The last convolutional block had the same number of filters as the first convolutional block. All convolutional layers in the network were followed by rectified linear unit (ReLU) activation. Each convolution layer used 3×3 kernels. In the last layer of the network, softmax activation was used to output membership scores for each of the tissue classes.

2.5. Network training

Preprocessed PD, T2, T1, and FLAIR images and their various combinations served as input to the network. Analysis was performed on 1000 image sets acquired at baseline in the CombiRx trial. In machine learning, it is common to divide the data into three parts: 1) training, 2) validation, and 3) testing [23]. The validation set is mainly used for determining the over- or under-fitting by comparing the changes in the loss function as a function of the number of epochs. The test data is used to evaluate how well the network performs without changing the network parameters following training. In this study, data were randomly partitioned into training (60% of the scans), validation (20%), and test (20%) sets.

The total number of combinations of image volumes used for input was 15; all 4 sequences (1 set), all combinations of 3 sequences (4 sets), all combinations of 2 sequences (6 sets), 1 sequence each (4 sets). Same training and evaluation approaches were used for all these different inputs.

For each set, the network was trained for 500 epochs (1 epoch = 1 iteration through training data) using the Adam optimizer [24], with an initial learning rate of 0.001. Network weights were initialized using Xavier algorithm [25]. A balanced version of Dice score coefficient [26] was selected to help alleviate imbalance between different tissue classes to avoid trapping at local minima. Data augmentation was used during training, including vertical and horizontal flips, rotation, translation, and zooming. The expert validated MRIAP segmentation served as the ground truth.

Training was implemented on the Maverick2 cluster at Texas Advanced Computing Center (TACC) with four NVIDIA GTX 1080 Ti graphics processing unit (GPU) cards using Keras [27] and Tensorflow [28].

2.6. Performance measures

Dice similarity coefficient was used to assess the agreement between network output and the ground truth:

$$DCS_k = 2 \times TP_k / (FP_k + 2 \times TP_k + FN_k)$$

TP_k , FP_k , and FN_k represent the number of true positive, false positive, and false negative classification for tissue class k , respectively. Lesion-wise TPR and FPR were also determined. We also investigated the effect of lesion size on the performance measures for various inputs to the network. For this purpose lesions were arbitrarily divided into 7 categories based on their volumes: 0–19 μl ; 20–34 μl ; 35–69 μl ; 70–137 μl ; 138–276 μl ; 277–499 μl ; 500 μl .

3. Results

Fig. 2 shows, as an example, the acquired FLAIR Fig. 2(A) and the CNN segmented images with different combination of input images (C–J) for one slice in a MS patient in the test set. For comparison, the corresponding ground truth segmented image is also shown in this Fig.

2(B). Overall, good quality segmentation can be observed, at least visually, on this figure when the input set contained FLAIR. In general, inputs that did not contain FLAIR did not produce satisfactory lesion segmentation. The worst lesion segmentation results were obtained with T1 images only as the input. However, segmentation of GM, WM, and CSF was satisfactory for all the inputs.

The results of quantitative analysis of segmentation, as assessed by DSC, for different inputs are summarized in Table 1. This table also reports lesion TPR and FPR for all inputs. As can be seen from this table, highest DSC was observed for GM, WM, and CSF when the input included all four image contrasts. However, all the other input combinations also provided excellent GM, WM, and CSF segmentation results.

The choice of input showed a significant effect on the quality of lesion segmentation (Table 1). The most important observations with regard to lesions based on the quantitative analysis are: 1) highest DSC was observed when images with all four contrasts were used in the input and the performance was comparable when only FLAIR was used as the input and 2) exclusion of the FLAIR images from the input resulted in poor segmentation, and 3) lesion FPR appeared to show much stronger dependence on the input than DSC and TPR, perhaps implying that FPR is a sensitive performance measure.

The DSC, TPR, and FPR for different lesion sizes are summarized in Table 2. For visualization, Fig. 3 shows the DSC, TPR, and FPR for different lesion sizes for 4 different image combinations that produced the best lesion segmentation: all four contrasts, FLAIR +T2+PD, FLAIR +T2, and FLAIR. The major observations based on this analysis are: 1) all input combinations performed rather poorly as reflected by DSC and/or FPR, TPR, in segmenting the smallest lesions (0–19 μ l), 2) highest DSC and TPR and lowest FPR were observed for input that includes all the four images for all lesion sizes > 20 μ l and the three performance measured improved with increasing lesion size; the results are comparable with only FLAIR as the input for DSC and TPR, but with higher FPR, 3) performance was poor for all lesion sizes when the input did not include FLAIR images, and 4) FPR showed stronger dependence on the input for all lesion sizes than DSC and TPR.

4. Discussion

There is a general agreement that segmentation based on multi-contrast images is superior to that based on a single contrast [5,7]. We believe that this is the first study on a large multi-center data that systematically investigated the effect of images with different contrasts on the segmentation of lesions, WM, GM, and CSF. The input to the network included various combinations of T1, T2, PD, and FLAIR images. These images are most commonly acquired in routine clinical scans in MS.

Our quantitative analysis indicates that the performance when the input contained all the four contrasts is similar to that observed when only FLAIR served as the input, but with improved FPR when all sequences are included. As seen from Table 1, without inclusion of FLAIR, the DSC values for lesions varied from 0.34 to 0.66, with the worst performance observed when only the T1 images were included as the input. The inclusion of FLAIR

resulted in substantially higher DSC values. However, DSC value alone does not provide complete picture about the network performance since it is relatively insensitive to segmentation of smaller lesions. FPR appears to be a more sensitive measure for evaluating smaller lesion segmentation as seen in its improvement when PD, T1 and T2 images were included with FLAIR.

While a few studies investigated the quality of WM hyperintense lesion segmentations using FLAIR alone [29–31], few compared the segmentation performance using different input sequences in MS. Using 3D CNN model, Brosch et al. [32] investigated the effect of different sequences on MS lesion segmentation. Specifically, these authors evaluated three contrast combinations: T1 + FLAIR, T1 + T2 + PD, and T1 + T2 + FLAIR + PD on lesion segmentation. Their results, consistent with ours, showed that the combination of 4 sequences provided the best lesion segmentation. They reported DSC of 0.638, FPR of 0.361, and TPR of 0.625. In contrast, our studies investigated the effect of 15 combinations on segmentation of GM, WM, CSF, and lesions. In our study, the DSC, FPR, and TPR values over all lesions using this sequence combination were 0.9, 0.28, and 0.81 respectively. This is perhaps one of the highest reported DSC value for MS lesions. Similarly, the four combination input resulted in DSC values of 0.96, 0.95, and 0.98 for WM, GM, and CSF, respectively. Part of the reason for the high DSC values was the large training data and accurate annotated image data. Fartaria et al. [33] segmented MS lesions using different combinations of 3D FLAIR, 3D MPRAGE, 3D MP2RAGE, and 3D double inversion recovery (DIR) images. These authors reported that the highest DSC was observed using the combination of FLAIR and MPRAGE. However, they did not perform segmentation based on FLAIR images alone. Even with the combination of FLAIR and MPRAGE, the DSC was < 0.6. Simões et al. [31] also have shown that results based on 3D FLAIR alone are comparable with those of state-of-the-art methods. These authors reported average TPR and FPR values obtained on the MICCAI challenge dataset of 0.304–0.352 and 0.213–0.182, respectively. When we used only FLAIR as the input, we obtained TPR and FPR of 0.82 and 0.47, respectively. A reason for this difference in TPR and FPR between ours and their study could be due to the fact that they used 3D FLAIR while all our analyses are based on 2D images.

Our finding that FLAIR performed better than FLAIR+T2 may appear surprising. But a possible reason for this result is that T2 images do not provide any additional image features that are not contained in the FLAIR images.

As can be seen from Table 2, the segmentation quality is degraded with decreasing lesion size, perhaps because noise has a stronger effect on smaller lesions. For example, even when all the four combinations were included, we observed the segmentation quality for lesions < 20 μ l to be poor as indicated by high FPR of 0.71 and relatively low TPR of 0.68 even when the DSC is reasonable (0.67). In contrast, for lesions $\geq$ 500 μ l, the TPR/FPR/DSC values improved dramatically to 0.99/0.02/0.91. This trend of improved segmentation with increased lesion size is consistent with that reported by Brosch et al. [32].

5. Limitations

Given the technical improvements in both software and hardware, it is now possible to acquire 3D multi-contrast MRI. This reduces partial volume averaging and could yield superior segmentation results. However, we are limited by the CombiRx protocol. Even though CombiRx is a multi-center clinical trial, all the patients were scanned using identical MRI protocol. For our conclusions to be generalizable, it is necessary to repeat these studies with images acquired with different scan parameters.

In this study, the ground truth was generated largely automatically by MRIAP which was manually edited by 2 experts. It would have been ideal if expert segmentations was not biased by automatic priors. Unfortunately, given the large number of images (1000 patients $\times$ (44 FLAIR $\times$ +44 T2 $\times$ +44 PD images)), it is impractical to segment these images completely manually. Lack of true “ground truth” is a problem with all the medical images.

In the CombiRx trial, 85% of the patients were scanned at 1.5 T and 15% at 3 T. Because of this large imbalance between the number of scans between these two field strengths, we did not stratify the results based on the field strength.

6. Conclusions

Our studies demonstrate that the combination of images with 4 different contrasts (FLAIR, T1, T2, and PD) as the input to the network provided the best segmentation results as assessed by all three performance measures of DSC, TPR, and FPR. FLAIR alone as the input yielded very comparable DSC and TPR results to that obtained with all the four contrasts combined, suggesting that acquisition of only FLAIR may be adequate for image segmentation if the moderate increase in FPR is acceptable. While DSC values for GM, WM, and CSF are relatively insensitive to the sequence combination, lesion segmentation appears to strongly depend on the combination of contrasts. The segmentation quality improves with the lesion size. However, for lesions $< 20 \mu\text{l}$, none of the image combinations provided satisfactory segmentation results.

Acknowledgments

We thank Fred Lublin, MD, the PI on the CombiRx trial, for proving access to the CombiRx data. This work is supported in part by NINDS of the National Institutes of Health (grant 1R56NS105857) and Chair in Biomedical Engineering Endowment. The content of this manuscript is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

References

- [1]. Rocca MA, Battaglini M, Benedict RHB, De Stefano N, Geurts JGG, Henry RG, et al. Brain MRI atrophy quantification in MS: from methods to clinical application. *Neurology* 2017;88:403–13. [PubMed: 27986875]
- [2]. Sastre-Garriga J, Pareto D, Rovira À. Brain atrophy in multiple sclerosis: clinical relevance and technical aspects. *Neuroimaging Clin* 2017;27:289–300.
- [3]. Moccia M, de Stefano N, Barkhof F. Imaging outcome measures for progressive multiple sclerosis trials. *Mult Scler J* 2017;23:1614–26.

- [4]. García-Lorenzo D, Francis S, Narayanan S, Arnold DLCD. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Med Image Anal* 2013;17:1–18. [PubMed: 23084503]
- [5]. Danelakis A, Theoharis T, Verganelakis DA. Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. *Comput Med Imaging Graph* 2018;70:83–100. [PubMed: 30326367]
- [6]. Datta S, Narayana PA. A comprehensive approach to the segmentation of multi-channel three-dimensional MR brain images in multiple sclerosis. *NeuroImage Clin* 2013;2:184–96. 10.1016/j.nicl.2012.12.007. [PubMed: 24179773]
- [7]. Anbeek P, Vincken KL, Van Bochove GS, Van Osch MJP, van der Grond J. Probabilistic segmentation of brain tissue in MR imaging. *Neuroimage* 2005;27:795–804. [PubMed: 16019235]
- [8]. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging* 2017;30:449–59. [PubMed: 28577131]
- [9]. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017;19:221–48. [PubMed: 28301734]
- [10]. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep Learning*. 1 MIT press Cambridge; 2016.
- [11]. Moeskops P, de Bresser J, Kuijf HJ, Mendrik AM, Biessels GJ, Pluim JPW, et al. Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI. *NeuroImage Clin* 2018;17:251–62. [PubMed: 29159042]
- [12]. Rachmadi MF, Valdés-Hernández M del C, Agan MLF, Di Perri C, Komura T. Segmentation of white matter hyperintensities using convolutional neural networks with global spatial information in routine clinical brain MRI with none or mild vascular pathology. *Comput Med Imaging Graph* 2018. doi:10.1016/j.compmedimag.2018.02.002.
- [13]. Laukamp KR, Thiele F, Shakirin G, Zopfs D, Faymonville A, Timmer M, et al. Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. *Eur Radiol* 2018. 10.1007/s00330-018-5595-8.
- [14]. Lindsey JW, Scott TF, Lynch SG, Cofield SS, Nelson F, Conwit R, et al. The CombiRx trial of combined therapy with interferon and glatiramer acetate in relapsing remitting MS: design and baseline characteristics. *Mult Scler Relat Disord* 2012;1:81–6.
- [15]. Narayana PA, Govindarajan KA, Goel P, Datta S, Lincoln JA, Cofield SS, et al. Regional cortical thickness in relapsing remitting multiple sclerosis: a multi-center study. *NeuroImage Clin* 2013;2:120–31. 10.1016/j.nicl.2012.11.009.
- [16]. Sajja BR, Datta S, He R, Mehta M, Gupta RK, Wolinsky JS, et al. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Ann Biomed Eng* 2006;34:142–51. 10.1007/s10439-005-9009-0. [PubMed: 16525763]
- [17]. Datta S, Sajja BR, He R, Wolinsky JS, Gupta RK, Narayana PA. Segmentation and quantification of black holes in multiple sclerosis. *Neuroimage* 2006;29:467–74. 10.1016/j.neuroimage.2005.07.042. [PubMed: 16126416]
- [18]. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging* 2000;19:143–50. [PubMed: 10784285]
- [19]. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans Pattern Anal Mach Intell* 1990;12:629–39. 10.1109/34.56205.
- [20]. Duda RO, Hart PE, Stork DG. *Pattern Classification*. John Wiley & Sons; 2012.
- [21]. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 2001;20:45–57. [PubMed: 11293691]
- [22]. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci* 2015. 10.1007/978-3-319-24574-4_28. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics).
- [23]. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology* 2018;286:800–9. [PubMed: 29309734]

- [24]. Kingma DP, Ba J. Adam: A method for stochastic optimization. ArXiv Prepr ArXiv14126980 2014.
- [25]. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Receiv Febr* 2010;12(2010):249–56.
- [26]. Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Deep learn. Med. Image anal. Multimodal learn. Clin. Decis. Support.* Springer; 2017 p. 240–8.
- [27]. Chollet F, others. Keras: Deep learning library for theano and tensorflow. URL <https://KerasIo/K2015;7>.
- [28]. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: a system for large-scale machine learning. *OSDI* 2016;16:265–83.
- [29]. Gibson E, Gao F, Black SE, Lobaugh NJ. Automatic segmentation of white matter hyperintensities in the elderly using FLAIR images at 3T. *J Magn Reson Imaging* 2010;31:1311–22. [PubMed: 20512882]
- [30]. Khademi A, Venetsanopoulos A, Moody AR. Robust white matter lesion segmentation in FLAIR MRI. *IEEE Trans Biomed Eng* 2011;59:860–71. [PubMed: 22203699]
- [31]. Simões R, Mönninghoff C, Dlugaj M, Weimar C, Wanke I, van C van Walsum A-M, et al. Automatic segmentation of cerebral white matter hyperintensities using only 3D FLAIR images. *Magn Reson Imaging* 2013;31:1182–9. [PubMed: 23684961]
- [32]. Brosch T, Tang LYW, Yoo Y, Li DKB, Traboulsee A, Tam R. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Trans Med Imaging* 2016;35:1229–39. 10.1109/TMI.2016.2528821. [PubMed: 26886978]
- [33]. Fartaria MJ, Bonnier G, Roche A, Kober T, Meuli R, Rotzinger D, et al. Automated detection of white matter and cortical lesions in early stages of multiple sclerosis. *J Magn Reson Imaging* 2016;43:1445–54. 10.1002/jmri.25095. [PubMed: 26606758]

Fig. 1.

Architecture of the U-net used in this study. The numbers next to the layers represent the image dimensions while the numbers at the bottom represent the number of filters.

Fig. 2.

Segmentation results for different input combinations in one slice in the test set in a MS patient. A: FLAIR; B: ground truth; C: FLAIR +PD+T1+T2; D: FLAIR +PD+T2; E: T1+T2+PD; F: PD+T2; G: FLAIR; H: PD; I: T2; J: T1. WM, GM, CSF, and lesions are shown in white, gray, cyan, and pink, respectively.

Fig. 3.

Variation of DSC, TPR, and FPR with lesion size for 4 different network inputs: combination of all 4 contrasts, FLAIR+T2+PD, FLAIR+T2, and FLAIR. The different shades of gray in this figure indicate different lesion sizes, with the lightest and darkest representing the smallest and the largest lesion groups, respectively.

Table 1

Dice similarity coefficient (DSC) for GM, WM, CSF, and lesions for various combinations of input images. The lesion true positive rate (TPR) and false positive rate (FPR) for lesions are also shown in this table.

Input sequences				DSC				Lesion TPR	Lesion FPR
FLAIR	PD	T1	T2	GM	WM	CSF	Lesions		
X	X	X	X	0.96	0.96	0.98	0.90	0.81	0.28
X	X	X		0.92	0.90	0.94	0.78	0.82	0.48
X	X		X	0.95	0.95	0.98	0.86	0.76	0.5
	X	X	X	0.92	0.94	0.91	0.66	0.70	0.46
X		X	X	0.92	0.90	0.98	0.73	0.85	0.58
X	X			0.92	0.92	0.93	0.84	0.8	0.47
	X		X	0.90	0.92	0.88	0.66	0.77	0.57
	X	X		0.88	0.88	0.88	0.65	0.77	0.60
X			X	0.93	0.92	0.98	0.85	0.84	0.52
X		X		0.85	0.84	0.89	0.75	0.8	0.47
		X	X	0.89	0.91	0.90	0.64	0.73	0.52
X				0.85	0.86	0.89	0.91	0.82	0.47
	X			0.89	0.92	0.85	0.60	0.69	0.54
			X	0.80	0.78	0.82	0.64	0.71	0.54
		X		0.89	0.92	0.88	0.34	0.46	0.72

Table 2

True positive rate (TPR), false positive rate (FPR), and Dice similarity coefficient (DSC) for the seven lesion sizes for different combinations of input images.

Input Sequences		TPR/FPR/DSC for different lesion sizes (shown below in μ l)									
FLAIR	PD	T1	T2	0-19	20-34	35-69	70-137	138-276	277-499	> 500	All Lesions
X	X	X	X	0.68/0.71/0.67	0.71/0.24/0.69	0.82/0.14/0.75	0.89/0.08/0.81	0.95/0.06/0.87	0.96/0.03/0.89	0.99/0.02/0.91	0.81/0.38/0.90
X	X	X		0.68/0.86/0.55	0.73/0.44/0.57	0.83/0.24/0.67	0.90/0.13/0.72	0.96/0.06/0.78	0.97/0.05/0.78	1.00/0.02/0.80	0.82/0.48/0.78
X	X	X	X	0.66/0.83/0.57	0.64/0.32/0.58	0.74/0.17/0.64	0.80/0.12/0.68	0.90/0.05/0.76	0.92/0.06/0.77	0.98/0.04/0.84	0.76/0.50/0.86
	X	X	X	0.43/0.74/0.32	0.58/0.47/0.40	0.75/0.36/0.47	0.84/0.19/0.53	0.92/0.16/0.61	0.95/0.09/0.62	0.99/0.09/0.71	0.70/0.46/0.66
X		X	X	0.75/0.91/0.54	0.77/0.52/0.56	0.86/0.28/0.63	0.89/0.14/0.68	0.95/0.07/0.73	0.97/0.05/0.73	0.99/0.01/0.75	0.85/0.58/0.73
X	X			0.67/0.82/0.59	0.71/0.32/0.61	0.81/0.19/0.69	0.87/0.11/0.75	0.95/0.05/0.81	0.96/0.05/0.83	1.00/0.02/0.85	0.80/0.47/0.84
	X	X	X	0.54/0.85/0.37	0.68/0.59/0.44	0.81/0.43/0.50	0.86/0.26/0.52	0.95/0.18/0.61	0.97/0.12/0.63	1.00/0.06/0.69	0.77/0.57/0.66
	X			0.57/0.88/0.35	0.70/0.64/0.40	0.79/0.47/0.45	0.86/0.33/0.48	0.92/0.23/0.53	0.94/0.15/0.58	0.98/0.05/0.61	0.77/0.60/0.65
X			X	0.72/0.87/0.62	0.75/0.43/0.62	0.85/0.21/0.69	0.91/0.14/0.75	0.95/0.05/0.81	0.97/0.03/0.81	1.00/0.01/0.86	0.84/0.52/0.85
X	X	X		0.73/0.92/0.55	0.76/0.55/0.56	0.86/0.30/0.64	0.91/0.17/0.69	0.96/0.09/0.75	0.99/0.04/0.74	1.00/0.02/0.77	0.80/0.47/0.75
	X	X	X	0.46/0.80/0.35	0.63/0.54/0.42	0.78/0.39/0.51	0.86/0.26/0.58	0.94/0.17/0.67	0.97/0.14/0.68	0.99/0.10/0.73	0.73/0.52/0.64
X				0.68/0.79/0.66	0.72/0.33/0.66	0.84/0.23/0.72	0.90/0.14/0.79	0.96/0.06/0.84	0.98/0.04/0.88	1.00/0.02/0.91	0.82/0.47/0.91
	X			0.45/0.80/0.34	0.58/0.57/0.40	0.69/0.40/0.43	0.79/0.31/0.48	0.90/0.19/0.56	0.93/0.15/0.59	0.99/0.12/0.66	0.69/0.54/0.60
			X	0.43/0.82/0.32	0.60/0.60/0.39	0.73/0.42/0.46	0.83/0.27/0.51	0.93/0.17/0.61	0.96/0.09/0.62	0.99/0.10/0.70	0.71/0.54/0.64
	X			0.24/0.93/0.15	0.29/0.89/0.17	0.40/0.80/0.22	0.51/0.70/0.26	0.70/0.51/0.36	0.77/0.27/0.39	0.94/0.09/0.40	0.46/0.72/0.34