White Matter Lesion Segmentation for Multiple Sclerosis Patients implementing deep learning*

Theofilos G. Papadopoulos, Evanthia E. Tripoliti, *Senior Member IEEE*, Daphne Plati, Styliani Zelilidou, Kostas Vlachos, Spiros Konitsiotis and Dimitrios I. Fotiadis, *Fellow IEEE*

***Abstract*— The aim of this work is to address the problem of White Matter Lesion (WML) segmentation employing Magnetic Resonance Imaging (MRI) images from Multiple Sclerosis (MS) patients through the application of deep learning. A U-net based architecture containing a contrastive path and an expanding path prior to the final pixel-wise classification is implemented. The data are provided by the Ippokratio Radiology Center of Ioannina and include Fluid-Attenuated Inversion Recovery (FLAIR) MRI images from 30 patients in three phases, baseline and two follow ups. The prediction results are quite significant in terms of pixel-wise classification. The implemented deep learning model demonstrates Dice coefficient 0.7292, Precision 75.92% and Recall 70.16% in 2D slices of FLAIR MRI non-skull stripped images.**

I. Introduction

Multiple sclerosis (MS) is a chronic and degenerative neurological disease that affects the human central nervous system. It causes disorder in motility, sensation, vision, coordination and cognitive decline as well. MS manifests by the presence of white matter lesions (WML) in the brain, consisting of damaged white matter (WM) tissues associated with the increase of cerebrospinal fluid content [1]. Segmentation of MS lesions in Magnetic Resonance Imaging (MRI) scans is a useful way to understand and evaluate the progression of the disease. This task can be achieved either manually by experts or automatically by the elaboration of machine learning (ML) techniques. Deep learning (DL) is a branch of ML based on learning representations of data with multiple levels of abstraction [2]. More particularly, convolutional neural networks (CNNs) [3] have demonstrated outstanding performance in biomedical image analysis. CNN-based biomedical image segmentation methods can be categorized into two different groups: patch-based and image-based. In patch-based methods, the image is scanned by a moving window generating local representations, while in image-based methods the entire image is processed as a whole exploiting its global structured information [4]. Furthermore, these methods can be categorized to 2D slice-based segmentation and 3D based segmentation.

In the previous years, Valverde *et al.* [5] proposed a pipeline relying on a cascade of two 3D patch-based CNNs for T1-weighted (T1w), T2-weighted (T2w) and Fluid-Attenuated Inversion Recovery (FLAIR) skull-stripped images. The first network was trained using all extracted patches and the second network was used to refine the training procedure utilizing misclassified samples from the first network. Roy *et al.* [6] proposed a 2D patch-based CNN for T1w, T2w, Proton Density weighted (PDw) and FLAIR images including two pathways. Different MRI modalities were used as input for each pathway and the outputs were concatenated to create a membership function for lesions. Aslani *et al.* [4] in order to process simultaneously T1w, T2w and FLAIR images proposed a 2D network based on residual networks (ResNet) and CNNs including multiple parallel down-sampling, multi-modal feature fusion and multiscale up-sampling. Hashemi *et al.* [7] proposed a method relying on a 3D patch-based CNN using the idea of a densely connected network. Gessert *et al.* [8] developed a two path CNN architecture with attention-guided interaction. Cerri *et al.* [9] proposed an architecture consisting of two symmetrical CNNs integrating a generative approach. Multiple datasets for multiple combinations of T1w, T2w and FLAIR images were tested. A U-net like encoder-decoder is proposed by Kruger *et al.* [10] and La Rosa *et al.* [11] using 3D CNNs. The U-net architecture based on CNNs and developed by Ronneberger *et al.* [12] for biomedical image segmentation, utilized middle-level features, which were shown to have a considerable impact on the segmentation performance. In the present work, a slice-based U-net segmentation method is applied, in which 3D images are converted to their 2D slices and then the entire image is processed individually. FLAIR MRI is preferred since it has the advantage of highlighting WML lesions robustly [13]. The MRI images feed the model without skull-stripping.

II. Materials and Methods

A. The dataset

The dataset comes from the Ippokratio Radiology Center of Ioannina and includes FLAIR MRI images from 30 patients diagnosed with multiple-sclerosis (MS). 59 volumes, 180 images each, at three different time-points,

*Research supported by the ProMiSi project that has received funding from the Operational Program Epirus 2014-2020 and is co-financed by Greece and the European Union (University Hospital of Ioannina,) Ethic Committee Name: The board of the University Hospital of Ioannina, Approval Code: 3/20-2-2019, Approval Date: 18/03/2019

T. G. Papadopoulos, E. Tripoliti, D. Plati, and S. Zelilidou are with the Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110, Ioannina, Greece (e-mail: tpapado2011@gmail.com, etripoliti@gmail.com, daphni.plati@gmail.com, stylianizelil@gmail.com).

K. Vlachos is with Ippokratio Ioanninon S. A., GR45333, Ioannina, Greece (e-mail: kovlaxos@gmail.com).

S. Konitsiotis is with the Dept. of Neurology, Medical School, University of Ioannina, GR45110, Ioannina, Greece (e-mail: skonitso@gmail.com

D. I. Fotiadis is with the Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, GR 45110, Ioannina, Greece and with the Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110, Ioannina, Greece (phone: +302651009006, e-mail: fotiadis@uoi.gr).

Manuscript 1326 submitted to 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Received January 31, 2022.

baseline (month 0), 1st follow-up (6th month), 2nd follow-up (12th month), have been acquired in sagittal plane at 1.5 Tesla Phillips – KM7S1KG scanner. The dimension of each image is 622x541 pixels. 4918 images, 2045 from baseline, 2188 from the 1st follow up and 685 from the 2nd follow up were pixel-wise annotated defining the lesion segments by white pixels in a black background by an expert (Fig. 1).

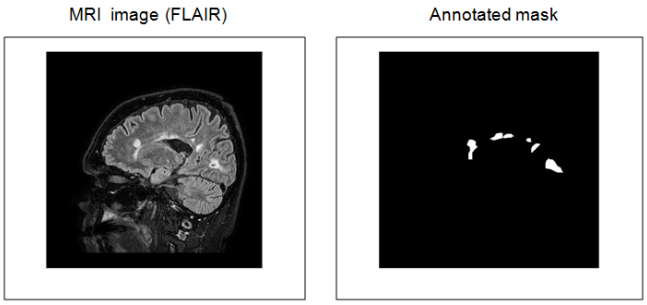

Figure 1. Raw FLAIR image (622x541) pixel by pixel annotated

B. The proposed methodology

The proposed method is based on the implementation of a U-Net [12] architecture. The U-net architecture integrates a contrastive path in which the number of feature maps increases and can be seen as analysis and an expanding path in which the number of the feature maps decreases and can be seen as synthesis [11]. Prior to feeding the network the only preprocessing of the input image is a size adaptation to the entry of the network by cropping the image to 256x256 and a scaling of the pixel values. Then, the generated output which has the same dimension with the input image, is compared to the annotated mask. The prediction problem of segmentation is posed as a pixel-wise classification problem.

Preprocessing

The initial images were cropped in order to remove the useless outer white edges and fit the model input which is designed to accept images 256x256x1 (Fig 2). The pixels of the images being stored as grayscale images take values in a range from 0 to 255. Zero is taken to be black and 255 to be white. A normalization is applied that transforms the range of the pixel's values to [0,1].

Model description

The model's architecture is illustrated in Fig. 3. It is a U-net [12] that contains two paths. The top down path on the left side is the encoding path, which captures the context of the image. It is a contrastive path in which down-sampling occurs. It consists of repeated convolution layers, followed by rectified units (ReLU) [14, 15] and pooling layers [2]. There are four levels of top down operations. At each top down level of operation, two consecutive convolutions with receptive field 3x3 and padding fixed as "same" are followed by 2x2 max pooling with stride 2 for down-sampling. At each down-sampling step the feature maps (i.e. the layer depth) are doubled by doubling the filters. In the lower level at the bottom, two convolutions are applied which increase only the layer depth. The bottom up path on the right is a decoding path symmetric to the encoding path which enables precise localization. It is an expanding path where up-sampling occurs by transposed convolutions [16].

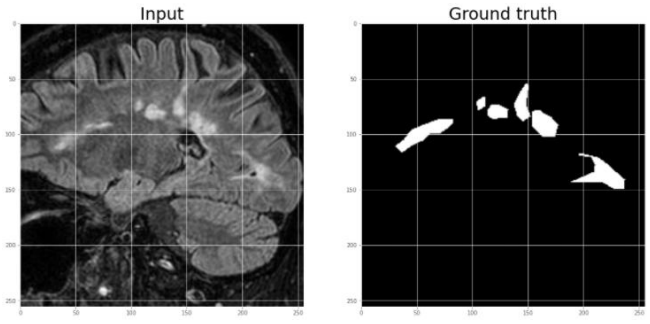

Figure 2. The image and the annotated mask as ground truth which feed the model. Cropped to fit the model input (256x256x1)

It consists of four operational levels. At each bottom up operational level a deconvolution operation (i.e. transposed convolution) [17, 18] is followed by a skip connection (i.e. concatenation) with the feature map of the corresponding top down level and two 3x3 convolutions. The skip connections are used to get better precise locations at every step of the decoder. At the end a 1x1 convolution with sigmoid activation is used to map the final layer to the output. The network is an end-to-end fully convolutional network [19] (i.e. only contains convolutional layers and does not contain any dense layer). In the encoder, the size of the image gradually reduces while the depth gradually increases starting from 256x256x1 to 16x16x256. In the decoder, the size of the image gradually increases and the depth gradually decreases starting from 16x16x256 to 256x256x1.

The model is compiled with Adam optimizer [20] and He initialization [21]. The binary cross entropy loss function is used pixel-wise since there are only two output classes (black and white). Threshold 0.5 is applied to decide whether to classify a pixel as 0 or 1. Then, learning rate decay is implemented from 0.001 to 0.00001 if the validation loss does not improve for 5 epochs. To prevent the model from overfitting dropout of rate 0.1 [22], batch normalization [23] and early stopping if the validation loss does not improve for 10 epochs, are implemented. Keras, Tensorflow and Python 3.8 are used to build, train and evaluate the model.

III. RESULTS

The proposed method is evaluated on a dataset of 4918 MRI FLAIR images originated from 30 patients. The dataset is split into train-test sets, 70%-30% accordingly. 10% of the training set is used for validating the model during training. Thus, 3098 instances are used for training, 344 for validation and 1476 for testing the model. The validation set is employed during training for early stopping and learning rate scheduling. The test set is never seen by the model during training. The batch size is set to 5. Early stopping occurs at the 18th epoch.

The results are evaluated in terms of metrics such as Accuracy, Intersection over Union (IoU), Dice coefficient, Positive Predictive Value (Precision) and True Positive Rate (Recall).

Accuracy measures the percent of pixels in the image which are correctly classified. Although simple conceptually, high pixel accuracy doesn't always imply superior segmentation ability. High classification ability of the majority class (in our case the black pixels) favors it.

Manuscript 1326 submitted to 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Received January 31, 2022.

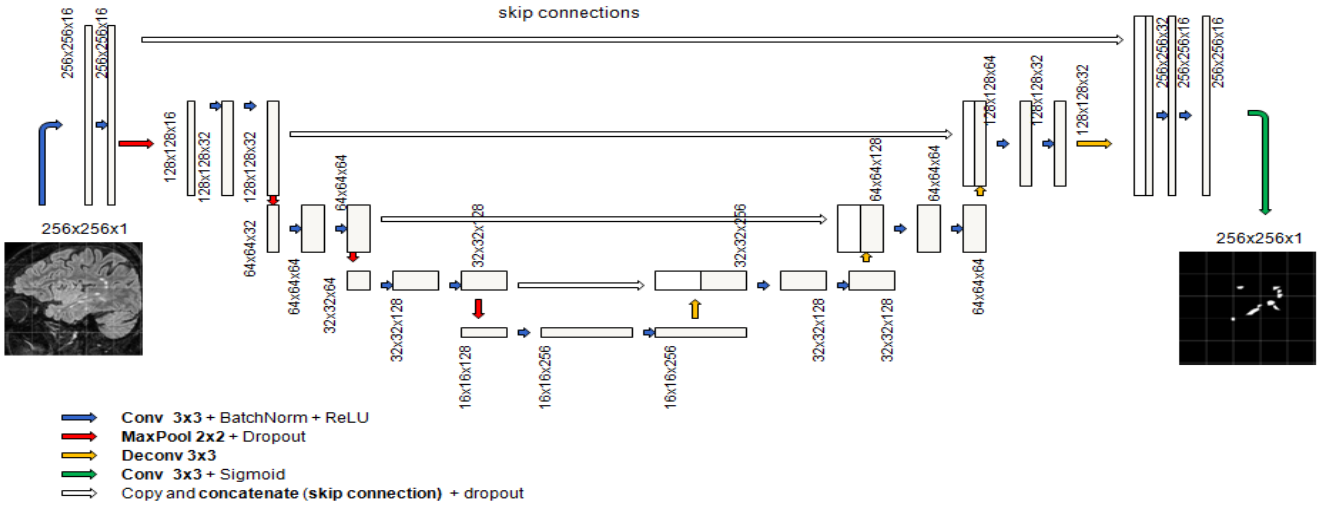

Figure 3. The U-net model with a down-sampling and an up-sampling path and the skip connections between them

IoU, also known as Jaccard index, measures the overlapping area between two segmentation masks A and B:

$$\text{IoU} = \frac{||A \cap B||}{||A \cup B||} \quad (1)$$

where ||A|| is the norm of A (for images, the area in pixels) and $\cap$, $\cup$ are the intersection and union operators respectively.

Dice coefficient is the area of overlap divided by the total number of pixels in both images:

$$\text{Dice}(A, B) = \frac{2 * ||A \cap B||}{||A|| + ||B||} \quad (2)$$

Both the Dice and IoU indices are bounded from 0 (when there is no overlap) to 1 (when A and B match perfectly).

Positive Predictive Value (PPV) or Precision is defined as the ratio of the total number of correctly classified positive pixels (in our case white pixels) divided by the total number of predicted as positive pixels. It tells us about when it predicts yes, how often is it correct.

True Positive Rate (TPR) or Recall is defined as the ratio of the total number of correctly classified positive pixels (white pixels) divided by the total number of the real positive pixels. It tells us about when it is actually yes, how often it predicts yes.

The testing results of the model in terms of the above metrics are reported in TABLE I. They are averaged over all the testing images. Within parentheses the standard deviation is reported as well.

TABLE I. EVALUATION RESULTS

Metrics evaluated on the test set of 1476 images never seen by the model	
Accuracy (pixel-wise)	99.63% (± 0.04)
Intersection over Union (IoU)	0.5739 (± 0.01)
Dice coefficient (DSC)	0.7292 (± 0.008)
Precision	75.92% (± 0.88)
Recall	70.16% (± 0.9)

IV. DISCUSSION

The key advantage of DL is the automated learning of features from raw data without the need of complex preprocessing and hand designed feature extractors [24]. In the current work, the performance achieved by the implemented U-net based deep model employing only 2D sliced FLAIR images, without patch-based data augmentation and skull stripping is quite high, compared to previous studies (TABLE II) mainly in terms of image overlapping which is well expressed by the Dice coefficient (0.7292). The trained model seems to have the ability to predict well the magnitude and the position of the lesions as it is illustrated in Fig. 4. Moreover, as it is known, in the nature of WML image segmentation problem there is an unbalance between the minority class of lesion pixels and the majority class of the background pixels which lead to predictions with high precision and low recall, biased towards the non-lesion class. For the same reason accuracy is not a reliable measure of segmentation because the majority class of background pixels favors it. In order to mitigate this problem, alternative more complex loss functions have been proposed in the literature [7]. Nevertheless, in our implementation a quite balanced

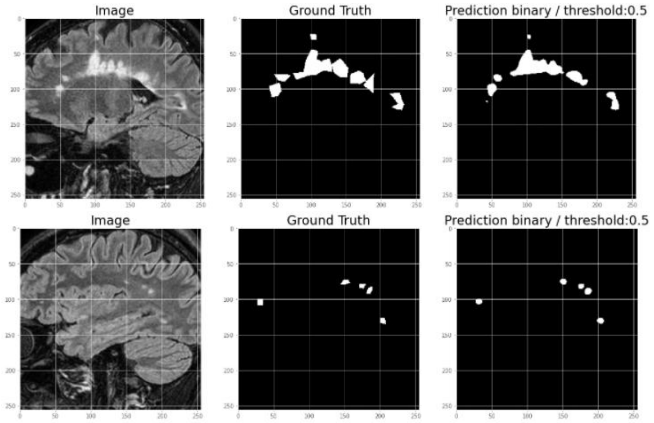

Figure 4. Illustration of the predicted results. The predicted masks are quite close to the annotated (considered as ground truth)

performance of Precision 75.92% vs Recall 70.16% is achieved using the standard binary cross-entropy loss.

TABLE II. STATE OF THE ART IN DEEP LEARNING FOR MS WML SEGMENTATION

Study	*Dataset*	*Evaluation measures*
Present Study	30 patients from Ippokration Radiology Center of Ioannina, Greece Only FLAIR images	IoU: 0.5739 **Dice: 0.7292** PPV(Precision): 0.7592 TPR (Recall): 0.7016
Valverde *et al.* [5]	MICCAI2008 Clinical data 35 subjects containing T1w, FLAIR, PDw 25 patients containing T1-w and FLAIR	MICCAI2008 TPR: 0.687 FPR: 0.46 Clinical data Dice: 0.56 TPR:0.682 FPR:0.336 PPV:0.661
Roy *et al.* [6]	MS-100 dataset (128 MS patients) T1-w, T2-w, FLAIR and PD images	Dice: 0.5639 PPV:0.604
Aslani *et al.* [4]	ISBI (14 patients) NRU A private dataset (37 MS patients) T1-w, T2-w, FLAIR images	ISBI 2015 PPV: 0.8992 Dice: 0.6114 NRU PPV: 0.8032 Dice: 0.6655
Kruger *et al.* [10]	3.253 patient data from 103 different MR scanners Zurich data / University Hospital of Zurich Dresden data /University Hospital CarlGustav Carus FLAIR images	Dice: 0.45 Sens :0.60 FPR :0.41
Hashemi *et al.* [7]	MSSEG 2016 (15 subjects from three scanners) T1-w, T2-w FLAIR, Gadolinium-enhanced T1-w, PD-w and T2-w MRIs	U-net Dice:0.5642/Sens:0.5685/ Spec:0.9993 FCDensenet Dice:0.703/Sens:0.7449 / Spec:0.9995
Gessert *et al.* [8]	89 MS cases from University Hospital of Zurich, Switzerland FLAIR images	Dice : 0.656 FPR : 0.264 TPR : 0.742
La Rosa *et al.* [11]	54 patients from Basel University Hospital 34 patients from Lausanne University Hospital T1-w, T2-w, FLAIR images	Dice : 0.62 PPV : 0.61 TPR : 0.76 FPR : 0.29
Cerri *et al.* [9]	**MSSeg:** 15 MS cases from three different scanners **Trio:** 40 MS cases at the Danish Research Center of Magnetic Resonance (DRCMR) **Achieva:** 50 MS cases and 25 healthy controls at DRCMR **ISBI** :14 patients T1w and FLAIR images	Dice MSSeg 0.65 Trio 0.58 Achieva 0.54 ISBI 0.58

V. CONCLUSION

The proposed U-net based model following a down-sampling and an up-sampling path interconnected with skip connections seems to understand "what" is there in the image and to learn "where" it is present. Thus, it achieves an overlapping performance of Dice coefficient 0.7292, Precision 75.92% and Recall 70.16%.

REFERENCES

[1] L. Steinman, Multiple sclerosis: a coordinated immunological attack against myelin in the central nervous system, Cell 85(3) (1996) 299-302.

[2] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86(11) (1998) 2278-2324.

[3] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, Nature 521 (2015) 436-44.

[4] S. Aslani *et al.*, Multi-branch Convolutional Neural Network for Multiple Sclerosis Lesion Segmentation, 2018, p. arXiv:1811.02942.

[5] S. Valverde *et al.*, Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach, Neuroimage 155 (2017) 159-168.

[6] S. Roy *et al*., Multiple Sclerosis Lesion Segmentation from Brain MRI via Fully Convolutional Neural Networks (2018).

[7] S.R. Hashemi *et al.*, Asymmetric Loss Functions and Deep Densely Connected Networks for Highly Imbalanced Medical Image Segmentation: Application to Multiple Sclerosis Lesion Detection, IEEE access : practical innovations, open solutions 7 (2019) 721-1735.

[8] N. Gessert *et al.*, Multiple sclerosis lesion activity segmentation with attention-guided two-path CNNs, Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society 84 (2020) 101772.

[9] S. Cerri *et al.*, A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis, NeuroImage 225 (2021) 117471.

[10] J. Krüger *et al.*, Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks, NeuroImage. Clinical 28 (2020) 102445.

[11] F. La Rosa *et al.*, Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE, NeuroImage: Clinical 27 (2020) 102335.

[12] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv e-prints (2015) arXiv:1505.04597.

[13] P.B. Gorelick *et al.*, Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the american heart association/american stroke association, Stroke 42(9) (2011) 2672-713.

[14] V. Nair, G. Hinton, Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair, 2010.

[15] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Neural Information Processing Systems 25 (2012).

[16] V. Dumoulin, F.J.a.e.-p. Visin, A guide to convolution arithmetic for deep learning, 2016, p. arXiv:1603.07285.

[17] H. Noh, S. Hong, B. Han, Learning Deconvolution Network for Semantic Segmentation, arXiv e-prints (2015) arXiv:1505.04366.

[18] M.D. Zeiler, G.W. Taylor, R.J.I.C.o.C.V. Fergus, Adaptive deconvolutional networks for mid and high level feature learning, (2011) 2018-2025.

[19] J. Long, E. Shelhamer, T. Darrell, Fully Convolutional Networks for Semantic Segmentation, arXiv e-prints (2014) arXiv:1411.4038.

[20] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, arXiv e-prints (2014) arXiv:1412.6980.

[21] K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1026-1034.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, Journal of Machine Learning Research 15 (2014) 1929-1958.

[23] S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, arXiv e-prints (2015) arXiv:1502.03167.

[24] Y. Bengio, A. Courville, P. Vincent, Representation Learning: A Review and New Perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8) (2013) 1798-1828.

Manuscript 1326 submitted to 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Received January 31, 2022.