Classification of Active Multiple Sclerosis Lesions in MRI Without the Aid of Gadolinium-Based Contrast Using Textural and Enhanced Features from FLAIR Images

Paulo G. L. Freire[1], Marcos Hideki Idagawa[2], Enedina Maria Lobato de Oliveira[2], Nitamar Abdala[2], Henrique Carrete Jr.[2], and Ricardo J. Ferrari[1(✉)]

[1] Department of Computing, Federal University of São Carlos, Rod. Washington Luis, Km 235, São Carlos, SP 13565-905, Brazil
{paulo.freire,rferrari}@ufscar.br
[2] Federal University of São Paulo, Rua Pedro de Toledo, 650, São Paulo, SP 04039-002, Brazil
https://www.bipgroup.dc.ufscar.br

Abstract. Multiple sclerosis (MS) is an autoimmune demyelinating disease that affects one's central nervous system. The disease has a number lesion states. One of them is known as active, or enhancing, and indicates that a lesion is under an inflammatory condition. This specific case is of interest to radiologists since it is commonly associated with the period of time a patient suffers most from the effects of MS. To identify which lesions are active, a Gadolinium-based contrast is injected in the patient prior to a magnetic resonance imaging procedure. The properties of the contrast medium allow it to enhance active lesions, making them distinguishable from nonactive ones in T1-w images. However, studies from various research groups in recent years indicate that Gadolinium-based contrasts tend to accumulate in the body after a number of injections. Since a comprehensive understanding of this accumulation is not yet available, medical agencies around the world have been restricting its usage to cases only where it is absolutely necessary. In this work we propose a supervised algorithm to distinguish active from nonactive lesions in FLAIR images, thus eliminating the need for contrast injections altogether. The classification task was performed using textural and enhanced features as input to the XGBoost classifier on a voxel level. Our database comprised 54 MS patients (33 with active lesions and 21 with nonactive ones) with a total of 22 textural and enhanced features obtained from Run Length and Gray Level Co-occurrence Matrices. The average precision, recall and F1-score results in a 6-fold cross-validation for active and nonactive classes were 0.892, 0.968, 0.924 and 0.994, 0.987, 0.991, respectively. Moreover, from a lesion perspective, the algorithm misclassified only 3 active lesions out of 157. These results indicate our tool can be used by physicians to get information about active

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12250, pp. 60–74, 2020.
https://doi.org/10.1007/978-3-030-58802-1_5

MS lesions in FLAIR images without using any kind of contrast, thus improving one's health and also reducing the cost of MRI procedures for MS patients.

Keywords: Multiple sclerosis · Classification · Supervised learning · Texture · Enhancement

1 Introduction

Multiple sclerosis (MS) is an autoimmune demyelinating disease that attacks one's central nervous system and affects more than 2.3 million people worldwide [21]. The most common form of MS is known as relapsing-remitting and is characterized by episodes of neurological dysfunctions that evolve over a period of time before plateauing and remitting with some degree of recovery [17]. Each MS lesion can be classified into a specific class, namely preactive, active, chronic active and chronic inactive [15]. The active - or enhancing - state indicates that one or more lesions are under an inflammatory condition and it is commonly associated with the period of time a patient suffers most from the effects of MS.

To identify which lesions are active, a Gadolinium-based contrast is injected in the patient prior to the procedure itself. The properties of the contrast enhance active lesions [17], making them distinguishable from other kinds of pathologies in T1-w images. However, recent works [4,14,27] indicate that Gadolinium-based contrasts tend to accumulate in one's brain, bones, skin and other parts of the body after a number of injections.

A comprehensive understanding of this accumulation is not yet available, which has made several agencies, such as the U.S. Food and Drug Administration and the European Medicines Agency, issue statements [10,11,24–26] restricting the usage of Gadolinium-based contrasts only to cases where it is absolutely necessary. There are indications that infrequent contrast administrations pose no threat [16]; however, this is not the case for MS patients, who must undergo a magnetic resonance imaging (MRI) procedure with contrast injections from time to time for the rest of their lives in order to assess how the disease is progressing.

Some works in the literature have used supervised and unsupervised techniques to segment and classify different types of MS lesions [18]. Since it is a challenging task, it is common for techniques to use different features from MR images besides gray level intensity. In [20], the authors used texture analysis to distinguish normal appearing white matter (NAWM) from active MS lesions in T2-w images. They used features extracted from both Gray Level Co-occurrence Matrix (GLCM) and Run Length Matrix (RLM) to help differentiate between the two classes. Using linear discriminant analysis (LDA), partial least squares (PLS) and logistic regression (LR), they were able to get different sets of textural features and results according to each technique. The authors achieved the highest accuracy when using PLS in a 6-texture parameter model with a sensitivity of 0.88 and specificity of 0.81. According to the authors, a limitation of their study concerned the limited number of patients (a total of 21) and

active lesions (a total of 58), indicating that further investigations should be conducted in order to confirm their findings in a larger set of data. Another important aspect to consider is that the goal of the authors was to distinguish active MS lesions from NAWM instead of active from nonactive lesions. Though the correct distinction between NAWM and other MS pathologies is crucial, it is not a sufficient condition to withdraw the usage of Gadolinium-based contrast.

In [1], a study was conducted in order to distinguish between three classes of lesions: non-enhancing lesions (NEL), enhancing lesions (EL) and persistent black holes (PBH). The authors used more than 300 textural features from GLCM, RLM, absolute gradient and wavelets extracted from T2-w images. Their dataset comprised 90 patients of which 39 of them had NELs, 32 had PBHs and 19 had ELs. The authors used LDA to classify two classes at a time. In this sense, they verified that 18 features were significant to differentiate NELs from ELs, 14 features to separate NELs and PBHs and other 18 to distinguish ELs from PBHs. Their model yielded perfect classification for NEL vs. EL and EL vs. PBH and a sensitivity and specificity of 0.943 and 0.963, respectively, for NEL vs. PBH. These results are a strong indication that texture analysis can be successfully applied to tell different MS lesion types apart without contrast injection.

Finally, in [8] the authors presented a pipeline for detection and segmentation of MS lesions, including active ones. To this end, they used pre- and post-contrast T1-w images to train their model using a Markov Random Field and textural features. They detected candidate active voxels with a Conditional Random Field classifier. Then, higher order features modeling the texture of the lesion and surrounding tissue were calculated for the patches containing the candidates. They also included an option to use longitudinal information, when available, in order to leverage any differences between time-points. They trained their algorithm with 1760 scans acquired at 180 sites. Applying their model to a test set of 2770 patients led to a sensitivity of 0.91 and average false positive counts of 0.46. A good point made by the authors regarded that active lesions can be as small as 3–10 voxels. This obviously poses a difficulty for segmentation algorithms and must be given special attention in the context of MS. In this sense, textures and patches can mitigate the problem by factoring in the neighborhood information.

Given this scenario, in this work we propose a supervised algorithm to distinguish active from nonactive lesions in Gadolinium-free fluid attenuation inversion recovery (FLAIR) images, thus eliminating the need for contrast injections altogether. The classification task was performed on a voxel level on 54 MS patients - 33 with active lesions and 21 with nonactive ones - using the XGBoost classifier [7]. We used a total of 22 features, including textures extracted from Run Length Matrix and Gray Level Co-occurrence Matrix. The average precision, recall and F1-score results in a 6-fold cross-validation for active and nonactive classes were 0.892, 0.968, 0.924 and 0.994, 0.987, 0.991, respectively. Moreover, from a lesion (instead of voxel) classification perspective, the algorithm misclassified only 3 active lesions out of 157, meaning mistakes on the voxel level are tolerable to some extent when we look at lesions as a whole. These results indicate our tool

can be used by physicians to get information about active MS lesions in FLAIR images without using any kind of contrast, thus improving one's health and also reducing the costs of MRI procedures for MS patients.

This paper is divided as follows. In Sect. 2 we present and describe our active and nonactive patient databases, along with explanations of the image features extracted from FLAIR images and the XGBoost classifier; in Sect. 3, we present our results and discuss them; finally, in Sect. 4, we present our conclusions and lay out future works that can be derived from our findings.

2 Materials and Methods

In this section we present the databases used in this work, the preprocessing pipeline applied to the images, the textural features used in the classification step, the classifier itself and the metrics to assess the results.

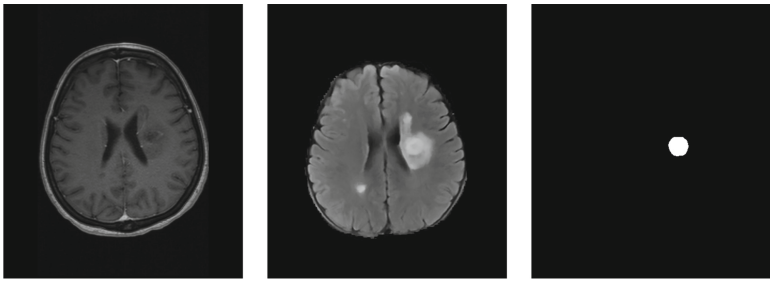
2.1 Databases

Nonactive Lesions. We used the training data of the Longitudinal MS Lesion Segmentation Challenge of the 2015 International Symposium on Biomedical Imaging [5] as our nonactive lesions database with scans from five patients and a total of 21 time-points (4 patients with 4 time-points and 1 patient with 5 time-points). The imaging sequences were adjusted to produce T1-w, T2-w, PD and FLAIR images, which were all skull-stripped, bias field corrected and registered to the same space. Each time-point included manual lesion annotations by two expert raters. Image dimensions were $181 \times 217 \times 181$ and voxel resolution was $1\,\text{mm}^3$.

Active Lesions. Our active lesions database came from the Demyelinating Diseases Outpatient Clinics - Neurology & Neurosurgery Department - Universidade Federal de São Paulo - (UNIFESP), Brazil. This database comprised scans of 33 patients with FLAIR and post-contrast T1-w images. An expert rater annotated active MS lesions in FLAIR using post-contrast T1-w as a reference for lesion location. Since this database existed prior to this work and the annotations were made on images as they were, 7 patients out of 33 had slightly different acquisition protocols regarding the number of slices per scan. While 26 patients were scanned with image dimensions $384 \times 512 \times 20$, the other 7 were scanned with image dimensions $384 \times 512 \times 25$. However, qualitatively speaking, the extra number of slices in this second group did not interfere with our region of interest (i.e., the brain itself). Voxel resolution was $0.44 \times 0.44 \times 6.5\,\text{mm}^3$ across all images. An example of a patient from this database is shown in Fig. 1.

2.2 Preprocessing

The preprocessing pipeline shown in Fig. 2 was applied to all FLAIR images from both databases.



(a) (b) (c)

Fig. 1. Patient with active lesions from the UNIFESP database. (a) T1, (b) FLAIR, (c) Active lesion mask annotated by an expert rater.

First, we applied a noise reduction step using the Non-local means technique [3] with $\sigma = 15$. This step is important because it helps mitigate intensity variations, thus creating a smoother brain profile across all images. After that, we applied the N4 algorithm [23] to reduce the bias field effect. Following that, a histogram matching [22] using two matching points was conducted to bring all images to a standard intensity domain. As a reference for this histogram matching we chose the FLAIR scan in the nonactive dataset with the heaviest lesion load (32,3 ml). The rationale behind this choice was to make lesions as representative as possible when matching histograms in both active and nonactive datasets. Finally, all images were rescaled to the $[0, 255]$ interval.

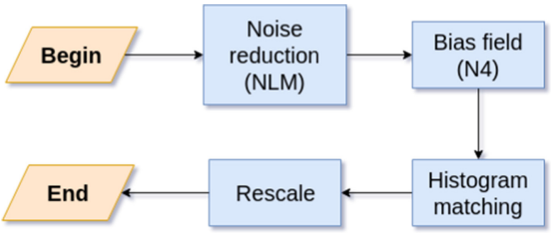


Fig. 2. Preprocessing pipeline.

2.3 Feature Extraction

After completing the preprocessing pipeline explained the previous section, we extracted features to aid in the distinction between active and nonactive lesions. Such features can be divided into two groups, as shown in Fig. 3.

The highlight branch was comprised of Sobel and two other feature images, enhanced and hyperintensity map, generated by the algorithm proposed in [12].

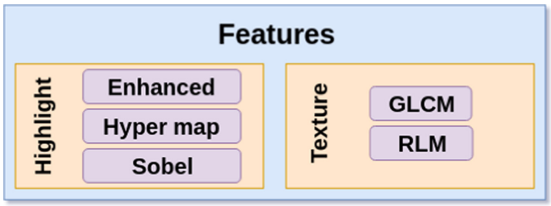


Fig. 3. Classes of extracted features.

In short, these enhanced and hyperintensity map feature images were created by first scattering the histogram in order to attenuate mean intensities and enhance "tail" ones, using it to further highlight hyperintensities through patch comparisons. An example of the images generated from this branch are shown in Fig. 4.

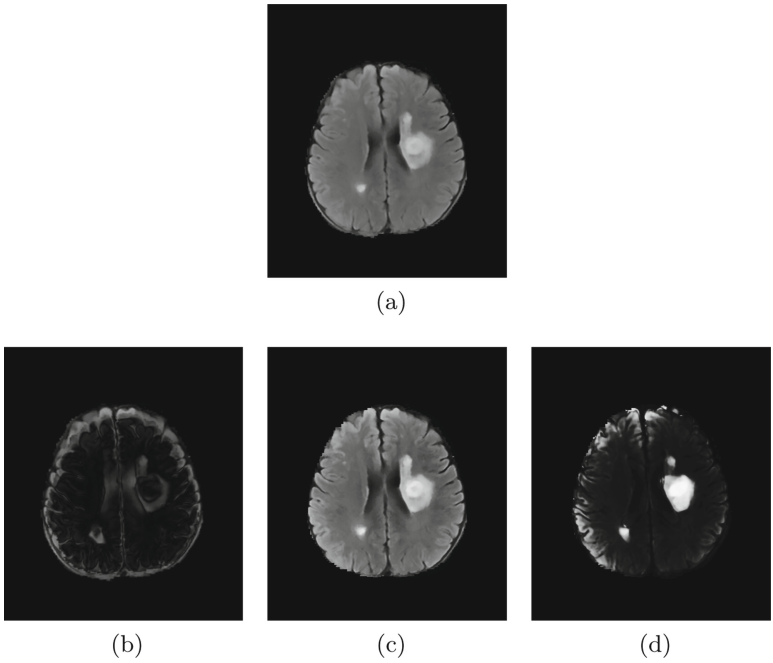


Fig. 4. Highlight features extracted from an active (a) FLAIR image yielding (b) Sobel image, (c) Enhanced image and (d) Hyperintensity map.

The other branch of feature extraction concerned textures and was divided into two techniques: GLCM and RLM. The former was first introduced by [13] and has been widely used in MRI segmentation and classification problems

[1,8,19,20]. The GLCM computes the joint probability of two voxels with the same, or co-occurring, gray level intensities within a distance d and direction θ. In this work, we set the distance to a radius of 2 and 13 directions, scaled analogously from the default 4 directions - 0, 45, 90 and 135° - used in 2D images. Each texture map was calculated as the average of all 13 directions. We extracted eight features from each GLCM, namely energy, entropy, correlation, inverse difference moment, inertia, cluster shade, cluster prominence and Haralick correlation. Similarly, RLM [6] counts the number of times two (or more) voxels with the same gray level intensity occur in a given direction. Features from RLM are mostly related to the fineness and coarseness of a given image represented by long runs and short runs, respectively. We extracted ten features from each RLM, namely short run emphasis (SRE), long run emphasis (LRE), grey level non uniformity (GLN), run length non uniformity (RLN), low grey level run emphasis (LGRE), high grey level run emphasis (HGRE), short run low grey level emphasis (SRLGE), short run high grey level emphasis (SRHGE), long run low grey level emphasis (LRLGE), long run high grey level emphasis (LRHGE).

An example of features extracted from GLCM and RLM are shown in Fig. 5.

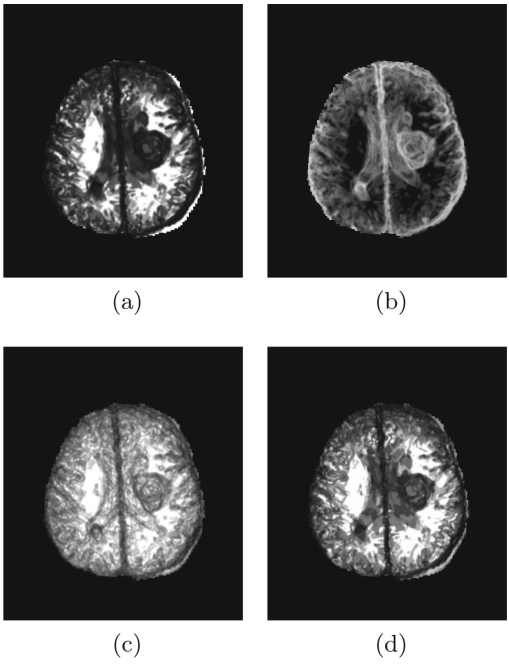


Fig. 5. Textures extracted from the patient with active lesions depicted in Fig. 4. (a) GLCM Energy, (b) GLCM Entropy, (c) RLM GLN, (d) RLM LRE.

After completing feature extraction, a total of 22 attributes were selected to be used in the classification step.

2.4 Classifier

The classifier used in this paper is the XGBoost [7], a gradient tree boosting algorithm. The main goal of boosting is to improve the accuracy of a given learning algorithm by using an ensemble of weak learners whose joint decision rule has an arbitrarily high accuracy on the training set [9]. We chose XGBoost due to its growing popularity in classification contests [2] and optimized structure that makes it scalable.

Mathematically, given a data set with n samples, indexed by $i = 1 \dots n$, and m features $D = \{(\mathbf{x}_i, y_i))\}$, where $|D| = n$, $\mathbf{x}_i \in \mathbb{R}^m$ and $y_i \in \mathbb{R}$, a tree ensemble model is comprised of the sum of K functions to predict the output

$$\hat{y}_i = \phi\left(\mathbf{x}_i\right) = \sum_{k=1}^{K} f_k\left(\mathbf{x}_i\right), \tag{1}$$

where $f_k \in \mathcal{F} = \left\{f\left(\mathbf{x}\right) = w_{q(\mathbf{x})}\right\}$ with $q : \mathbb{R}^m \to T$ and $w \in \mathbb{R}^T$ is the space of trees. In this context, q is related to the arrangement of each tree that maps a sample to its corresponding leaf index and T is the number of leaves in a tree. Therefore, each f_k is related to one particular tree q and leaf weights w.

In order to learn the functions that will describe the model itself, the following loss function is used as the objective

$$\mathcal{L}\left(\phi\right) = \sum_{i} l\left(\hat{y}_i, y_i\right) + \sum_{k} \Omega\left(f_k\right), \tag{2}$$

where

$$\Omega\left(f_k\right) = \gamma T + \frac{1}{2}\lambda \|w\|^2, \tag{3}$$

l is a differentiable loss function and Ω is a penalty term to avoid selecting complex trees. The goal is to keep the model simple and highly predictive at the same time. We can think of gradient boosting in the following way:

1. Fit a model to the data, $\mathcal{F}_1\left(\mathbf{x}\right) = y$.
2. Fit a model to the residuals (i.e., the loss), $h_1\left(\mathbf{x}\right) = y - \mathcal{F}_1\left(\mathbf{x}\right)$.
3. Create a new model, $\mathcal{F}_2\left(\mathbf{x}\right) = \mathcal{F}_1\left(\mathbf{x}\right) + h_1\left(\mathbf{x}\right)$.
4. Repeat steps (1)–(3) until a number M of trees or a sufficiently small difference between loss functions of subsequent steps is reached.

In other words, the algorithm trains successive component classifiers and the output for a test sample $\mathbf{x}$ is based on the outputs of these very same components. The scalability of XGBoost is achieved by analyzing and optimizing cache access patterns, data compression and sharding, which are described in more detail in [7]. Regarding parameters, we set the maximum depth of each tree to one and the number of estimators to 5000 in order to avoid specialization of learners and provide enough of them to get an accurate joint decision rule.

2.5 Metrics

To assess the classification of active and nonactive voxel lesions we used the precision, recall and F1-score metrics defined in Table 1.

Table 1. Metrics used to assess classification of active and nonactive voxel lesion. TP, FP and FN are true positives, false positives and false negatives, respectively.

Metric	Evaluation
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	$2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$

The ratio of active to nonactive lesions in our databases was roughly 1:4, i.e., the active voxel class had a representativity of approximately 25%. Given this imbalance and the paramount importance of correctly detecting active lesions, the recall and F1-score metrics bear more meaning to the assessment of classification performance. This is due to the fact that the former is related to the actual ground truth and the latter provides a better grasp on the overall performance - which can be distorted by the uneven distribution among both classes of interest.

We also conducted a lesion-level analysis. To do so, the active and nonactive voxels were grouped into connected components, thus indicating how many actual lesions were present in each patient. Given the classification probability of each lesion voxel, we calculated the average probability of any given lesion component being active or nonactive. Let C be a lesion component and p the probability of a given lesion voxel inside C. If the average probability of component C was equal to or greater than 50%, we ranked it as being active; otherwise, we ranked it as nonactive, as indicated in Eq. 4.

$$\text{Class} = \begin{cases} \text{Nonactive}, & \frac{1}{\|C\|}\sum p_{p\in C} < 0.5 \\ \text{Active}, & \text{otherwise.} \end{cases} . \tag{4}$$

3 Results and Discussion

In this section we present the voxel-level classification results based on precision, recall and F1-score, as well as the accuracy of our method on a lesion-level basis. Before discussing the results per se, two observations regarding the number of folds are deemed necessary.

First, we set the number of patients per fold to five as a trade-off between the size of the training and test sets in order to make the classifier general enough to avoid overfitting problems. However, since the number of patients with active lesions was 33, the last fold would comprise either 3 or 8 patients. We chose

the latter for the sake of generalization. The results using only 3 patients in the training set indicated a strong overfitting, especially for the active lesion class. Hence, we decided to include these 3 patients in the last 5-patient fold. Since this experiment design choice implied in less samples for the training set, thus making the classification task "harder", we decided it was a valid approach. This is the reason why the sixth fold included eight instead of five patients from each class.

Second, as mentioned in Sect. 2.1, the nonactive database was comprised of 5 patients, four of them with four time points and one with five. The rationale to choose which patients with nonactive lesions would go into each fold was to randomly choose one time point per patient when the fold size was five. For the last fold, with eight patients, we simply chose three additional random time points from the first three patients in the nonactive dataset (i.e., the last fold had two time points from the first three patients and one time point from the rest).

The voxel-level classification results are shown in Table 2. The nonactive classification yielded higher results when compared to the active class, especially regarding the precision metric. However, as mentioned in Sect. 2.5, recall and F1-score are more significant when it comes to analyze imbalanced datasets as the one we used in this work. In this sense, we can observe that these two metrics were similar between the active and nonactive classes. Moreover, we verified that most misclassifications occurred when a nonactive voxel was mistaken for an active one. Even though this is not ideal, this result is still better for a physician - who can discard these misclassifications when reviewing the outcome of our approach - than the other way around.

On a lesion-level perspective, 154 out of 157 lesions were correctly classified as active and the classification was perfect for the nonactive class. As mentioned in Sect. 2.5, we averaged the probability of each lesion as a whole and labeled them according to the majority class. We can see that even though the active class did not have precision, recall and F1-score as high as the nonactive class in a voxel-level analysis, the mistakes were compensated by the surrounding voxels comprising the lesion itself. In other words, the voxel-level classification provided a margin of safety for errors, making them less significant when viewed from a coarser granularity. Since the outcome of interest for neurologists and physicians is on the lesion-level, this margin indicates there is room for mistakes on a voxel-level that, to a certain extent, do not interfere on the lesion-level results.

Only one patient, ID 32, did not have their lesions identified as active. Some features from this particular patient are shown in Fig. 6. Visual analysis indicate that not only were the lesions small, they also had an intensity profile too similar to that of nonactive lesions. Moreover, the quality of the FLAIR image itself was rather poor. Even though textural and enhanced features aid in the distinction between both classes, there are cases where the classification step fails to yield the expected output due to conditions that are out our reach but yet have influence on the final outcome.

Table 2. Precision, recall and F1-score metrics for the 6-fold cross-validation.

Patients	Class	Results		
		Precision	Recall	F1-score
1–5	Active	0.945	1	0.972
	Nonactive	1	0.993	0.997
6–10	Active	0.7	0.996	0.823
	Nonactive	1	0.966	0.983
11–15	Active	0.765	1	0.867
	Nonactive	1	0.977	0.988
16–20	Active	0.99	1	0.995
	Nonactive	1	0.998	0.999
21–25	Active	0.887	0.863	0.875
	Nonactive	0.978	0.982	0.98
26–33	Active	0.999	0.958	0.978
	Nonactive	0.991	1	0.995
Weighted avg.	Active	0.892	0.968	0.924
	Nonactive	0.994	0.987	0.991

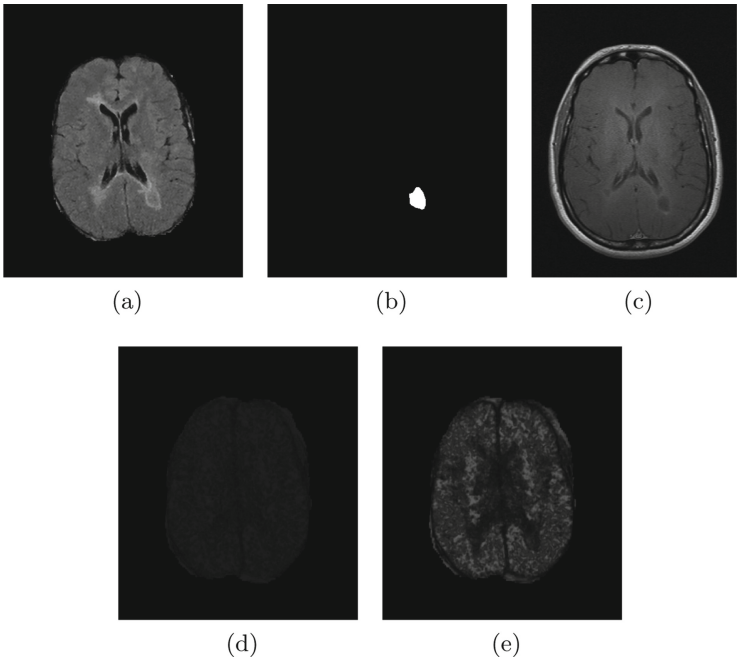
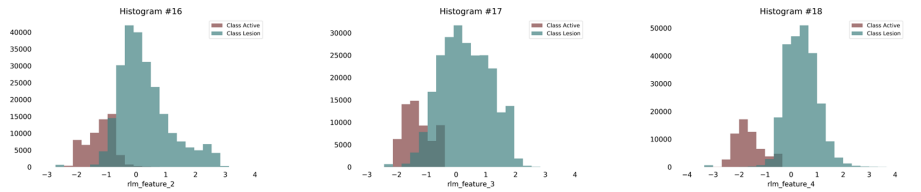


Fig. 6. Example of slices from patient ID 32 who had zero lesion hits. (a) FLAIR, (b) active lesion annotation, (c) T1 with Gadolinium enhancement, (d) RLM LRE, (e) RLM GLN.

We can also observe that the standard deviation across all lesions was quite small and the average lesion probability for most patients, both active and nonactive, was close or equal to one. It indicates our classifier was able to accurately distinguish both classes with a high degree of certainty, which is important when dealing with a sensitive healthcare scenario as this one.

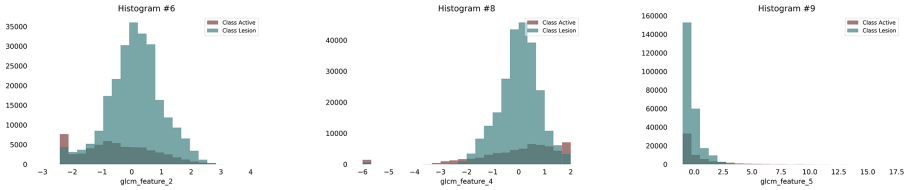
To analyze the effect of the features in the classification step, we plotted histograms for all 22 attributes extracted from all images of both active and nonactive datasets. Doing so we observed the RLM features RLE, RLN, GLN showed a significant distinction between both classes, whereas GLCM features, in general, had an overlay between one class and the other. Histograms from RLM and GLCM features are shown in Figs. 7 and 8.

The enhancement effect mentioned in Sect. 2.3 can be seen in histograms shown in Fig. 9. We can see that the enhanced and hyperintensity map features offset and spread the class histograms apart, making them more distinguishable. Extracting good features is of paramount importance for the supervised learning pipeline, since they have a direct effect on the classification step. Therefore, combining dissimilar and relevant textural and enhanced features creates a domain space that is adequate for a proper class separation. However, it is also relevant to note that features with significant overlays cannot be promptly discarded because their combination with other more key/distinct features can be useful for the classification step as a whole, especially when the overlay areas between these features are complementary (i.e., the overlay area between classes in one histogram is the most distinctive area in another).



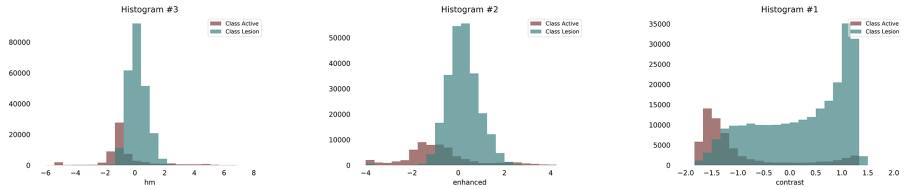
(a) (b) (c)

Fig. 7. RLM feature histograms: (a) LRE, (b) RLN and (c) GLN. Class "Lesion" represents the nonactive lesions.



(a) (b) (c)

Fig. 8. GLCM feature histograms: (a) Entropy, (b) Inverse Difference Moment and (c) Inertia. Class "Lesion" represents the nonactive lesions.



(a) (b) (c)

Fig. 9. Enhanced feature histograms: (a) Histogram matched (no enhancement), (b) Enhanced and (c) Hyperintensity Map. Class "Lesion" represents the nonactive lesions.

Overall, the classification accuracy on a lesion-level for the active class was over 98%, indicating that it is possible to identify this particular kind of MS lesion using only FLAIR images and without any kind of contrast agent. In addition, RLM features proved to be the most distinctive ones in this scenario. Performing the classification step on a voxel-level provided us with enough samples to properly train our classifier and also allowed for a margin of safety to take place, "smoothing out" the effects of misclassifications on a lesion-level.

4 Conclusions

This paper presented a supervised classification pipeline to distinguish between active and nonactive MS lesions in FLAIR images without using any contrast agent based on textural and enhanced features to create a feature space where both classes could be easily distinguished using the XGBoost classifier. Our dataset comprised 54 patients, 33 with active lesions and 21 with nonactive ones, with varying lesion loads, shapes, sizes and locations.

Though active lesions represented a small portion of samples when compared to nonactive ones, the results indicate that is possible to tell these two classes apart using information from FLAIR images and without the aid of any sort of contrasts. Out of the 157 active lesions, 154 were correctly identified as so. This is relevant because it unburdens patients from the accumulation of heavy metal based contrasts in their systems, makes it possible for patients with kidney problems to get an assessment of their active lesion progress (since they cannot be injected with contrasts in the first place) and also lowers the cost of the MRI procedure with the elimination of contrast shots.

There are limitations in this work. A larger number of patients would be required to assess the classification performance on a broader set of lesions. Also, lesions were not rated by an independent neuroradiologist, and that could have introduced bias in our study. And finally, we lacked an external validation set, which would make our data more robust. Nonetheless, we were able to show that is very possible to correctly identify active MS lesions without using Gadolinium-based contrasts.

We expect to further investigate these findings by conducting a feature selection and analysis to verify which of them are the most relevant for the classifica-

tion step and which can be discarded without affecting accuracy. We also intend to create a fully automatic pipeline to segment MS lesions and classify them into active and nonactive without any kind of contrast agent.

Acknowledgements. This work was supported by the São Paulo Research Foundation - FAPESP (grant numbers 2016/15661-0 and 2018/08826-9).

References

1. Ardakani, A.A., Nabavi, S., Farzan, A., Najafabad, B.: Quantitative MRI texture analysis in differentiating enhancing and non-enhancing T1-hypointense lesions without application of contrast agent in multiple sclerosis. Czech and Slovak Neurol. Neurosurg. **113**(6), 700–707 (2017)
2. Bennett, J., Lanning, S.: The Netflix prize. In: KDD Cup and Workshop in Conjunction with KDD (2007)
3. Buades, A., Coll, B., Morel, J.-M.: A non-local algorithm for image denoising. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 60–65 (2005)
4. Burke, L., Ramalho, M., AlObaidy, M., Chang, E., Jay, M., Semelka, R.: Self-reported gadolinium toxicity: a survey of patients with chronic symptoms. Magn. Reson. Imaging **34**, 1078–1080 (2016)
5. Carass, A., et al.: Longitudinal multiple sclerosis lesion segmentation: resource and challenge. NeuroImage **148**, 77–102 (2017)
6. Castellano, G., Bonilha, L., LM, L., Cendes, F.: Texture analysis of medical images. Clin. Radiol. **59**(12), 1061–1069 (2004)
7. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2016, pp. 785–794. ACM, New York (2016)
8. Doyle, A., Elliott, C., Karimaghaloo, Z., Subbanna, N., Arnold, D.L., Arbel, T.: Lesion detection, segmentation and prediction in multiple sclerosis clinical trials. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 15–28. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_2
9. Duda, O.R., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience, Hoboken (2000)
10. European Medicines Agency: EMA reviewing gadolinium contrast agents used in MRI scans (2016). http://bit.ly/EuropeanMedicinesAgency-2016. Accessed 02 Apr 2017
11. European Medicines Agency: EMA's final opinion confirms restrictions on use of linear gadolinium agents in body scans, July 2017. http://bit.ly/EuropeanMedicinesAgency2017. Accessed 01 July 2019
12. Freire, P.G.L., Ferrari, R.J.: Multiple sclerosis lesion enhancement and white matter region estimation using hyperintensities in FLAIR images. Biomed. Signal Process. Control **49**, 338–348 (2019)
13. Haralick, R., Shanmugam, K., Dinstein, I.: Textural feature for image classification. IEEE Trans. Syst. Man Cybern. **3**(6), 610–621 (1973)
14. Hu, H.H., Pokorney, A., Towbin, R.B., Miller, J.H.: Increased signal intensities in the dentate nucleus and globus pallidus on unenhanced T1-weighted images: evidence in children undergoing multiple gadolinium MRI exams. Pediatr. Radiol. **46**(11), 1590–1598 (2016). https://doi.org/10.1007/s00247-016-3646-3

15. Jonkman, L., et al.: Can MS lesion stages be distinguished with MRI? A post-mortem MRI and histopathology study. J. Neurol. **262**(4), 1074–1080 (2015). https://doi.org/10.1007/s00415-015-7689-4
16. Kromrey, M., et al.: Intravenous injection of gadobutrol in an epidemiological study group did not lead to a difference in relative signal intensities of certain brain structures after 5 years. Eur. Radiol. **27**(2), 772–777 (2016). https://doi.org/10.1007/s00330-016-4418-z
17. Lewis, P.A., Spillane, J.E.: Chapter 7 - Multiple Sclerosis. In: Lewis, P.A., Spillane, J.E. (eds.) The Molecular and Clinical Pathology of Neurodegenerative Disease, pp. 221–251. Academic Press (2019)
18. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
19. Loizou, C.P., Petroudi, S., Seimenis, I., Pantziaris, M., Pattichis, C.S.: Quantitative texture analysis of brain white matter lesions derived from T2-weighted MR images in MS patients with clinically isolated syndrome. J. Neuroradiol. **2015**(42), 99–114 (2014)
20. Michoux, N., Guillet, A., Rommel, D., Mazzamuto, G., Sindic, C., Duprez, T.: Texture analysis of T2-weighted MR images to assess acute inflammation in brain MS lesions. PLoS One **10**(12), e0145497 (2015)
21. Multiple Sclerosis International Federation: 2013 Atlas of MS (2013). http://www.msif.org/about-us/advocacy/atlas/. Accessed 01 Dec 2015
22. Nyul, L.G., Udupa, J., Zhang, X.: New variants of a method of MRI scale standardization. IEEE Trans. Med. Imaging **19**(2), 143–150 (2000)
23. Tustison, N., Gee, J.: N4ITK: Nick's N3 ITK implementation for MRI bias field correction. Penn Image Computing and Science Laboratory (2009)
24. U.S. Food and Drug Administration: FDA Drug Safety Communication: FDA evaluating the risk of brain deposits with repeated use of gadolinium-based contrast agents for magnetic resonance imaging (MRI) (2015). http://www.fda.gov/Drugs/DrugSafety/ucm455386.htm. Accessed 07 May 2018
25. U.S. Food and Drug Administration: FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings, May 2017. https://www.fda.gov/media/109825/download. Accessed 01 July 2019
26. U.S. Food and Drug Administration: Update on FDA approach to safety issue of gadolinium retention after administration of gadolinium-based contrast agents, September 2018. https://www.fda.gov/media/116492/download. Accessed 01 July 2019
27. Beomonte Zobel, B., Quattrocchi, C.C., Errante, Y., Grasso, R.F.: Gadolinium-based contrast agents: did we miss something in the last 25 years? Radiol. Med. **121**(6), 478–481 (2015). https://doi.org/10.1007/s11547-015-0614-1