Edges-enhanced Convolutional Neural Network for Multiple Sclerosis Lesions Segmentation

Gustavo Ulloa-Poblete[1], Héctor Allende-Cid[2],
Alejandro Veloz[3], Héctor Allende[1]

[1] Universidad Técnica Federico Santa María,
Chile

[2] Pontificia Universidad Católica de Valparaíso,
Chile

[3] Universidad de Valparaíso,
Chile

gustavo.ulloa@usm.cl, hallende@inf.utfsm.cl

Abstract. Multiple sclerosis (MS) segmentation is a crucial task that helps to monitor the progression of that condition and to investigate how efficient is the treatment provided to a patient. Convolutional Neural Networks (CNN) have been successfully employed in MS lesion segmentation in recent years, but still have problems in segmenting voxels in the boundaries of the lesions. In this work, we present a modified CNN that assign more importance in learning hard-to-classify voxels close to the boundaries of the MS lesions. During the training process, we performed a stratified sampling to dynamically increase the penalization of voxels in the neighborhood around MS lesions boundaries. We prove that the stratified sampling strategy increases the representation of voxels near to the neighborhood of the edges and retrieves more accurate results in terms of Dice similarity coefficient compared to existing methods that uses uniform sampling. To test our approach, the 2015 Longitudinal MS Lesion Segmentation Challenge dataset was used, obtaining $\text{Dice} > 0.7$, which is comparable to the performance of human experts.

Keywords. Convolutional neural networks, focal loss, multiple sclerosis, lesions segmentation, magnetic resonance imaging.

1 Introduction

Multiple sclerosis (MS) is a chronic autoimmune inflammatory neurological disease affecting the central nervous system (CNS) in which the autoimmune system attacks the myelin sheath, myelin-producing cells and axons present in the white matter [7].

This can lead to a progressive loss of sensory, visual, motor and cognitive brain function in people with this disorder.

Magnetic Resonance Imaging (MRI) is used to detect MS white matter lesions using T1-weighted (T1-w), T2-weighted (T2-w), PD-weighted (PD-w), and fluid-attenuated inversion recovery T2 (FLAIR) sequences.

MS lesions appear on MRI as small areas of hyperintensity on T2-weighted, PD-w, and FLAIR MR sequences and small areas of hypointense on T1-weighted MR sequences.

Medical professionals need quantitative data on the amount of lesions in the myelin sheath for diagnosis, treatment and prognosis. They also need this data to test new drugs [8, 20]. From their perspective, image processing techniques such as segmentation help radiologists accurately identify lesions caused by myelin damage. Computerized segmentation of lesions – related to Multiple Sclerosis – dramatically reduces the time it takes radiologists to analyze the progression of this condition. And it decreases the intra and inter-expert variability when assessing the progressions.
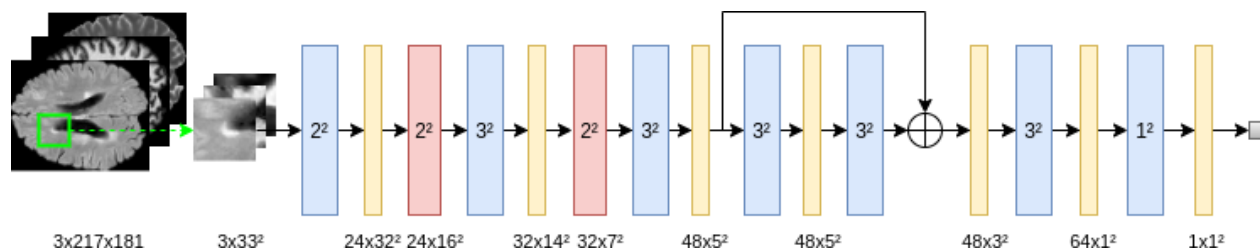
Fig. 1. Red blocks: 2×2 maxpooling layers, Blue blocks: convolutional layers, Yellow blocks: Leaky-ReLU activation functions, except the classification layer use sigmoidal activation. One residual connection present in the network

In the last years, using deep learning architectures based on Convolutional Neural Networks (CNNs), state-of-the-art results have been obtained in most of visual recognition tasks; in particular in image classification, object detection and semantic segmentation [13, 14].

The use of deep learning in the health area allows not only to detect pathologies, it has also allowed to support the medical team in the prognosis of the evolution of diseases through the use of information or expert knowledge extracted from images whose modalities can vary from MRIs, microscopy, photoacoustic and color images [12, 6, 13].

In the MRIs of patients with multiple sclerosis, methods have been introduced that consider probabilistic analysis [19], characteristics such as the hyperintensity of the lesions in the FLAIR channel [1] and spatial and spectral information of the lesions [18].

The usual way to perform the segmentation task of MS-lesions is by using as input a patch surrounding the interest voxel to classify in one of the defined classes.

A voxel is understood as an element of a three-dimensional array whose intensity represents a volume which may or may not be isometric. Typically, the patches correspond to a set of intensities determined by uniform sampling in a square neighborhood centered in the voxel of interest.

Segmenting MS-related lesions is difficult due to the partial volume effect, variability of MS lesions' intensities and the overlapping between MS lesions and normal brain tissues distributions [5].

This makes the accurate determination of MS lesions borders a problem difficult to solve for professionals and machines. Indeed, segmentation is hard to determine at a glance by physicians [2, 16].

These difficulties causes that most segmentation algorithms present a high proportion of false positive and false negative errors in MS lesions domains close to the borders.

In this work, a methodology is proposed that allows addressing the problem of segmentation of multiple sclerosis lesions in an efficient way.

A preprocessing of the MRIs is carried out that corrects the excessive truncation of the hyperintensity voxels of the FLAIR modality used in [2], since it is in these voxels where the CNN presents the highest false positive rates.

Next, a stratified sampling of voxels difficult to classify near the edges is carried out in order to increase their representation in the training set.

Finally, the focal loss function is used, which allows the penalty to be focused on the most difficult examples of the training set.

In this paper, the methodology proposed in [15] is used, to which is added the use of a convolutional neural network architecture that includes a residual block (ResNet) with the aim of increasing the depth of CNN and preserving the gradient backpropagated to the more superficial layers.

In the case of the Dice similarity coefficient (4), a t-test of mean is performed to determine the best combination of parameters of the proposed methodology. The training procedure is modified, a cross-validation training of 100 epochs is performed without using early stopping due to the low number of MRI volumes of patients as
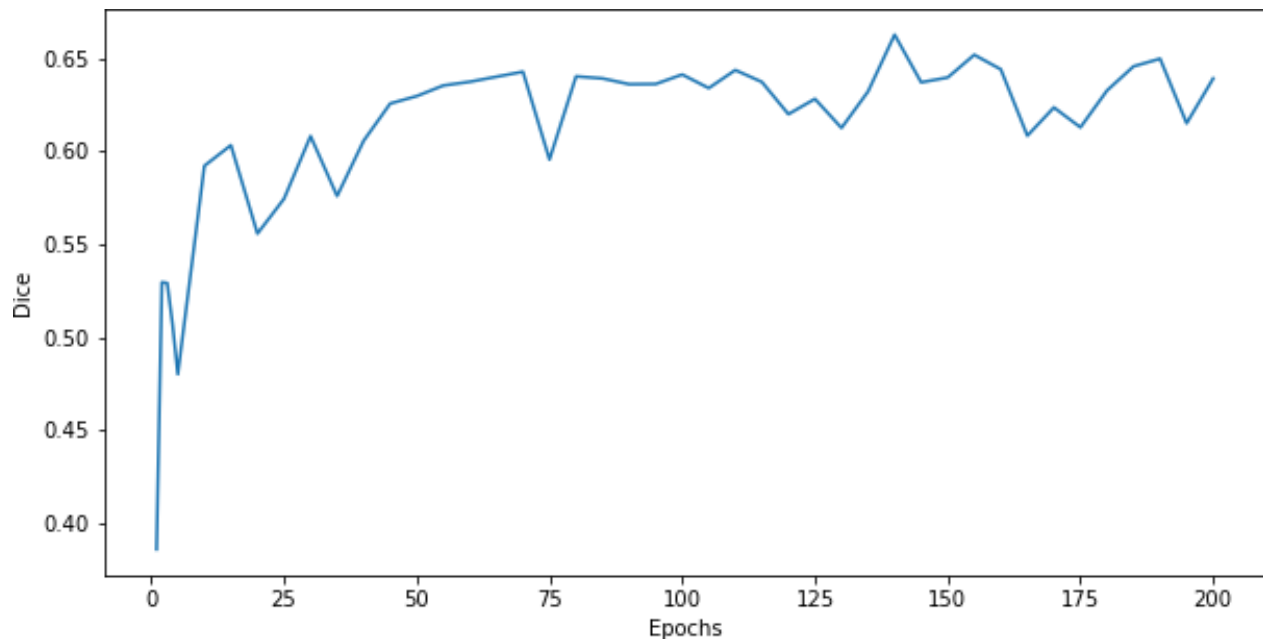
Fig. 2. Evolution of the Dice similarity coefficient during training using a validation example

examples. The paper is organized as follows: In Section 2 we expose the methodology, the steps used for preprocessing, the architecture of the convolutional network model with focal loss.

Section 3 presents discussions about the results obtained on the 2015 Longitudinal MS Lesion Segmentation Challenge dataset. Finally, in Section 4 we present conclusions and future works.

2 Materials and Methods

2.1 Data Set

To show how the proposed method performs we used the longitudinal multiple sclerosis lesion segmentation data set [3].[1] The training set contains 4-5 volumes of multi-modal MR images of five patients.

The available modalities were acquired on a 3T MRI scanner and corresponds to T1-weighted, T2-weighted, PD-weighted and FLAIR.

---

[1]Publicly available at https://smart-stats-tools.org/lesion-challenge.

Successive volumes were acquired in time intervals that allow to reflect disease progression. For each image, the corresponding manual lesion segmentation is also provided.

2.2 Preprocessing Pipeline

The preprocessing steps that were applied correspond to the standard preprocessing pipeline typically performed to prepare MR images, i.e. correction of the MR intensity inhomogeneities, skull-stripping, dura mater stripping, and rigid-body registration to a 1 mm isotropic MNI template [3].

In addition to the previously described standard MRI preprocessing pipeline, a series of additional operations were performed with the purpose of generating high quality image patches to train the CNN model.

These operations are summarized below. The image intensities were truncated to the quantiles within the range $[0.01, 0.9995]$, and then scaled to the range $[0, 1]$.

The truncation quantiles were determined experimentally in such a way that they allow to achieve better correspondence between MR image

Table 1. Group cross validation sets

Trainning		Testing	
Patients	#MRIs	Patients	#MRIs
2,3,4,5	17	1	4
1,3,4,5	17	2	4
1,2,4,5	16	3	5
1,2,3,5	17	4	4
1,2,3,4	17	5	4

intensities, especially in high intensity voxels in T2 and FLAIR images which are potentially part of MS-related lesions.

We found that this process allows to achieve results even better than other approaches used to maximize the contrast between image intensities (e.g., histogram matching).

The quantiles were determined experimentally in this work. We found values relatively close to quantiles found by other researchers [2]. To train the CNN model, around of 650.000 patches were generated from the available images.

This was done by selecting image patches that contain voxels that satisfy two conditions: (i) the patch contains hyper-intense voxels on FLAIR images which have the potential of being a MS-related lesion [5], and (ii) the patch contains white matter.

To satisfy the condition (i) a set of supra-threshold voxels were determined from the FLAIR images, and to satisfy the condition (ii) the subset of voxels belonging to white matter were determined.

This was done by registering the FLAIR images to the ICBM452 probabilistic atlas using a rigid-body registration algorithm implemented in FSL [9].

2.3 Network Architecture and Training

The CNN used has 7 convolutional layers. The input of the network consists of three patches of size 33×33. The selected CNN architecture was initially inspired by the architecture presented in [2], but the number of hidden layers was increased.

This increment of the model's hidden layers was motivated by our empirical findings in terms of the Dice similarity criterion. The inputs of the model correspond to patches generated from the T1-weighted, T2-weighted, and FLAIR images.

The PD-weighted images were discarded as they do not contribute to the CNN model's performance according to our empirical study. Each convolutional layer (except for the output layer) uses Leaky ReLU activation function with negative slope coefficient $\beta = 0.3$.

The output layer corresponds to a 1-dimensional sigmoidal activation function that retrieves the probability of a voxel belonging to the MS-related lesion class.

By evaluating a whole image, a probability map can be generated with the purpose of solving the segmentation problem as a voxel-to-voxel binary classification task. The network contains 2 layers of MaxPooling of 2×2 with stride 2.

Dropout after each convolutional layer was used as regularization method. In order to increase the depth of the network without generating adverse effects such as the problem of the vanishing gradient, a single ResNet block was included into the model.

We empirically found that by increasing the number of ResNet blocks the overall model's performance is not improved. Figure 1 summarizes the network architecture. The image sampling performed to generate patches is highlighted with green squares.

Considering that our CNN model has a one dimensional output, the loss functions used for training was the binary cross entropy and focal loss with parameter values $\gamma = 1.0$, $\gamma = 2.0$, $\gamma = 3.0$ and $\gamma = 5.0$.

In addition, a mechanism to improve the CNN model performance in difficult voxels near the edges of the lesions was implemented. This mechanism consisted in balancing the number of patches of the non-lesion class and lesion class with the ratio $1 : 1$.

A stratified sampling process was performed considering a proportion p for voxels located at a distance not greater than 4 mm to the edges of the MS lesions and a proportion $1 - p$ for voxels randomly sampled from voxels candidates that are

Table 2. The results of experiments with various choices of the parameters proportion of stratified sampling near the edges (s) and γ parameter of focal loss

LF	s	γ	bt	HD(sd)	HD95(sd)	ASSD(sd)	Sensitivity(sd)	Precision(sd)	Dice(sd)	p-value
BCE	-	-	0.99	35.2901(1.9499)	18.7378(1.3450)	1.6591(0.1311)	0.6887(0.0186)	0.7346(0.0143)	0.6961(0.0108)	4.2×10^{-8}
FL	-	1.0	0.92	34.1959(1.2904)	18.8834(0.9477)	1.6966(0.0731)	0.7006(0.0139)	0.7177(0.0125)	0.6941(0.0067)	1.4×10^{-11}
FL	-	2.0	0.84	33.3430(1.7256)	18.4407(0.9899)	1.6486(0.1049)	0.6959(0.0206)	0.7280(0.0142)	0.6977(0.0106)	2.1×10^{-7}
FL	-	3.0	0.77	34.8955(1.4089)	18.7555(1.1646)	1.7146(0.1161)	0.7002(0.0196)	0.7241(0.0129)	0.6958(0.0095)	6.2×10^{-9}
FL	-	5.0	0.68	34.2176(1.4506)	19.361(1.3372)	1.6948(0.1423)	0.6908(0.0182)	0.7478(0.0127)	0.7003(0.0119)	6.6×10^{-5}
BCE	0.2	-	0.97	39.5863(2.4359)	22.5021(0.9233)	1.7167(0.1546)	0.6920(0.0168)	0.7587(0.0174)	0.7077(0.0110)	8.9×10^{-3}
FL	0.2	1.0	0.86	39.0989(1.9471)	21.8893(0.5873)	1.7305(0.1235)	0.7143(0.0127)	0.7442(0.0130)	0.7143(0.0084)	3.9×10^{-1}
FL	0.2	2.0	0.78	39.5909(1.8391)	22.0459(0.7287)	1.7083(0.1002)	0.7014(0.0138)	0.7544(0.0092)	0.7118(0.0099)	1.3×10^{-1}
FL	0.2	3.0	0.72	38.6024(1.6650)	21.9691(0.7630)	1.6930(0.1277)	0.7023(0.0156)	0.7593(0.0146)	0.7150(0.0094)	–
FL	0.2	5.0	0.64	39.2317(1.0066)	22.5822(0.6221)	1.8491(0.1692)	0.6984(0.0153)	0.7535(0.0143)	0.7064(0.0119)	8.3×10^{-3}
BCE	0.4	-	0.95	44.1661(1.5941)	24.7171(0.6144)	2.0616(0.1768)	0.7012(0.0146)	0.7401(0.0165)	0.7031(0.0118)	2.7×10^{-4}
FL	0.4	1.0	0.84	43.3542(2.2077)	24.2018(0.6012)	1.9777(0.1352)	0.6954(0.0193)	0.7482(0.0164)	0.7040(0.0086)	1.0×10^{-4}
FL	0.4	2.0	0.75	42.4496(1.6333)	24.4972(0.4589)	1.9677(0.1245)	0.6969(0.0135)	0.7486(0.0108)	0.7058(0.0091)	8.9×10^{-4}
FL	0.4	3.0	0.69	43.0914(1.1723)	24.5159(0.7112)	2.0125(0.1257)	0.6970(0.0141)	0.7441(0.0127)	0.7032(0.0072)	1.4×10^{-5}
FL	0.4	5.0	0.63	44.5579(1.7654)	24.7383(0.5421)	2.1408(0.1826)	0.6871(0.0160)	0.7488(0.0165)	0.6978(0.0093)	1.2×10^{-7}

distant more than 4 mm to the MS lesions. The training of the CNN model was made by using the stochastic gradient descent strategy.

The learning rate selected was 0.01 with decay learning rate of 1×10^{-6}, the momentum parameter was 0.9 with Nesterov momentum and batch size of 64. Dropout with probability 0.1 was used.

The implementation and training of the CNN was done using the Keras Deep Neural Network Library [4] with TensorFlow as backend numerical engine. The model was trained using NVIDIA GeForce GTX 1080TI Graphic Card.

To compare the segmentation performance, we implemented a group five-fold cross validation with five identical models. For each fold, the training and testing sets were determined is such a way no patches of the same patient belongs to the testing and training sets simultaneously.

The details are presented in Table 1. The number of epochs was determined in such a way that the Dice similarity coefficient reaches a stable value in the validation set (a total of 100 epochs was applied for training, see Figure 2).

2.4 Focal Loss Function

Our model was trained using the focal loss function that was proposed in [10], with the purpose of addressing the class imbalance problem, specifically in CNN models trained in one-stage.

The focal loss function is based on the binary cross entropy loss function, which is given by:

$$\text{BCE}(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}), \quad (1)$$

where $y \in \{0, 1\}$ is the ground truth, $y = 0$ and $y = 1$ correspond to the non-MS and MS-lesion classes, respectively, and $\hat{y} \in [0, 1]$ is the probability of the class with label $y = 1$ estimated by the model. For notation convenience, $\hat{y}_t$ is defined as:

$$\hat{y}_t = \begin{cases} \hat{y} & \text{if } y = 1, \\ 1 - \hat{y} & \text{otherwise.} \end{cases} \quad (2)$$

Thus, for the t-th input instance, the BCE can be rewritten as: $\text{BCE}(\hat{y}_t) = -\log(\hat{y}_t)$.

In the focal loss function, a weighting term given by $(1 - \hat{y}_t)^\gamma$ is added to the binary cross entropy loss function. According to this, the focal loss function can be expressed as:

$$\text{FL}(\hat{y}_t, \gamma) = -(1 - \hat{y}_t)^\gamma \log(\hat{y}_t), \quad (3)$$

where γ is a parameter that modulates the weighting effect provided to this loss function.

The focal loss weights dynamically the training examples losses, down-weighting easy-to-classify training examples and assigning higher losses to examples with low prediction accuracy, i.e. hard-to-classify training examples.
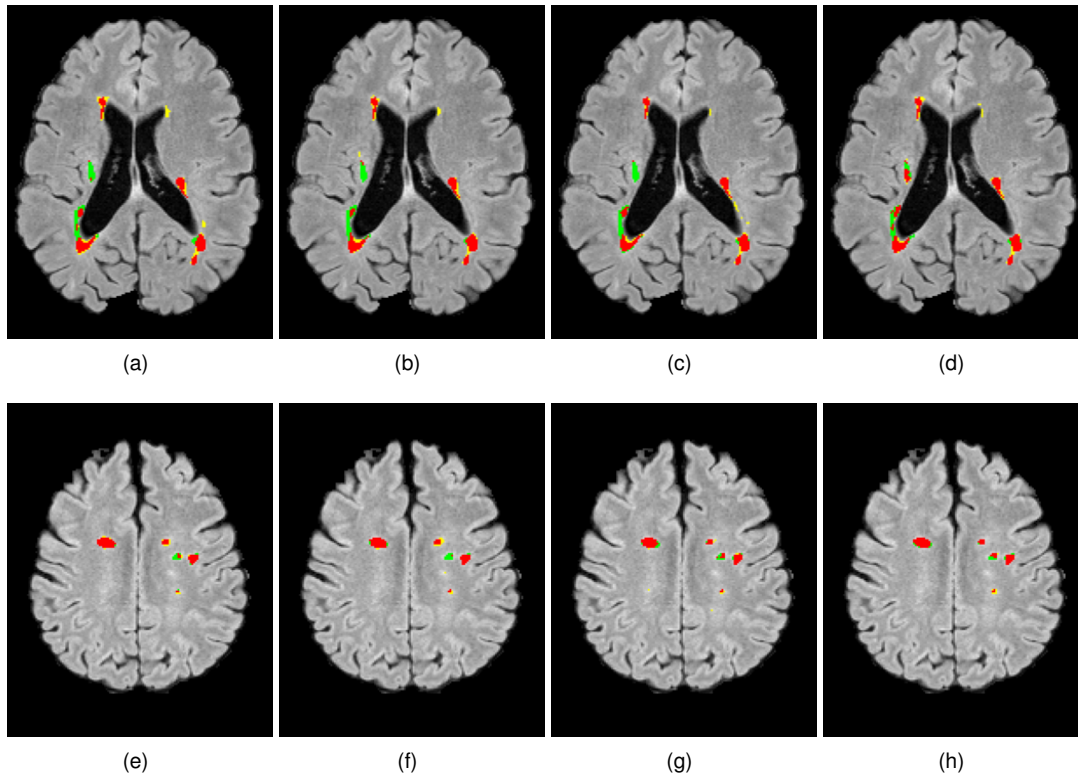
(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3. Automatic segmentation: (a)-(d) pacient 1 (e)-(f) pacient 5

3 Experimental Results

3.1 Evaluation

The accuracy of the CNN model was computed by comparing the segmentation mask predicted by the model S and the gold standard G manually segmented by the expert. The evaluation metrics used in this work aim to highlight the spatial-overlap and spatial-distance and are defined below [11].

Spatial-overlap metrics:

— The Dice similarity coefficient or F_1-score is an overlap measure between two binary label masks:

$$\text{Dice}(G, S) \quad = \quad \frac{2\,\text{TP}}{2\,\text{TP} + \text{FN} + \text{FP}}, \quad (4)$$

where TP, FN and FP correspond to the amount of true positive, false negative and false positive values.

— Sensitivity or True Positive Rate indicates the rate of voxels correctly segmented as lesions:

$$\text{Sensitivity}(G, S) \quad = \quad \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (5)$$

— Precision or Positive Predictive Value indicates the rate of segmented voxels from ones estimated as lesions:

$$\text{Precision}(G, S) \quad = \quad \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (6)$$

Spatial-distance metrics:

— The Hausdorff distance (HD) indicates the largest segmentation error considering the edges or borders between two objects present

in binary images. Intuitively, it is the longest distance one has to travel from a point in one of the two sets to its closest point in the other set:

$$HD(G, S) = \max\left(hd(G, S), hd(S, G)\right), \quad (7)$$

where:

$$hd(G, S) = \max_{g \in G} \min_{s \in S} \|g - s\|_2. \quad (8)$$

The Hausdorff distance is usually adversely affected by the presence of outliers since in its construction it includes the max statistic, which is non-robust. Due to this fact, the Hausdorff distance and the percentile 95% are usually used in the literature.

— Average symmetric surface distance (ASSD) is the average of the distance from the lesions in G to the nearest lesion identified in S plus the distance from the lesions in S to the nearest lesion identified in G [17]:

$$ASSD(G, S) = \frac{\sum_{x \in \partial G} d(x, \partial S) + \sum_{y \in \partial S} d(y, \partial G)}{|\partial G| + |\partial S|}, \quad (9)$$

where $d(x, \partial S) = \min_{y \in \partial G} \|x - y\|_2$ and $d(y, \partial G) = \min_{x \in \partial S} \|x - y\|_2$, $|\partial G|$ and $|\partial S|$ denote the number of points on the surfaces of G and S, respectively.

3.2 Results and Discussion

Table 2 presents the results obtained by our proposed CNN network (Figure 1) using stratified sampling together with the binary cross entropy loss functions (BCE) and focal loss (FL).

The parameter s corresponds to the proportion of voxels sampled near to the lesions edges respect to the total number of voxels of the non-lesion class and γ is the focal loss parameter (see Eq. (3)).

The metrics used correspond to, Hausdorff distance (Eq. ((7)), Average symmetric surface distance (Eq. (9)), Sensivity (Eq. (5)), Precision (Eq. (6)) and Dice (Eq. (4)).

Regarding the results obtained by the spatial-distance metrics, it can be noted from Table 2 that the HD and ASSD metrics are sensitive to the presence of outliers, due to the fact that the mere presence of a lesion not detected by CNN or a lesion detected erroneously can generate a large number of voxels greatly distant from the lesions in the gold standard.

That is why this metric is often used in evaluating the image segmentation of big and high-contrast organs. For metrics based on spatial-overlap, the best result about the sensitivity metric corresponded to $0.7143(0.0127)$ using $s = 0.2$ and $\gamma = 1.0$.

For the precision metric, the best result was $0.7593(0.0146)$ obtained using the parameters $s = 0.2$ and $\gamma = 3.0$. These results show that this methodology reduces false negative and false positive errors.

In the particular case of false positive errors they are reduced thanks to the stratified sampling strategy that increases the representation of difficult voxels close to the MS lesions.

The best results in terms of the Dice similarity coefficient corresponded to $0.7150(0.0094)$, obtained using $s = 0.2$ and $\gamma = 3.0$, which indicates that using focal loss, the increased representation of difficult-to-classify voxels by the stratified sampling strategy will affect positively the model's performance compared to uniform sampling, which has $s = 0.026$ (in average).

The best result in the Dice similarity coefficient was compared with all other combinations of parameters by means of a t-test with the assumption of different and unknown variances (Welch's test), where the null hypothesis corresponds to the assumption that there are no differences of the Dice means with the different combinations.

The alternative hypothesis is that this result has a higher Dice mean. It must be noted in Table 2 that the p-values were less than 0.01 when using a uniform sample ($s = -$), i.e. a non-stratified sample, independent of γ parameter of the focal loss.

A low p-value is also observed in stratified samples $s = 0.2$ and $s = 0.4$ only when binary cross entropy loss (BCE) since using focal loss increases the value of the Dice similarity

ISSN 2007-9737

coefficient, thus decreasing the power of the test (p-value> 0.1), except for $s = 0.4$ where FL has a marginal incremental effect.

We also observed that the use of focal loss drastically influences the decrease of the binarization threshold (bt) necessary to maximize the results of the spatial-distance metrics and the Dice similarity coefficient, especially when stratified sampling was used.

This increases the proportion of non-lesion voxels close to the edges MS-lesions in the training set, which supports the hypothesis that these areas contain most difficult-to-classify voxels.

Some segmentation results are shown in Figure 3. Axial FLAIR MRI views of a patient with multiple sclerosis are presented with the segmentation mask obtained with our proposed algorithm overlayed. True positive are represented in red, false negative errors are presented in green and false positive errors in yellow.

Figure 3 (a)-(d) corresponds to images from patient 1 and (e)-(h) corresponds to images from patient 5. In Figure 3 (a) and (e) the parameters are $s = 0.0$ and $\gamma = 0.0$ and in Figure 3 (b) and (f) the parameters are $s = 0.0$ and $\gamma = 3.0$. It can be noted that in Figure 3 (b) there was a slight decrease in false negative errors which is attributable to the application of focal loss.

In Figure 3 (c) and (g) the results of automatic segmentation are presented using the parameters $s = 0.2$ and $\gamma = 0.0$ and in Figure 3 (d) and (h) $s = 0.2$ and $\gamma = 3.0$. In Figure 3 (d) and (h) an increase in true positive and a decrease in false negative errors and false positive errors with respect to all the previous combinations of parameters. This is attributable to the use of the stratified sampling mechanism and the proper of the γ parameter.

4 Conclusion and Future Work

This paper highlights the importance of an efficient design in the task of segmentation of multiple sclerosis lesions, where the application of a loss function that penalizes difficult voxels can improve the segmentation results.

In this work we empirically verified that the focal loss function requires an adequate representation of the difficult voxels near the edges and that the targeting of the penalty in these improves the results of the Dice similarity coefficient (Dice).

We also conclude that by improving the preprocessing stage it is possible to represent in a better way these difficult voxels near the edges of the lesions and thus reduce false positives and false negatives, where the best result obtained for the Dice similarity coefficient corresponded to $\text{Dice} = 0.7150(0.0094)$ with the parameters $s = 0.2$ and $\gamma = 3.0$, taking into account that for the medical image segmentation task a value greater than $\text{Dice} > 0.7$ is considered comparable to that obtained by another human expert.

The possibility of using non-uniform 3D sampling patches is also raised, which is expected to generate an increase in the results in the Dice similarity coefficient considering the 3D nature of the MRIs.

Acknowledgments

References

1. **Acar, Z. Y., Başçiftçi, F., Ekmekci, A. H. (2022).** A convolutional neural network model for identifying multiple sclerosis on brain FLAIR MRI. Sustainable Computing: Informatics and Systems, Vol. 35, pp. 100706. DOI: 10.1016/j.suscom.2022.100706.

2. **Birenbaum, A., Greenspan, H. (2017).** Multi-view longitudinal CNN for multiple sclerosis lesion segmentation. Engineering Applications of Artificial Intelligence, Vol. 65, pp. 111–118. DOI: 10.1016/j.engappai.2017.06.006.

3. **Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C. H., et al. (2017).** Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. NeuroImage, Vol. 148, pp. 77–102. DOI: 10.1016/j.neuroimage.2016.12.064.

4. **Chollet, F., et al. (2015).** Keras.

5. **Danelakis, A., Theoharis, T., Verganelakis, D. A. (2018).** Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. Computerized Medical Imaging and Graphics, Vol. 70, pp. 83–100. DOI: /10.1016/j.compmedimag.2018.10.002.

6. **Daoudi, K., Kersten, B. E., Vonk, M., de Korte, C. L. (2019).** High-frequency photoacoustic and ultrasound imaging of systemic sclerosis patients. Opto-Acoustic Methods and Applications in Biophotonics IV, pp. 1107707. DOI: 10.1117/12.2527050.

7. **Filippi, M., Bar-Or, A., Piehl, F., Preziosa, P., Solari, A., Vukusic, S., Rocca, M. A. (2018).** Multiple sclerosis. Nature Reviews Disease Primers, Vol. 4, No. 1. DOI: 10.1038/s41572-018-0041-4.

8. **Giorgio, A., Stefano, N. D. (2018).** Effective utilization of MRI in the diagnosis and management of multiple sclerosis. Neurologic Clinics, Vol. 36, No. 1, pp. 27–34. DOI: 10.1016/j.ncl.2017.08.013.

9. **Jenkinson, M., Bannister, P., Brady, M., Smith, S. (2002).** Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage, Vol. 17, No. 2, pp. 825–841. DOI: 10.1006/nimg.2002.1132.

10. **Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P. (2020).** Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, No. 2, pp. 318–327. DOI: 10.1109/TPAMI.2018.2858826.

11. **Nai, Y. H., Teo, B. W., Tan, N. L., O'Doherty, S., Stephenson, M. C., Thian, Y. L., Chiong, E., Reilhac, A. (2021).** Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset. Computers in Biology and Medicine, Vol. 134, pp. 104497. DOI: 10.1016/j.compbiomed.2021.104497.

12. **Ochoa-Montiel, R., Olague, G., Sossa, H. (2020).** Expert knowledge for the recognition of leukemic cells. Applied Optics, Vol. 59, No. 14, pp. 4448–4460. DOI: 10.1364/AO.385208.

13. **Ochoa-Montiel, R., Sossa, H., Olague, G., Chan Ley, M., Menendez, J. (2021).** Symbolic learning using brain programming for the recognition of leukemia images. Computación y Sistemas, Vol. 25, No. 4. DOI: 10.13053/cys-25-4-4045.

14. **Tovar-Corona, B., Flores-Alonso, S. I., Luna-García, R. (2022).** Convolutional neural network for improvement of heart valve disease detection. Computación y Sistemas, Vol. 26, No. 3. DOI: 10.13053/cys-26-3-4202.

15. **Ulloa, G., Veloz, A., Allende-Cid, H., Monge, R., Allende, H. (2022).** Efficient methodology based on convolutional neural networks with augmented penalization on hard-to-classify boundary voxels on the task of brain lesion segmentation. Pattern Recognition, pp. 338–347. DOI: 10.1007/978-3-031-07750-0_31.

16. **Valverde, S., Cabezas, M., Roura, E., González-Villa, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Àlex Rovira, Oliver, A., Lladó, X. (2017).** Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach. NeuroImage, Vol. 155, pp. 159–168. DOI: 10.1016/j.neuroimage.2017.04.034.

17. **Yeghiazaryan, V., Voiculescu, I. (2018).** Family of boundary overlap metrics for the evaluation of medical image segmentation. Journal of Medical Imaging, Vol. 5, No. 1. DOI: 10.1117/1.jmi.5.1.015006.

18. **Yılmaz Acar, Z., Başçiftçi, F., Ekmekci, A. H. (2022).** Future activity prediction of multiple sclerosis with 3d mri using 3d discrete wavelet transform. Biomedical Signal Processing and Control, Vol. 78, pp. 103–940. DOI: 10.1016/j.bspc.2022.103940.

19. **Zaleta Gallardo, C. A., Garca Herrera, M. F., Vera Acosta, A., Mendoza Flores, D. A., Vera Santos, O. A., Sanchez Avalos, J. M., Perez Careta, M. C., Perez Careta, E. (2018).** 3D Multiple Sclerosis Image Analysis Based on Probabilistic Methods for a 4F Array. DOI: 10.1364/FIO.2018.JTu2A.139.

20. **Zhou, T., Ruan, S., Canu, S. (2019).** A review: Deep learning for medical image segmentation using multi-modality fusion. Array, Vol. 3-4, pp. 100–004. DOI: 10.1016/j.array.2019.100004.