A NEW CLASSIFIER FEATURE SPACE FOR AN IMPROVED MULTIPLE SCLEROSIS LESION SEGMENTATION

X. Tomas-Fernandez, Simon K. Warfield

Computational Radiology Laboratory. Department of Radiology, Boston Children's Hospital.

ABSTRACT

Multiple Sclerosis lesion classifiers are usually based in anatomical Magnetic Resonance Imaging intensity features. The intensity overlap between lesions and healthy brain tissues results in a low sensitivity and specificity lesion segmentation. We propose a new, extended classifier feature space that is based in spatial locations, the intensity of which is abnormal when compared to the expected values in a healthy population. Validation on the new feature space shows an improvement in both sensitivity and specificity in lesion segmentation.

Index Terms— Multiple Sclerosis, Magnetic Resonance Imaging, Segmentation

1. INTRODUCTION

A widely used approach for MS lesion segmentation involves modeling the distribution of intensities in healthy brain MRI as a Gaussian Mixture Model (GMM) and then segmenting lesions as outliers of this model. Van Leemput [1] pioneered this strategy by introducing typicality weights for each voxel and class as well as intensity and contextual constraints to segment MS lesions as outliers.

Lesion and normal tissue MRI-derived intensity values overlap, thus making the MRI-derived intensity feature space a poor choice for discriminating brain tissue types in the case of MS patients. Despite these limitations, all proposed classifiers for MS rely on this kind of feature space. Since MRI-derived intensity features are sensitive, but not specific to MS lesions, brain tissue classifier results are affected by an increased number of lesion false positives. Several approaches have been proposed to mitigate this phenomenon. Algorithms modeling lesions as intensity outliers as in [1], apply a set of *ad hoc* neighboring rules to reject those voxels classified as outliers, but which do not correspond to MS lesions. In [2], the classifier feature space is extended with a distance map generated from an anatomical template. Although this approach reduces the number of false positives; it is also sensitive to misregistration error that may exist between the anatomical template and the MS patient. In a similar way, Shiee et al. [3] propose the use of a topological atlas to infer brain regions where lesions are more likely to occur within the classifier. More recently, Geremia et al. [4] applied a multi-channel and context-rich random forest classification that also makes use of a discriminative symmetry feature.

In this paper, we propose a novel, multichannel local characterization of MS lesions based on a voxel-level intensity profile across a set of MRI scans. In contrast to the classical approach where lesions are characterized by their intensity distribution along the brain; our new MS lesion model aims to distinguish locations in the brain with an abnormal intensity level as compared to the expected value in the same location in a healthy subject population. To incorporate this knowledge into a tissue brain classifier, we propose a novel feature space created by extending the common intensity classifier feature space using a measure of the typicality of a brain location intensity value. This is accomplished by comparing, voxel-by-voxel, the intensity level in a MS patient with the expected level in a set of healthy subjects by means of the Mahalanobis distance. We evaluate the specificity and sensitivity of this new, extended feature space with a set of clinical images provided by the MS Challenge segmentation at MICCAI08 database which shows that the proposed MS lesion characterization enables an improved discrimination among tissues when compared to the original intensity feature space.

2. METHODS

2.1. Overview

We propose a novel approach to local MS lesion characterization for segmenting MS patients brain MRI. Our goal is to take aligned T1w, T2w, and FLAIR MR images of an MS patient, detect voxels with an abnormal intensity level when compared to the expected value in a population of healthy subjects.

Our algorithm starts with a library of MRI images representing typical healthy subjects comparable in age to the study subject. Each reference subject image is then registered to the MS patient being segmented. Because MR image intensity varies from scan to scan due to the MRI acquisition process, intensity levels from each reference subject were normalized to match the intensity distribution in the MS patient. Finally, intensity levels between the study subject and the reference population were compared, voxel-by-voxel, by means of the Mahalanobis distance.

The local comparison between the study subject and a reference population of intensity values yields a measurement of the MS subject voxel intensity typicality that outlines brain locations with an unexpected intensity value such as MS lesions.

2.2 Reference Healthy Population Data

To evaluate the typicality of voxel intensity value from an MS subject, a group of 15 volunteers was used as the normal reference database.

The data from volunteers were acquired on a 3T clinical MR scanner from GE Medical Systems (Waukesha, WI, USA) using an 8-channel receiver head coil. For the reference database, MR scans were acquired using three different pulse sequences: a T1-weighted MPRAGE (Magnetization Prepared Rapid Acquisition Gradient Echo) sequence; a T2-weighted scan from an FSE (Fast Spin Echo) sequence; and a FLAIR scan, also run with an FSE sequence. The T1w sequence was axially acquired; T2w and FLAIR sequences were sagitally acquired. All sequences were acquired with a matrix size of 256x256 and a field of view of 28 cm. Slice thickness was 1.3 mm for the T1w-MPRAGE sequence; 1 mm for the T2w-FSE sequence; and 2 mm for the FLAIR-FSE sequence. The MPRAGE parameters were TR 10/TE 6 ms with a flip angle of 8. For the FSE, the paramenters were TR 3000/TE 140 ms with a train length of 15.

After the volunteer was scanned, T2w and FLAIR images were aligned to the T1w scan. In addition, all acquired MR images were re-oriented to an axial orientation and then the intra-crainial l volume was manually segmented by a radiologist.

2.3 Healthy Population Non-Rigid Registration

To evaluate a voxel's intensity value typicality, both the study subject and the reference database have to be aligned to the same reference space. Due to anatomical variability among subjects and the presence of MS lesions, a robust, block-matching based non-linear registration was used; this method extended the rigid registration algorithm proposed by [5].

The registration algorithm follows an iterative framework using a multi-resolution scheme. At each iteration *i*, pairings are computed between the images using Block-Matching. A correction displacement field δT is then interpolated from the sparse pairings *Ui* using the similarity values of the pairings as confidence parameters. An outlier rejection is then performed by comparing δT and *Ui*. If the norm of their difference is greater than an automatically defined threshold, then the pairing is considered an outlier and removed. A correction is then computed from the remaining pairings and composed with the current estimate of the transformation.

2.4 Intensity Normalization

MR image intensity varies from scan to scan due to variation in the physical state of the scanner hardware, interactions between the detector and the patient's body and pulse sequence parameters. This variability makes direct comparison of intensity levels among scans unfeasible.

Because of the anatomical abnormalities (MS lesions) present in the scans we want to normalize, an algorithm that does not rely on any pre specified model was chosen. To this purpose, we used the intensity normalization method proposed in [6]. Weisenfeld et al. developed a novel approach for normalizing the intensities within an image to best match a supplied histogram model, which can be generated by any representative subject, allowing to proceed without assumptions about the shape of the histogram or the specific contribution of a given class of tissue.

For our purposes, a histogram model was provided to the intensity normalization algorithm for each of the MR modalities (T1w, T2w and FLAIR) from the MS patient. Finally, intensity levels from each reference subject will be normalized to match the given histogram model.

2.5 Classifier Extended Feature Space Generation

Once the scans from the healthy reference dataset were aligned to the MS subject, and their intensity levels were normalized, a voxel-by-voxel comparison of the intensity levels between the subject under study and the reference population was performed. Intensity similarity was evaluated by means of the Mahalanobis distance [7].

Mahalanobis distance is a measure between two data points in the space defined by relevant features. Since it accounts for unequal variances as well as correlations between features; it will adequately evaluate the distance by assigning different weights or importance factors to the features of data points. This weighting will assign components with high variability *less* weight than components with low variability. The Mahalanobis distance is written as:

$$M(x_i) = \sqrt{(x_i - \mu_i)^T \Sigma_i^{-1} (x_i - \mu_i)}$$

Equation 1*: Mahalanobis distance*

where x_i is a vector composed by the MS subject, T1w, T2w, and FLAIR scans intensity values at voxel *i*, μ_i is the mean intensity vector at the healthy reference population, and Σ_i is the healthy population intensity covariance matrix at the location of voxel *i*. An image based on this equation is depicted in Figure 1.

3. VALIDATION

To evaluate the performance of this novel extended feature space, training subjects from the MS MICCAI 2008 database were used. The training database consists of 20 subjects provided by the University of North Carolina at Chapel Hill (UNC) and Children's Hospital Boston (CHB). For all subjects, the database contains a T1w, a T2w, and a FLAIR scan as well as manual lesion segmentation. All

scans were re-oriented to axial orientation. The T2w and FLAIR scans were rigidly registered to the corresponding T1w image. Finally, scans were resliced at 0.5x0.5x0.5 mm resolution with cubic spline interpolation.

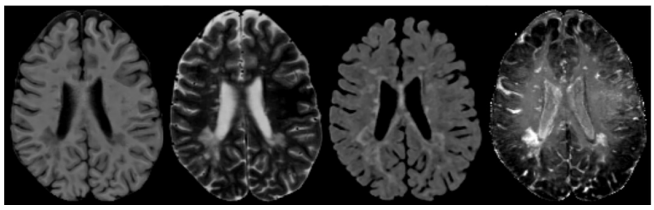

Figure 1: *Demonstration of the proposed extended feature space. From left to right: T1w scan, T2w scan, FLAIR scan, and Mahalanobis distance map. Notice how higher values in the Mahalanobis distance map correspond with MS lesions (hyper- or hypo-intensity voxels in anatomical MRI).*

To evaluate the assumption of MS lesion intensity values being outliers from healthy tissue, 11 randomly chosen CHB subjects were semi-automatically classified into CSF, GM, WM and MS lesion categories by using a manually trained kNN. The resulting segmentations were manually corrected for tissue misclassification. Next, for each of the 11 subjects, the Mahalanobis distance map was calculated, voxel-by-voxel, using Equation 1, where x_i is a vector composed by the T1w, T2w, and FLAIR intensity values at voxel *i*; μ_{GM} and Σ_{GM} is the mean intensity value; and the intensity covariance matrix for GM in the intensity space is defined by T1w, T2w and FLAIR scans. MS lesions were defined as locations in the brain where the intensity level surpassed a Mahalanobis distance of 1.4 when compared to the estimated GM intensity distribution. The estimated MS lesion mask was compared to the lesion manual segmentation to calculate the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) obtained. Last, the quality of the estimated MS lesion segmentation was evaluated by means of sensitivity and specificity, as reported in Figure 2.

MS lesion sensitivity, specificity of the intensity, and extended feature spaces were assessed for each subject included within the training set from the MS Grand Challenge database by using the full-knowledge of MS lesion intensity distribution provided by the available manual lesion segmentation. To this purpose end, for each of the subjects, both lesion and healthy tissue posterior probabilities were estimated from intensity histograms derived from the T1w, T2w, and FLAIR (in the case of the intensity feature space); T1w, T2w, FLAIR, and Mahalanobis distance map (in the case of the extended feature space); and from the tissue class for a given voxel's location obtained from the MS lesion manual segmentation. These composite histograms represent the number of times a given feature vector was both labeled as a lesion and *not* labeled as a lesion. Using the Bayes optimal decision boundary on the histograms, the number of MS lesion TP, TN, FP and FN were estimated. Figure 3 and Figure 4 reports sensitivity and specificity as well as PPV and FDR values for each patient when the intensity feature space or the extended version were used.

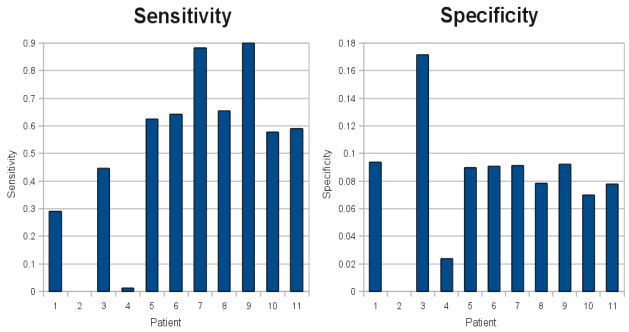

Figure 2: *Intensity outliers MS lesion classification sensitivity and specificity values.*

4. DISCUSSION

Although MRI intensities is the feature of choice in classifying brain tissue, the obvious overlap between lesions and healthy tissue intensity values in MS patients raises the question of whether a classifier feature space constituted by anatomical MRI (T1w, T2w, and FLAIR) intensities is the most appropriate choice for this task. By using the optimal Bayes decision boundary, a full understanding of lesions and healthy tissue intensity distribution can be achieved. This forms the basis upon which intensity values are inspected and precise MS lesion segmentations are obtained. (Figure 4 reports an average PPV of 0.98±0.02 and 0.86±0.05 respectively for the CHB and UNC sets). Tissue intensity values, however, are not sensitive enough to detect the full extent of MS lesion volume. (Figure 3 reports an average sensitivity value of 0.64±0.19 and 0.42±0.13 respectively for the CHB and UNC training sets). The lack of MS lesion sensitivity reported by the optimal Bayes classifier clearly reflects the overlap between lesion and healthy tissue intensity values.

In spite of the low MS lesion sensitivity reported by intensity based classifiers, one of the most popular MS brain segmentation strategies consists of modeling lesions as outliers when compared to the expected GM intensity distribution [1]. By applying this strategy we were able to show a lack of sensitivity (Figure 2 reports a mean sensitivity value of 0.51±0.30), effectively establishing that almost half of MS lesion voxels have an intensity level below the threshold at which outliers are typically defined. Clearly, this approach fails to detect the entire extent of the disease burden. Moreover, because of the intensity overlap that exists among brain tissues, over 90% of voxels labeled as lesions correspond to healthy tissue whose intensity is also over the defined threshold. (Figure 2 reports a mean specificity of 0.08±0.04).

Commonly, intensity-based MS lesion segmentation methods try to overcome low sensitivity and specificity by limiting the extent of lesions that appear in the WM. However, this strategy requires the use of anatomical templates that are sensitive to alignment error between the template and the MS patient, and/or post-processing steps based on experimentally tuned morphological operators, connectivity rules, size thresholds, etc. As an additional limitation, these steps often have to be re-tuned and tailored to each subject, which is time-consuming and inefficient. Instead, our approach to MS brain tissue segmentation does not rely on prior anatomical information or ad-hoc rules but rather on mimicking the manual segmentation of MS lesions, effectively producing a highly novel tool by which locations in the brain with abnormal intensity levels may be identified and easily compared to values expected in a healthy population.

In summary, our new extended feature space allows to achieve optimal MS segmentation beyond what is typically obtained solely by evaluating MRI intensity features. Further, with this technique, we have shown quantifiable improvements in sensitivity. (Figure 3 reports an average sensitivity of 0.65±0.19 and 0.51±0.19 for the CHB and UNC datasets respectively.) We have also made significant improvements in the specificity of the resulting segmentation. (Figure 4 reports an average PPV of 99.36±0.94 and 94.14±5.67 for the CHB and UNC datasets).

Acknowledgements

Xavier Tomas Fernandez was supported by a research grant from Caja de Ahorros y Pensiones de Barcelona “La Caixa” and a DuPre Fellowship from the Multiple Sclerosis International Federation (MSIF). This investigation was supported in part by the National Multiple Sclerosis Society under Grant RG 3478A2/2 and Grant RG 4032A1/1 , and by NIH grants R01 RR021885, R01 EB008015, R03 EB008680 and R01 LM010033.

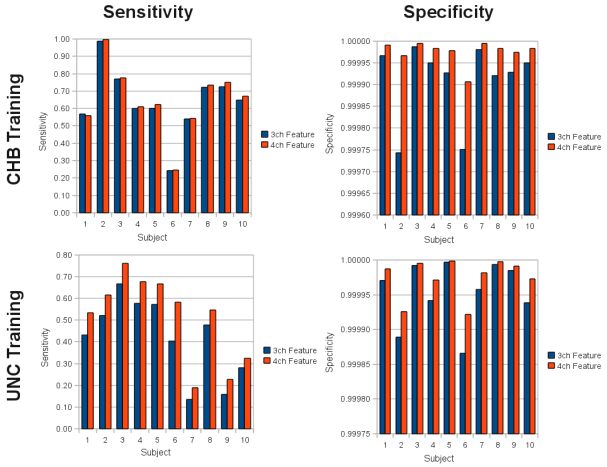

Figure 3: *MS lesion sensitivity and specificity values. Reported in blue intensity feature space results; in orange the extended feature space results. Consistently along all the subjects, the extended feature space allows to obtain improve the sensitivity and specificity of MS lesion segmentation.*

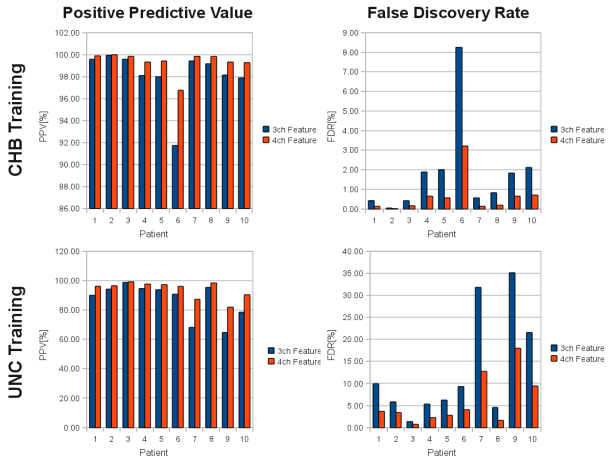

Figure 4: *MS lesion PPV and FDR values. Reported in blue are intensity feature space results; in orange, the extended feature space results. It is obvious that the extended feature space allows the radiologist to obtain a much more precise MS lesion segmentation.*

6. REFERENCES

[1] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, “Automated model-based tissue classification of MR images of the brain”, IEEE Trans Med Imaging 18, pp. 897-908, 1999.

[2] S.K. Warfield, M. Kaus, F.A. Jolesz, and R. Kikinis, “Adaptive, template moderated, spatially varying statistical classification”, Med. Image Anal 4, pp 43-55, 2000.

[3] N. Shiee, P. Bazin, A. Ozturk, D.S. Reich, P.A. Calabresi, and D.L. Pham, “A topology-preserving approach to the segmentation of brain images with multiple sclerosis”, Neuroimage 49, pp. 1524-1535, 2010.

[4] E. Geremia, B.H. Menze, O. Clatz, E. Konukoglu, A. Criminisi and N. Ayache, “Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel MR Images”, MICCAI, pp. 111-118, 2010.

[5] S. Ourselin, A. Roche, G. Subsol, X. Pennec, and N. Ayache, “Reconstructing a 3D structure from serial histogram sections”, Image and Vision Computing 19, pp. 25-31, 2001.

[6] N. Weisenfeld, and S.K. Warfield, “Normalization of Joint Image-Intensity Statistics in MRI using the kullback-leibler divergence”, ISBI, pp. 101-104, 2004

[7] P. Mahalanobis, “On the generalized distance in statistics”, Proceedings of the National Institute of Sciences of India 2, pp 49-55, 1936.