

[dx.doi.org/10.17488/RMIB.41.3.3](https://doi.org/10.17488/RMIB.41.3.3)

E-LOCATION ID: 1050

Probabilistic Multiple Sclerosis Lesion Detection using Superpixels and Markov Random Fields

Detección Probabilística de Lesiones de Esclerosis Múltiple usando Superpíxeles y Campos Aleatorios de Markov

Alejandro Reyes, Alfonso Alba, Martín O. Méndez, Edgar R. Arce-Santana, Ildelfonso Rodríguez-Leyva

Universidad Autónoma de San Luis Potosí

ABSTRACT

Multiple Sclerosis (MS) is the most common neurodegenerative disease among young adults. Diagnosis and monitoring of MS is performed with T2-weighted or T2 FLAIR magnetic resonance imaging, where MS lesions appear as hyperintense spots in the white matter. In recent years, multiple algorithms have been proposed to detect these lesions with varying success rates, which greatly depend on the amount of a priori information required by each algorithm, such as the use of an atlas or the involvement of an expert to guide the segmentation process. In this work, a fully automatic method that does not rely on a priori anatomical information is proposed and evaluated. The proposed algorithm is based on an over-segmentation in superpixels and their classification by means of Gauss-Markov Measure Fields (GMMF). The main advantage of the over-segmentation is that it preserves the borders between tissues, while the GMMF classifier is robust to noise and computationally efficient. The proposed segmentation is then applied in two stages: first to segment the brain region and then to detect hyperintense spots within the brain. The proposed method is evaluated with synthetic images from BrainWeb, as well as real images from MS patients. The proposed method produces competitive results with respect to other algorithms in the state of the art, without requiring user assistance nor anatomical prior information.

KEYWORDS: Multiple Sclerosis; Lesion Detection; Superpixels; GMMF, Image Segmentation

RESUMEN

La Esclerosis Múltiple (MS) es una de las enfermedades neurodegenerativas más comunes en adultos jóvenes. El diagnóstico y su monitoreo se realiza generalmente mediante imágenes de resonancia magnética T2 o T2 FLAIR, donde se observan regiones hiperintensas relacionadas a lesiones cerebrales causadas por la MS. En años recientes, múltiples algoritmos han sido propuestos para detectar estas lesiones con diferentes tasas de éxito las cuales dependen en gran medida de la cantidad de información a priori que requiere cada algoritmo, como el uso de un atlas o el involucramiento de un experto que guíe el proceso de segmentación. En este trabajo, se propone un método automático independiente de información anatómica. El algoritmo propuesto está basado en una sobresegmentación en superpíxeles y su clasificación mediante un proceso de Campos Aleatorios de Markov de Medidas Gaussianas (GMMF). La principal ventaja de la sobresegmentación es que preserva bordes entre tejidos, además que tiene un costo reducido en tiempo de ejecución, mientras que el clasificador GMMF es robusto a ruido y computacionalmente eficiente. La segmentación propuesta es aplicada en dos etapas: primero para segmentar el cerebro y después para detectar las lesiones en él. El método propuesto es evaluado usando imágenes sintéticas de BrainWeb, así como también imágenes reales de pacientes con MS. Con respecto a los resultados, el método propuesto muestra un desempeño competitivo respecto a otros métodos en el estado del arte, tomando en cuenta que éste no requiere de asistencia o información a priori.

PALABRAS CLAVE: Esclerosis Múltiple; Detección de Lesiones; Superpíxeles; GMMF; Segmentación de Imágenes

Corresponding author

TO: Alfonso Alba

INSTITUTION: Universidad Autónoma de San Luis Potosí

ADDRESS: Facultad de Ciencias, UASLP.

Av. Chapultepec #1570, Col. Privadas del Pedregal,

C. P. 78295, San Luis Potosí, San Luis Potosí, México

E-MAIL: alfonso.alba@uaslp.mx

Received:

28 March 2020

Accepted:

23 August 2020

INTRODUCTION

Neurodegenerative diseases are one of the most critical issues for the health sector. Not only elderly people are the most affected by neurodegenerative diseases, but also young people can suffer from them. Multiple sclerosis (MS) is a neurodegenerative disease that mainly affects people between 20 and 40 years old, with high incidence in the general population. In fact, it is the second in incidence, epilepsy being the first^[1]. The cause of MS is controversial but seems to depend on genetic and environmental factors and may also have a strong auto-immune component. The diagnosis and prognosis are well established nowadays by neurologists. The symptoms are described by the patient, and evidence is found through physical examination. In clinics, the neurologist verifies the symptoms of the patient (for example weakness, blurred vision, ataxia, etc.), and then requests a brain imaging study. Magnetic resonance imaging (MRI) is highly recommended in this case, with the most common protocols for this purpose being: T1-w, T2-w, Proton Density (PD) and Fluid Attenuated Inversion Recovery (FLAIR). In brain images, the neurologist manually annotates MS lesions, which in FLAIR images are shown as high-intensity spots on the white matter. It is important to mention that manual annotation and counting of hyper-intense spots is often an extensive and tedious process because the clinician needs to check dozens of images and may find several spots for a single patient. Hence, there is a large interest in designing algorithms that can automatically detect MS lesions or assist the expert during the process^{[2][3]}. In the past decades, various methods to segment MS lesions in MRI images have been published; however, some of these methods suffer from low accuracy or have such a large number of parameters to tune that they are not user-friendly. Other approaches rely heavily on the user's participation, or need atlas databases, requiring additional preprocessing time (for instance, to align the atlas with the input images). For these reasons, it is interesting to develop fully automatic and user-friendly methods to aid in the detection of MS lesions.

There are a few reviews of methods for MS lesion segmentation in the literature^{[4][5]}. Some of those methods are based on probabilistic approaches^{[6][7]}, support vector machines (SVM)^[8], region growing^[9], K-Nearest Neighbors [10] or neural networks^[11], while some methods may also use additional prior information such as atlases^[12] or clinical information^[13]. Many of these methods use pre-processing steps to prepare the input images for MS lesion detection, such as image denoising and non-uniformity correction. Also, since some non-brain tissues such as scalp and optic nerve are also shown with high intensity on T2-w images, a skull stripping step is often required; to this end, several methods use the Brain Extraction Tool (BET)^[3]. Among the algorithms for MS lesion detection, several methods are based on Expectation-Maximization (EM) to segment MS lesions^{[14][15]} due to its good accuracy and easy implementation. In the EM-based method proposed by Garcia-Lorenzo *et al.*^[15], the algorithm is divided in three steps. In the first and second steps, there is a non-uniformity correction of the input image, followed by a skull stripping process. Finally, in the last step, the MS lesion detection is applied using clinical rules to select potential regions with good results (reported Specificity of 0.9954). Another interesting method, which uses Markov Random Fields (MRF), is proposed by Khayati *et al.*^{[6][7]}. They developed an MS lesion detector by estimating a conditional probability density function for each class, which was trained using the adaptive mixtures method (AMM). For validating their results, in^[6] they use a cross-validation approach where the first MS reader was used as the gold standard, leading to very accurate results since they achieved an average Dice Similarity Coefficient (DSC) of 0.75^{[16][17]}. Other proposal was developed by Lao *et al.*^[18], where they first perform affine registration of T1-w, T2-w PD and FLAIR images by maximization of the mutual information^[19] and skull stripping based on affine registration using the BET algorithm^[3]. In the training process, the proposed method combines T1-w, T2-w, PD and FLAIR in an

attribute vector (AV) for each voxel, along with the information of the neighboring voxels. A set of manually segmented scans is used to train a support vector machine (SVM) using the AdaBoost method [8] [20]; once trained, the SVM outputs, for each voxel, a scalar measure of abnormality that is binarized by applying a tuned threshold to discriminate lesions from normal tissue. Although similarity values (e.g., Dice index) were not reported, specificity and sensitivity look very promising which are also complemented by good visual results. Finally, in [21] the proposed method is based on Artificial Neural Networks (ANN), and the implemented software is freely available online at <https://med.inria.fr/the-app/downloads>. After selecting the segmentation option in this interface, the user can upload data and interactively click on the MS lesion to enhance them. One disadvantage of this software is that sometimes the algorithm segments the entire white matter region, particularly when MS lesions have blurred borders. Despite the diversity of algorithms for MS lesion detection, it is difficult to find one that fulfills the requirements of public health institutions, particularly when images are acquired with low resolution, few slices, and using a single modality to reduce costs. These requirements include border preservation, robustness to noise and blurred borders, good accuracy with single-modality (T2-w or T2-FLAIR) images, a reduced number of tuning parameters, and an implementation that does not rely on atlases or databases.

In this paper, a new method for automatic MS lesion segmentation is proposed. The method works with single-modality low resolution images, it does not need prior information related to anatomical structures or user annotations, and it only requires a few parameters to be tuned. In order to achieve good border preservation and robustness, an over-segmentation in superpixels (SPs) is performed as the first step [22], followed by a post-processing stage to achieve connectedness and eliminate spurious SPs produced by noise. Each SP

is then classified by means of a Gauss-Markov Measure Field (GMMF) model [23]. This segmentation approach is applied twice: first to isolate the brain region, and a second time to segment MS lesions in the brain. This paper is organized as follows: Section 1 contains the introduction and a description of the goals of this work. Section 2 presents the details of the proposed segmentation algorithm, which is called SP-GMMF, and the methodology for MS lesion segmentation. Results from applying the proposed method to the analysis of synthetic and real MR images are shown in Section 3, with a comparison against a state-of-the-art algorithm based on Expectation Maximization (EM). Finally, the conclusions of this work are presented in Section 4.

MATERIALS AND METHODS

In this proposal, segmentation of MS lesions is achieved in two stages: the first stage is to isolate the region of interest (in this case, the brain) from the rest of the image (skull, meninges, or bone cavities); then, in the second stage, MS lesions within the brain region are detected. Each stage employs a segmentation algorithm. The segmentation method proposed here is a combination of an over-segmentation of the image in superpixels using the Simple Linear Iterative Clustering (SLIC) [22] and a probabilistic labeling of the superpixels by means of the GMMF [23]. Besides the main steps of the segmentation process, there are other pre and post-processing steps that are important for the efficiency of the proposed algorithm. The details of the complete algorithm are described below.

The SLIC method is a clustering algorithm that combines spatial location and intensity information to subdivide an image in a relatively small number of groups of pixels that have similar color and are spatially coherent, commonly called superpixels [22] [24] [25]. This algorithm is a variant of the k-means algorithm that uses a reduced search space to form each SP, and whose main advantages are high computational speed

and border-preserving SPs. The SLIC method works as follows: let us define $l(\vec{r})$ as the image over the lattice L (that is, $l(\vec{r}) \in L$), and each superpixel is defined as $S_k = [C_k, D_k]$ where C_k and D_k are the average intensity and the geometric center of the superpixel, respectively. Indexes of superpixels are denoted by k , that is $k = 1, 2, \dots, K$, with K as the number of superpixels over L . The initialization of each D_k is given by a regular hexagonal grid in L , and C_k as the intensity at pixel site C_k . For each superpixel S_k , a square neighborhood of size $2M \times 2M$ is defined with center at D_k and $M = \sqrt{|L|/K'}$, where K' is the desired number of SPs given by the user; as a rule of thumb, this parameter can be defined as the total number of pixels in the image divided by the area (in pixels) of the smallest lesion that is observed; for instance, we are using $K' = 3000$ SPs for all images in this study. The actual number of superpixels K will vary across all the steps in the algorithm. Only those pixels that belong to the neighborhood of D_k can be assigned to S_k according to the distance measure:

$$\delta_k(\vec{r}) = \sqrt{m^{-2}(C_k - l(\vec{r}))^2 + \left(\frac{\gamma^2}{M^2}\right) \|D_k - \vec{r}\|^2} \quad (1)$$

where the first term measures the intensity distance, and it is normalized by the dynamic range of the data m , which can be computed as $m = \max(l(\vec{r})) - \min(l(\vec{r}))$. The second term is the spatial distance term and it is normalized by the size of the neighborhood. Finally, γ is a hyperparameter that weighs the importance between both terms. Once every pixel is assigned to some S_k , C_k and D_k are updated with the average intensity and spatial position of the pixels that belong to the k -th SP, and δ_k is computed again; this process is iterated until convergence. In our experience, the algorithm converges in 5 to 10 iterations. Although SPs adhere well to borders, for an adequate choice of γ , the SLIC algorithm can sometimes produce fragmented superpixels, which may lead to unconventional neighborhoods for the Markovian Fields process [23]. For that

reason, a relabeling process by means of connected component algorithm is applied [26], so that each connected fragment of a fragmented SP is considered as a new, individual superpixel. Once the relabeled field has been obtained, small SPs which are often due to noise, are fused to the most similar (in terms of average intensity) neighboring superpixel. In our case, a superpixel is considered too small when its area is less than 3% of the SP average area M^2 ; the 3% threshold was found experimentally as the percentage for which the number of SPs after fusion approximates better the number of desired SPs K' . With this fusion, not only the number of variables, but also the noise was reduced. Once the SPs are obtained, the next step for the segmentation process is to classify them with respect to their intensity. For that purpose, we consider that each class is defined by a Gaussian distribution of pixel intensities with a given mean and variance. Note that, although we aim for a binary classification at each stage (brain vs non-brain in the first stage, lesion vs non-lesion in the second stage), there are several types of tissues represented in the images; for this reason, multiple classes must be considered using the GMMF model [23], the goal is to estimate, for each SP S_k , the probability $p_j(S_k)$ that it belongs to class j , for each $j = 1, \dots, C$, where C is the number of classes. Under this model, one can obtain the probability field for each class j by minimizing the following energy function U given by:

$$U(p_j) = \sum_{k=1}^K (p_j(k) - g_j(k))^2 + \lambda \sum_{k=1}^K \sum_{i \in S_k} (p_j(k) - p_j(i))^2 \quad (2)$$

where the first term is a data term which enforces $p_j(k)$ to be similar to the normalized likelihood between the k -th superpixel and the j -th class, which is given by $g_j(k) = (v_j(k))/\sum_i v_i(k)$, where $v_j(k)$ is a likelihood function which measures how well the k -th SP fits in class j . Assuming that classes follow a Gaussian distribution, the likelihood can be obtained as:

$$v_j(k) = \frac{1}{\sqrt{2\pi} \sigma_j} \exp\left\{-\kappa(C_k - \mu_j)^2 / 2\sigma_j^2\right\} \quad (3)$$

where μ_j and σ_j are the mean and variance of the j -th class, respectively. Notice that the μ can be automatically initialized by the k-means method. Additionally, there is a hyper-parameter κ which controls the overall variance of all classes. The second term is a regularization term that promotes the similarity in the neighborhood. The neighborhood N_k for the k -th superpixel can be obtained, first, inspecting the borders of the image of superpixel labels and then inspecting the image of labels; i.e., when a border is detected, the labels in that point (SP) are added to its SP neighborhood. Finally, λ is a hyper-parameter that weighs the importance between terms. To solve (2), one can calculate its derivative and equal it to zero to obtain a linear equation system, which can be iteratively solved by the Gauss-Seidel method where the solution of $p_j(k)$ is given by:

$$p_j(k) = \frac{g_j(k) - \lambda \sum_{i \in N_k} p_j(i)}{1 + \lambda |N_k|} \quad (4)$$

After each Gauss-Seidel iteration, the means and variances of all classes are updated with a forgetting factor $\alpha = 0.2$ (i.e. $\mu_t \rightarrow (1-\alpha) \mu_{t-1} + \alpha \mu_t$, where t is the current iteration). This method generally stops until convergence is obtained, but in our experience, it converges from 5 to 10 iterations. Once convergence is achieved, each superpixel S_k is assigned to the class given by: $c(S_k) = \arg \max_j p_j(k)$. Figure 1 shows the flow chart of the SP-GMMF Segmentation proposal.

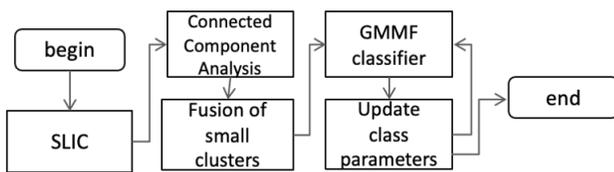


FIGURE 1. Block diagram of the proposed method SP-GMMF.

MS Lesion Classification

The SP-GMMF segmentation proposal takes advantages from SLIC and GMMF methods to implement a general-purpose segmentation method. In the proposed two-stage algorithm for MS lesion detection, which is illustrated in Figure 2, the first stage is oriented to automatically isolate the parenchyma region from the rest of the image. For that purpose, this proposal takes the SP-GMMF segmentation $c(S_k)$ of a T2-weighted or T2-FLAIR MR brain image (Figures 2a and 2b) to obtain a binary mask that represents the isolated brain. The fusion of small superpixels into bigger ones is depicted in Figure 2c, where some of the smaller SPs (circled in green) were fused with a neighboring larger SP, and the final segmentation is shown in Figure 2d.

It is well known that in several slices of an MRI sequence, the brain region and the background have the largest regions in almost every MRI slice. Under this assumption and knowing that the background has the lowest intensity in the MRI image, the brain mask will be obtained by choosing the largest area that does not belong to the lowest-intensity class. Notice that, in the case of the axial images, this does not always occur because the hemispheres might appear disconnected from each other. For example, in some supratentorial MRI Axial images the ratio of areas between hemispheres will be closest to one because their areas are similar (i.e. a ratio threshold above 0.7 may indicate that areas are similar). On the other hand, in some supratentorial and most infratentorial MRI images where the hemispheres are joined into a single region, the next largest region will be another structure whose area will be significantly smaller than the largest one (with a ratio less than 0.7), and thus the algorithm will consider it as part of the brain. Once the brain area mask has been obtained, a hole-filling algorithm is applied to it in order to recover any dark structure within the brain that might have been discarded by the previous operation.

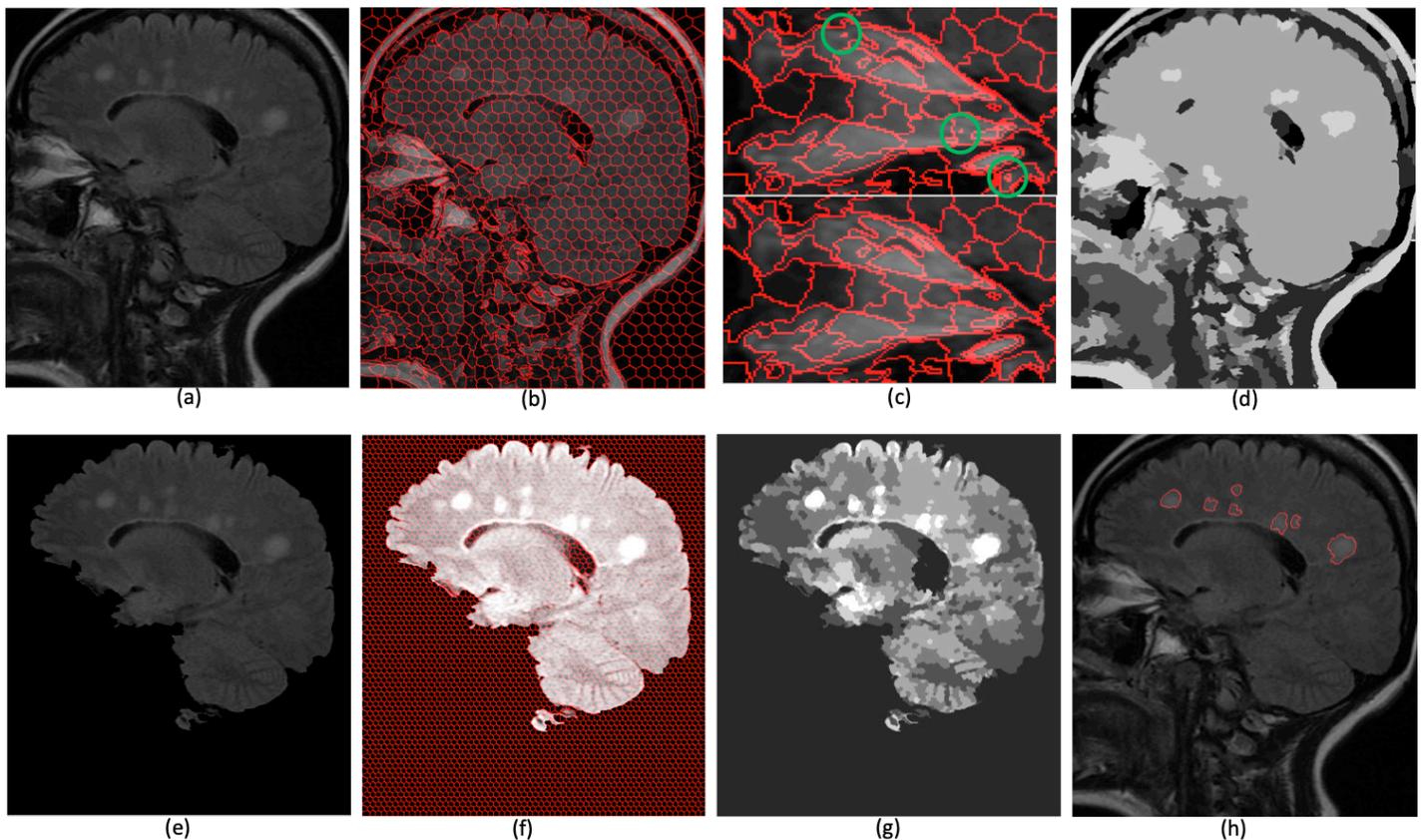


FIGURE 2. Complete process for detection of lesions in the brain. (a) Sagittal MRI input image of 520x459 pixels, (b) image segmented in superpixels with $\gamma = 0.1$ and 1000 desired clusters (resulting in $K = 972$ clusters), (c) Zoomed region where small clusters are shown circled in green (top), result of the fusion process where small regions were merged (bottom), (d) Result of GMMF segmentation with 10 iterations, $\lambda = 0.1$ and $k = 0.1$, (e) Isolated brain area, (f) Segmentation of the intensity-adjusted brain region in 5000 super-pixels, resulting in $K = 5453$ clusters after connected component were found, (g) GMMF segmentation of the brain region, (h) final result detecting 7 lesions in the white matter with eccentricity less than 0.9 and area of at least 80.

This mask, applied to the MRI image, automatically isolates the brain for the rest of the image (Figure 2e). On the other hand, MS lesions appear in T2 images as hyperintense spots; for this reason, an intermediate step is to apply a contrast-enhancing intensity adjustment to the isolated brain image to saturate the hyperintensities. This operation scales the voxel intensities by a factor such that the lowest 3% percentile will be saturated to zero, and the highest 3% percentile will be saturated to 255 (in an 8-bit grayscale image), which will facilitate the MS lesion detection for the next segmentation procedure.

In the second stage, the SP-GMMF segmentation is applied to the intensity adjusted brain image with a finer resolution of the SPs; that is, a larger desired number of SPs K' (Figures 2f and 2g). To obtain the binary mask that contains the MS lesions, the algorithm now isolates the regions that correspond to the highest intensity class on the segmentation and labels them using a connected component analysis, from which the area and eccentricity of each spot can be also obtained, as well as the number of potential lesions found in the image. Finally, a filtering procedure designed to reduce the number of false-positives

is applied, in which hyper-intense spots are reported as MS lesions (Figure 2g) only if they fulfill the following criteria: they must not be located in the external border of the brain (which corresponds to gray matter), their area must be sufficiently large (approximately 30 pixels) and they must not be too eccentric (maximum eccentricity of 0.85).

Roughly speaking, each stage of the process performs a binary segmentation of the image: in the first stage (brain peeling) the interest is to distinguish between background/non-background pixels, whereas in the second stage the interest is to distinguish between lesion and non-lesion pixels. However, at each stage a larger number of classes is considered so that a multimodal distribution of one of the binary classes (in this case non-background and non-lesion) can be modeled as the superposition of multiple classes.

EM* Classification

A revision of the state-of-the-art algorithms for MS lesion detection shows that a number of methods are based on the popular Expectation Maximization (EM) algorithm [4]. The main reason for choosing EM is due to a good balance between simplicity, popularity and good results in this task. In general, this approach assumes that intensities belonging to the structures on MRI images follow a Rician distribution that can be fairly approximated by a Gaussian distribution [13] [14].

Therefore, each image contains a finite mixture of Gaussian distributions and thus the goal is to find the parameters that define these distributions and their proportions in the mixture. Given the number of Gaussian mixture components T , and their respective parameters $\theta_j = (\mu_j, \sigma_j)$ as well as their weight in the mixture β_j , for each class $j = 1, \dots, T$, the intensity distribution for any pixel in the image can be expressed as:

$$p(l(\vec{r})|\theta) = \sum_{j=1}^T \beta_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(l(\vec{r}) - \mu_j)^2}{2\sigma_j^2}\right\} \quad (5)$$

In the first step, all the proportions β_j are equal to $1/T$, corresponding to a uniform distribution of classes; then, the parameters of the distributions θ_j are computed, which is called the Expectation step. After that, the Maximization step is applied, which consists in fixing the θ_j and then estimating the weights β_j . Both steps compose a single Expectation-Maximization iteration, and the EM method stops until convergence is achieved. Once the algorithm has converged, each pixel is classified with the class label which minimizes the Mahalanobis distance between the pixel intensity and the corresponding Gaussian component. That is, $c(\vec{r}) = \arg \min_j \{(l(\vec{r}) - \mu_j)^2 / (2\sigma_j^2)\}$.

This popular EM classifier can be used to segment MS lesions within the proposed two-step framework for comparison purposes against the SP-GMMF classifier. Additionally, another proposal for MS lesion segmentation based on EM is presented here. Specifically, the two-step algorithm can be implemented, first using SP-GMMF to segment the brain, and then using EM at pixel level classification to segment MS lesions. For simplicity, this algorithm will be called EM*. Notice that, both EM* and EM methods follow the same MS lesion classification of the proposed SP-GMMF method. In other words, EM* and EM methods, in the first step of the algorithm obtain a mask of the brain, then an intensity adjustment operation is applied to the masked region, and finally in the second step, the EM algorithm is applied to the intensity-adjusted brain, and the same discrimination criteria is applied to obtain the MS lesions.

RESULTS AND DISCUSSION

For comparison purposes, both SP-GMMF and EM* were tested and compared against EM using both synthetic and real brain MRI images. A brief summary of the details for each algorithm is shown in Table 1. Experimental results were obtained using the BrainWeb dataset [27] and several MRI images (axial and sagittal sequences) from two subjects positively

diagnosed with MS of the Central Hospital in San Luis Potosí, México. Accuracy and reliability of the results were primarily measured using the Dice Similarity Coefficient (DSC) defined as:

$$DSC = 2 |GT \cap R| / (|GT| + |R|) \quad (6)$$

where R is the estimated set of pixels corresponding to MS lesions obtained from the segmentation process, and GT corresponds to the Ground Truth. In the case of the BrainWeb set the ground truth was available on the BrainWeb site, and for real images, the GTs were manually obtained by one expert MS physician from the Central Hospital and validated by another expert from the same institution. In addition to DSC, the Sensitivity (SEN) and the average absolute difference between the number of lesions (ADNL) were computed as well.

TABLE 1. Block diagram of the proposed method SP-GMMF.

Algorithm	Brain Peeling	WML Detection	Resolution
EM	EM	EM	Pixel
EM*	SP-GMMF	EM	Pixel
SP-GMMF	SP-GMMF	SP-GMMF	Superpixel

The parameters for the first step (brain peeling) with SP-GMMF and EM* methods were: $K' = 1000$ desired number of SPs, $\gamma = 0.1$, $\kappa = 0.1$ and $\lambda = 0.1$ with 10 Gauss-Seidel iterations. The number of classes was $C = 5$, and their means were initialized by the k-means algorithm; the class with lowest mean intensity corresponds to the background, while the other classes correspond to the different anatomical structures in the image; however, it is important to recall that in this stage of the process we are interested in the largest connected non-background region, so the non-background classes are merged into a single class for this purpose. For the MS lesion detection stage, the parameters for the SP-GMMF and EM* were $K' = 5000$ for a finer resolution, $\kappa = 0.1$, $\lambda = 2$ and $\gamma = 5$, with $C = 7$,

with class means also initialized by the k-means algorithm; here, the class with highest mean intensity is considered to represent MS lesions. With respect to the EM method, the only parameter is the number of the classes, which we set $C = 5$ for the brain peeling stage and $C = 7$ for the MS lesion detection stage; class means are also initialized by k-means. Finally, in the False-Positive discrimination step, the criteria were a minimum area of 30 pixels and minimum eccentricity of 0.85 for all the algorithms. All parameters, including the number of classes considered at each stage, were optimized experimentally to maximize the DSC for real images from Subject 2 (see below).

Results from synthetic images

Data for these experiments consisted on simulated $181 \times 217 \times 181$ T2-weighted MRI volumes from the MS Lesion Brain Database in Brain Web [27] with slice thickness of 1mm, noise levels of 0%, 1%, 3%, 5%, 7% and 9%, and intensity non-uniformity (INU) of 0% and 20%. A total of 12 volumes were used for the experiment. The experiment with synthetic MRI images was designed to test the robustness to noise and INU. For evaluation purposes, experiments were performed selecting only the four slices that contain the highest number of MS lesions, corresponding to slices 31 to 34. BrainWeb does not provide T2-FLAIR images so T2 images were used instead for this experiment; however, the ventricles appear with high intensity in T2 images and are easily confounded with MS lesions; for this reason, only four slices that do not contain the ventricles were used.

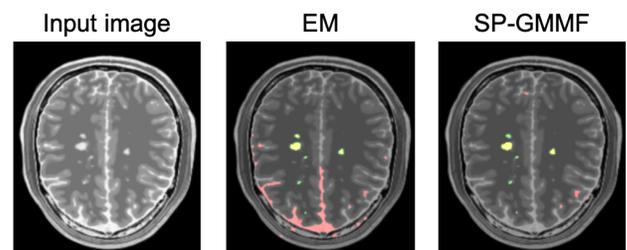


FIGURE 3. Application of the EM and SP-GMMF algorithms to slice 32 of the BrainWeb volume.

For illustration purposes, Figure 3 shows the results from the application of EM and SP-GMMF to slice 32 of the BrainWeb volume, where yellow regions correspond to true positive estimations, green regions correspond to false negatives and red regions correspond to false positives. Average results for these images are summarized in Figure 4.

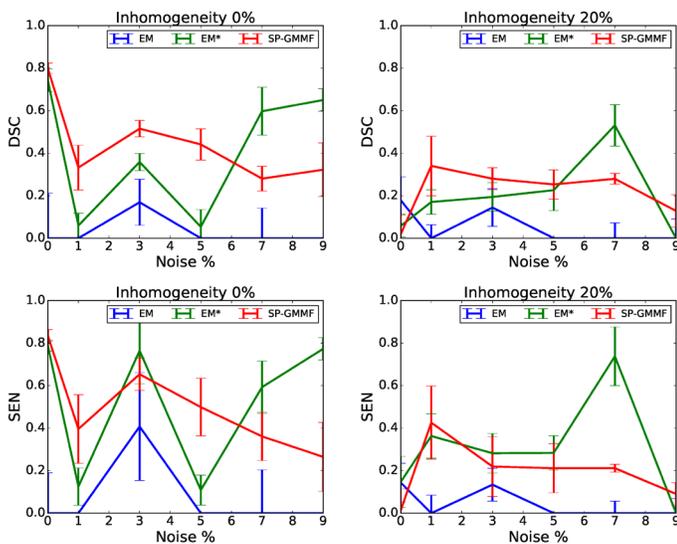


FIGURE 4. Average results of DSC and Sensitivity (SEN) from applying the EM, EM* and SP-GMMF methods on the BrainWeb images (slices 31 to 34).

Under low noise conditions (noise level $\leq 3\%$) and uniform intensity levels (INU = 0%), SP-GMMF achieves DSC values between 0.49 and 0.79 in average. However, when INU is increased to 20%, DSC values decrease approximately by a half, except in the case for 1% noise, in which the DSC values are maintained.

Additionally, a statistical analysis has been performed in order to determine if there exist significant differences in the performance indices (DSC and Sensitivity) between the different methods under study.

For this analysis, results for all noise levels were grouped so that each group consists of 24 data points (4 slices with 6 noise levels) for each method and INU level.

First, the Kruskal-Wallis non-parametric test was performed to see if there were differences in DSC or Sensitivity among the three methods for different INU levels; in all cases, significant differences ($p < 0.05$) were found. Post-hoc testing was then performed using the Kolmogorov-Smirnov and Mann-Whitney U tests, in order to determine which methods showed performance differences. The resulting p-values are summarized in Table 2.

TABLE 2. Results (p-values) from the Kruskal-Wallis (KW), Kolmogorov-Smirnov (KS) and Mann-Whitney U (MW) statistical tests to determine the existence of significant differences in performance measures (DSC and Sensitivity) between the methods under study (SP-GMMF, EM and EM*). P-values lower than 0.05 are shown in bold face, indicating significant differences among the methods in the corresponding row.

Dice Similarity Coefficient				
	Methods	All INU	0% INU	20% INU
KW	All methods	0.0000	0.0000	0.0390
KS	SP vs EM	0.0000	0.0000	0.0120
	SP vs EM*	0.1614	0.6860	0.2628
	EM vs EM*	0.0000	0.0000	0.1398
MW	SP vs EM	0.0000	0.0000	0.0089
	SP vs EM*	0.1479	0.3513	0.1045
	EM vs EM*	0.0000	0.0000	0.0504
Sensitivity				
	Methods	All INU	0% INU	20% INU
KW	All methods	0.0000	0.0000	0.0093
KS	SP vs EM	0.0000	0.0000	0.0120
	SP vs EM*	0.0996	0.9942	0.0299
	EM vs EM*	0.0000	0.0000	0.0120
MW	SP vs EM	0.0000	0.0000	0.0100
	SP vs EM*	0.2743	0.3706	0.2100
	EM vs EM*	0.0000	0.0000	0.0017

Furthermore, in order to determine the effect of intensity non-uniformity, a paired Wilcoxon test was performed to determine if there exist significant differences between INU levels 0% and 20% for each

method. Significant differences in DSC values were found for EM* ($p=0.009$) and SP-GMMF ($p=0.002$), and also in Sensitivity for EM* ($p=0.036$) and SP-GMMF ($p=0.002$), whereas EM did not show significant differences.

Results from real images

Another set of tests was performed with T2-FLAIR volumes from two MS patients acquired at the Central Hospital in San Luis Potosí, México, with a 1.5T MRI scanner. Images from Subject 1 (S1) were sampled approximately at 6.1 mm per slice, with each slice having a size of 512×512 pixels. For this test, four slices with visible MS lesions were selected: two axial and two sagittal.

The data set from the second subject (S2) is comprised by eight 256×256 axial images (A) and eight 288×288 sagittal images (S). Average results from the segmentation of these images using the methods under testing (EM, EM* and SP-GMMF) are shown in Table 3. Figure 5 shows boxplots of the DSC and SEN results.

TABLE 3. Average results (\pm standard error of mean) for real T2-FLAIR brain MRI images. Computed indices are: Dice Similarity Coefficient (DSC), Sensitivity (SEN) and Absolute Difference in Number of Lesions (ADNL - with the number of total lesions found shown in parentheses). Best results for each column are shown in bold face.

		S1	S2	
		A&S	A	A
DSC	EM	0.577±0.101	0.476±0.090	0.270±0.091
	EM*	0.610±0.076	0.479±0.084	0.401±0.099
	SP-GMMF	0.697±0.007	0.576±0.081	0.454±0.081
SEN	EM	0.592±0.136	0.576±0.116	0.559±0.115
	EM*	0.593±0.086	0.569±0.091	0.559±0.115
	SP-GMMF	0.719±0.039	0.651±0.099	0.574±0.108
ADNL	EM	3.25±0.48 (44)	1.50±0.46 (21)	2.00±0.63 (19)
	EM*	3.75±1.11 (40)	1.25±0.45 (17)	1.125±0.58 (24)
	SP-GMMF	2.25±0.63 (30)	1.75±0.37 (15)	1.5±0.38 (23)

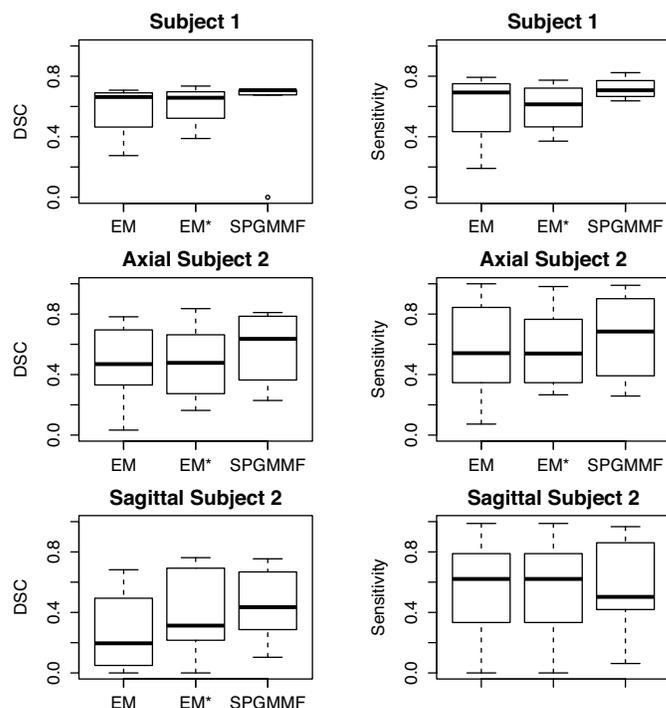


FIGURE 5. DSC and Sensitivity for both test subjects using the EM, EM* and SP-GMMF.

Figure 6 shows the results of the EM* and SP-GMMF algorithms for a few selected slices. Results are shown using the green channel for the true lesions and the red channel for the estimated lesions, so that yellow regions correspond to true positives, and green regions to false negatives. Statistical tests were also performed to determine if significant differences exist in the performance measures (DSC, Sensitivity and ADNL) among the three methods. In this case, however, the Kruskal-Wallis test did not report any evidence of significant differences; therefore, no post-hoc tests were applied.

Discussion

Experiments to test the accuracy of the proposed algorithm were performed using both synthetic and real MRI images of MS. The statistical analysis of the results for synthetic BrainWeb images suggest that INU highly affects the performance of the two proposed methods, SP-GMMF and EM*, since the DSC and sensitivity were strongly reduced when INU was set to 20%.

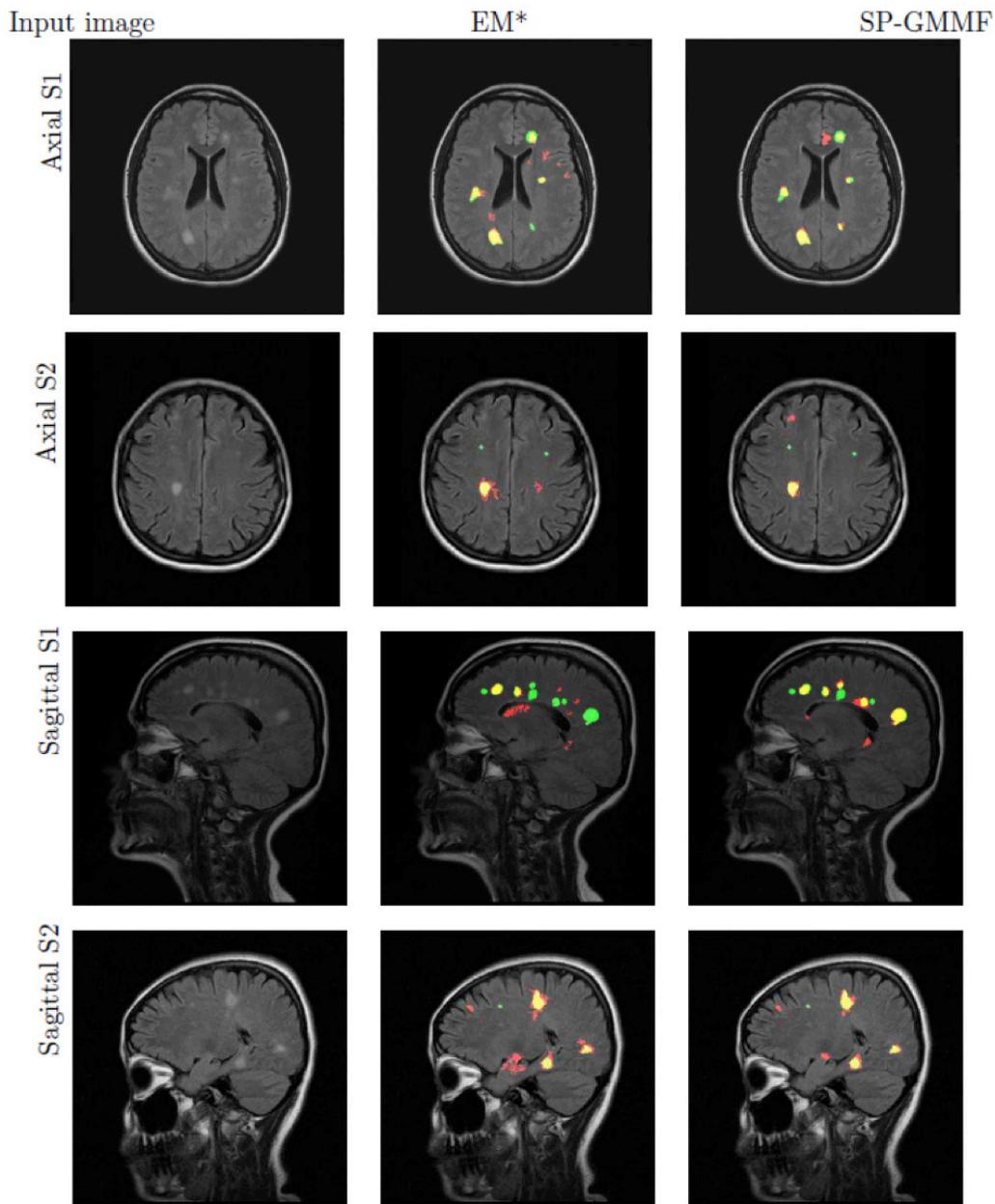


FIGURE 6. Example of the results obtained with EM* and SP-GMMF. Yellow regions correspond to the true positive estimations, where the automatic segmentation coincides with the expert segmentation. Green regions correspond to false negatives (lesions that the automatic methods did not detect), and Red regions correspond to false positives (regions that the automatic algorithm incorrectly reports as lesions).

With respect to the noise level, SP-GMMF shows a more consistent behavior with a mild decrease in DSC and Sensitivity as the noise increases, suggesting that the proposed algorithm is fairly robust to noise. On the other hand, it is difficult to characterize the behavior

of EM and EM* under noisy conditions as their results are highly variable, but it seems that the pixel-wise approach employed by EM* can be of benefit in the lesion segmentation step under higher noise levels. A statistical analysis of the performance indices, whose

results are summarized in Table 2, shows that, in general, EM has a significantly lower performance than EM* and SP-GMMF, possibly due to errors in the brain peeling stage. On the other hand, no significant differences between EM* and SP-GMMF were found.

In the literature, some methods are applied to the BrainWeb data [24] [28] [29] [30] [31] with a high performance in DSC. However, their good results are not clearly explained, for example, some of them do not mention the slice (or slices) used for the tests. In [29], experiments using T1 and T2 images with 3% and 5% of noise are shown, obtaining a DSC of 0.782 in the best case. These experiments illustrate the main advantage of using prior information from an atlas; in general, using an atlas is a good option, however, algorithms without an atlas could reach similar performance; for instance, in [29] they report a DSC = 0.74 with an atlas and DSC = 0.75 without atlas. Moreover, they do not indicate which atlas is used or how the subject brain is registered to the atlas, which suggests they might have used the healthy BrainWeb data as atlas in order to avoid the registration process; in that case, applying the method proposed in [29] to real images would require further pre-processing steps which could introduce additional errors. In [31], the authors show a methodology based in Gauss-Markov model followed by curve evolution. Experiments were achieved using 61 slices (from the 60th to the 120th) using the T1, T2 and PD images. They obtained very good results since DSC is over 0.78 for 3% to 9% of noise. The results obtained by Garcia-Lorenzo *et al.* [28] present an interesting behavior for INU = 0%. The method yields a lower DSC = 0.24 for low noise levels (1%), increases its performance for 3% and 5% of noise (DSC of 0.65 and 0.79 respectively), and then decreases at 7% and 9% with DSC = 0.6 and DSC = 0.25, respectively. These results, along with our own results from Figure 4, suggest that some algorithms perform better when there is a mild amount of noise in the images. For instance, the EM and EM* algorithms estimate the mean and

standard deviation of gray intensities for each class; in the absence of noise, the standard deviation will approach zero, which may cause numerical instabilities; on the other hand, the algorithm in [28] is gradient-based and may also be affected by untextured, noiseless regions. Table 4 summarizes the DSC results for various methods mentioned above, along with SP-GMMF for comparison purposes.

TABLE 4. DSC values reported for different MS lesion detection algorithms for synthetic (BrainWeb) and real MRI images. For each method, it is indicated if the method requires multi-modal images (T1, T2, PD/FLAIR), and if it requires prior information in the form of an atlas or user assistance.

Method	DSC for BrainWeb Images	DSC for Real Images	Requires multiple modality	Requires prior data
TOADS [12]	0.79	0.63	X	X
Graph Cuts [28]	~0.7	0.63	X	
AWEM [30]	0.7		X	
CGMM+CE [31]	0.78		X	X
Khayati [6] [7]		0.75		
SP-GMMF	0.49 - 0.79 (low noise)	~0.6		

Despite the sensitivity to INU and noise, experiments with real images suggest that the proposed methods can be used for a real application with encouraging results. Figure 5 shows the boxplots of the DSC and the sensitivity for the EM, EM* and the SP-GMMF methods. For Subject 1, SP-GMMF shows better DSC and sensitivity than the EM based methods. For Subject 2, the results show a considerable variability between slices; in general, SP-GMMF presents a higher median DSC but a lower sensitivity in the case of the sagittal images. Table 3 summarizes the average DSC, sensitivity and ADNL of the proposed methods, where the best result is highlighted in boldface. In this case, the best DSC performance is obtained by the SP-GMMF method with a competitive average DSC of 0.6968 for Subject 1 and average DSC of 0.5762 and 0.4544 for Subject 2 axial and sagittal images, respectively. SP-GMMF also

presents the highest average sensitivity. Nevertheless, the best ADNL is obtained by the EM* method; this is interesting because the number of lesions is one of the main clinical indices to characterize the progress of the MS disease. To our knowledge, only a few works in the literature report results with real images. In [30], they report results from three different patients for which they can obtain DSC values of 82%, 56%, and 52%, respectively. In [28], experiments with 10 patients yield DSC values between 40% and 75%. Considering that the authors of these works use multi-modal images (combining T1-w, T2-w, FLAIR and PD) with high resolution, we consider our results (DSC between 45% and 69%) to be competitive for low-resolution single modality imaging.

Finally, we discuss the differences in the false positive lesions reported by the proposed methods, as shown in Figure 6. Ideally, the algorithm should report no false negatives while keeping the number of false positives as low as possible. For subject one, SP-GMMF shows more accuracy because there are more yellow regions (true positives) and less red regions (false positives) than the segmentations obtained by the EM*, particularly in the sagittal image. For subject two, the advantage of SP-GMMF versus EM* is not clearly evident as they show almost the same yellow regions. Nevertheless, a detailed inspection of these images suggests that the EM* segmentation tends to generate a higher number of false positive (red regions), possibly due to the lack of regularization, such as the one induced by using superpixels, at the MS lesion detection stage.

CONCLUSIONS

An automatic algorithm to detect and segment Multiple Sclerosis lesions from T2 MR images was presented. The proposed method is based on a segmentation process where the image is subdivided into connected clusters (superpixels) which are then labeled according to their average intensity using Gauss-Markov Measure Field model. The segmentation pro-

cess is applied two-fold: first to isolate the brain region, and then to detect hyperintense spots within the brain region. Finally, some of the false positives are discarded depending on their area and eccentricity. An experimental test using synthetic images from the BrainWeb database shows that the proposed segmentation has strong advantages against the popular EM method, even under noisy conditions. While SP-GMMF is fairly robust to noise, it is very sensitive to non-uniformity, so additional pre-processing might be required for some images to deal with the spatial intensity variations. In the case of real MRI images, SP-GMMF maintains its advantages against EM, although using EM for detecting hyperintense spots (once the brain region has been isolated) has shown benefits, such as a more accurate lesion count. In brief, the SP-GMMF produces competitive results with single-modality, low-resolution images; it is also fully automatic and does not depend on prior anatomic information. Thus, SP-GMMF and EM* could be adequate image processing tools for low-resource institutions. We are currently working on implementing the proposed algorithms as part of an interactive application that can be used for clinical purposes, where an expert physicist can manually refine the segmentation by setting post-processing parameters in real-time and deleting spurious lesions, and then obtain indices such as the number and volume of lesions.

AUTHOR CONTRIBUTIONS

A.R. Contributed to the implementation of the proposed methodology and in conducting the experiments. A.A. Contributed in the direction of the project and development of the methodology. M.O.M. Contributed in the experimental design, validation of the results, reviewing the manuscript and management activities. E.R.A.S Contributed in the development of the methodology, validation of results and reviewing the manuscript. I.R.L. Contributed in the raise of the problem, development of the methodology, data acquisition and expert validation of the real cases (ground truth).

REFERENCES

- [1] Benito-León J, Morales JM, Rivera-Navarro J, Mitchell AJ. A review about the impact of multiple sclerosis on health-related quality of life. *Disability and Rehabilitation*. 2003;25(23):1291-1303. <https://doi.org/10.1080/09638280310001608591>
- [2] Manjón JV, Coupé P. volBrain: An Online MRI Brain Volumetry System. *Frontiers in Neuroinformatics*. 2016; 10:30. <https://doi.org/10.3389/fninf.2016.00030>
- [3] Smith SM. Fast robust automated brain extraction. *Human Brain Mapping*. 2002;17(3):143-155. <https://doi.org/10.1002/hbm.10062>
- [4] Mortazavi D, Kouzani AZ, Soltanian-Zadeh H. Segmentation of multiple sclerosis lesions in MR images: a review. *Neuroradiology*. 2012; 54(4): 299-320. <https://doi.org/10.1007/s00234-011-0886-7>
- [5] Garcia-Lorenzo D, Francis S, Narayanan S, Arnold DL, Collins DL. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical Image Analysis*. 2013;17(1):1-18. <https://doi.org/10.1016/j.media.2012.09.004>
- [6] Khayati R, Vafadust M, Towhidkhalaf F, Nabavi M. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and markov random field model. *Computers in Biology and Medicine*. 2008;38(3): 379-390. <https://doi.org/10.1016/j.compbiomed.2007.12.005>
- [7] Khayati R, Vafadust M, Towhidkhalaf F, Nabavi M. A novel method for automatic determination of different stages of multiple sclerosis lesions in brain MR FLAIR images. *Computerized Medical Imaging and Graphics*. 2008;32(2):124-133. <https://doi.org/10.1016/j.compmedimag.2007.10.003>
- [8] Vapnik VN. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*. 1999;10(5): 988-999. <https://doi.org/10.1109/72.788640>
- [9] Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesion in MR images: method and validation. *IEEE Transactions on Medical Imaging*. 1994;13(4):716-724. <https://doi.org/10.1109/42.363096>
- [10] de Boer R, van der Lijn F, Vrooman HA, Vernooij MW, Ikram MA, Breteler MMB, Niessen WJ. Automatic segmentation of brain tissue and white matter lesions in MRI. In 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Arlington: IEEE;2007:652-655. <https://doi.org/10.1109/ISBI.2007.356936>
- [11] Awad M, Chehdi K, Nasri A. Multicomponent Image Segmentation Using a Genetic Algorithm and Artificial Neural Network. *IEEE Geoscience and Remote Sensing Letters*. 2007; 4(4): 571-575. <https://doi.org/10.1109/LGRS.2007.903064>
- [12] Shiee N, Bazin P-L, Ozturk A, Reich DS, Calabresi PA, Pham DL. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage*. 2010;49(2):1524-1653. <https://doi.org/10.1016/j.neuroimage.2009.09.005>
- [13] Aït-Ali LS, Prima S, Hellier P, Carsin B, Edan G, Barillot C. STREM: A Robust Multidimensional Parametric Method to Segment MS Lesions in MRI. In Duncan JS, Gerig G (eds.). *Medical Image Computing and Computer-Assisted Intervention MICCAI*. Berlin: Springer. 2005;3749:409-416. https://doi.org/10.1007/11566465_5
- [14] Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via EM Algorithm. *Journal of the Royal Statistical Society*. 1977;39(1):1-22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- [15] García-Lorenzo D, Prima S, Morrissey SP, Barillot C. A robust Expectation-Maximization algorithm for Multiple Sclerosis lesion segmentation. *MICCAI Workshop: 3D Segmentation in the Clinic: A Grand Challenge II, MS lesion segmentation*. 2008:1-8.
- [16] Bartko JJ. Measurement and Reliability: Statistical Thinking Considerations. *Schizophrenia Bulletin*. 1991;17(3):483-489. <https://doi.org/10.1093/schbul/17.3.483>
- [17] Powers D. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *Journal Machine Learning Technologies*. 2011;2(1):37-63.
- [18] Lao Z, Shen D, Liu D, Jawad AF, Melhem ER, Launer LJ, Bryan RN, Davatzikos C. Computer-Assisted Segmentation of White Matter Lesions in 3D MR images using Support Vector Machine. *Academic Radiology*. 2008;15(3):300-313. <https://doi.org/10.1016/j.acra.2007.10.012>
- [19] Viola P, Wells WM. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*. 1997;24(2):137-154. <https://doi.org/10.1023/A:1007958904918>
- [20] Wang XY, Wang T, Bu J. Color image segmentation using pixel wise support vector machine classification. *Pattern Recognition*. 2011;44(4):777-787. <https://doi.org/10.1016/j.patcog.2010.08.008>
- [21] Toussaint N, Souplet JC, Fillard P. MedINRIA: Medical Image Navigation and Research Tool by INRIA. In *Proceedings of MICCAI Workshop on Interaction in Medical Image Analysis and Visualization*. Brisbane: MICCAI. 2007;4791:1-8.
- [22] Achata R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis Machine Intelligence*. 2012;34(11):2274-2282. <https://doi.org/10.1109/TPAMI.2012.120>
- [23] Marroquin JL, Velasco FA, Rivera M, Nakamura M. Gauss-Markov measure field models for low-level vision. *IEEE Transactions on Pattern Analysis Machine Intelligence*. 2001;23(4):337-348. <https://doi.org/10.1109/34.917570>
- [24] Cheng J, Liu J, Xu Y, Yin F, Kee-Wong DW, Tan NM, Tao D, Cheng CY, Aung T, Wong TY. Superpixel Classification Based Optic Disc and Optic Cup Segmentation for Glaucoma Screening. *IEEE Transactions on Medical Imaging*. 2013;32(6):1019-1032. <https://doi.org/10.1109/TMI.2013.2247770>

- [25] Ren CY, Reid I. gSLIC: a real-time implementation of SLIC superpixel segmentation. Technical Report [Internet]. 2011:1-6. Available from: http://www.carlyuheng.com/pdfs/gSLIC_report.pdf.
- [26] Haralick RM, Shapiro LG. Computer and Robot Vision. Boston, United States: Addison-Wesley Longman Publishing;1992:28-48p.
- [27] Cocosco CA, Kollokian V, Kwan KS, Pike GB, Evan AC. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. NeuroImage. 1997;5:425.
- [28] García-Lorenzo D, Lecoeur J, Arnold DL, Collins DL, Barillot C. Multiple Sclerosis Lesion Segmentation Using an Automatic Multimodal Graph Cuts. In Yang G-Z, Hawkes D, Rueckert D, Noble A, Taylor C (eds.). Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009:584-591. https://doi.org/10.1007/978-3-642-04271-3_71
- [29] Bricq S, Collet Ch, Armspach JP. Lesions detection on 3D brain MRI using trimmed likelihood estimator and probabilistic atlas. 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Paris; IEEE. 2008:93-96. <https://doi.org/10.1109/ISBI.2008.4540940>
- [30] Forbes F, Doyle S, Garcia-Lorenzo D, Barillot C, Dojat M. Adaptive weighed fusion of multiple MR sequences for brain lesion segmentation. 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Rotterdam: IEEE. 2010:69-72. <https://doi.org/10.1109/ISBI.2010.5490413>
- [31] Freifeld O, Greenspan H, Goldberger J. Multiple Sclerosis Lesion Detection Using Constrained GMM and Curve Evolution. International Journal of Biomedical Imaging. 2009: 715124. <https://doi.org/10.1155/2009/715124>