Statistics and Its Interface Volume 12 (2019) 193–202

An integrative classification model for multiple sclerosis lesion detection in multimodal MRI

Fengqing (Zoe) Zhang*, Ji-Ping Wang, and Wenxin Jiang

We study a classification problem of multiple sclerosis (MS) lesions in three dimensional brain magnetic resonance (MR) images. Segmentation of MS lesions is essential for MS diagnosis, assessment of disease progression and evaluation of treatment efficacy. Accurate identification of MS lesions in MR images is challenging due to variability in lesion location, size and shape in addition to anatomical variability between subjects. We propose a supervised classification algorithm for segmenting MS lesions, which integrates the intensity information from multiple MRI modalities, the texture information, and the spatial information in a Bayesian framework. A multinomial logistic regression is employed to learn the posterior probability distributions from the intensity information, combined from three MRI modalities. Texture features are selected by the Elastic Net model. The spatial information is then incorporated using a Markov random field prior. Finally, a maximum a posteriori segmentation is obtained by the graph cuts algorithm. We illustrate the effectiveness of our proposed model for lesion segmentation using both the synthetic BrainWeb data and the clinical neuroimaging data.

AMS 2000 subject classifications: Primary 62H30; secondary 62H35.
Keywords and phrases: Supervised classification algorithm, Multiple sclerosis, Segmentation, Multimodal MRI.

1. INTRODUCTION

Multiple sclerosis (MS) is an autoimmune disease that affects the brain and spinal cord. MS lesions may form in the white matter of central nervous system at any location, and hence clinical presentations may vary. Continuing lesion formation in MS often brings about physical disabilities in different organs of the body and, sometimes, to cognitive decline. There are about 250,000 to 350,000 MS patients in the United States alone. Though it is unknown what exactly causes this disease, a virus or gene defect or both are considered as the most probable causes. Magnetic Resonance Imaging (MRI) is the most sensitive technique to detect MS lesions and has proved to be an important paraclinical tool for MS diagnosis, assessment of disease progression, and evaluation of the efficiency of drug therapy. A variety of MRI modalities have been widely used in assessment of MS lesions, including T1-weighted (T1-w), T2-weighted (T2-w), proton density-weighted (PD-w), and fluid-attenuated inversion recovery (FLAIR) MRI. In MS clinical trials, measures derived from MRI, such as lesion load, have been established as standard outcome markers to monitor the natural history and treatment effects of the disease. Correct segmentation of MRI images with MS lesions is the first essential step in characterizing the MS lesion load, and in calculating and interpreting more specialized measures of damage. The process of segmenting MRI images with MS lesions is to assign an anatomically meaningful label to each voxel of the image. Gray matter (GM), white matter (WM), cerebrospinal fluid (CSF) and MS lesion (MSL) are the four main brain tissue types of interest.

Segmentation of MRI images with MS lesions is challenging due to several reasons. First, MRI images contain various noise artifacts such as intratissue noise and thermal noise, and thus have a much lower signal-to-noise ratio than regular images. Second, the geometry of the brain tissues is complex and hence MRI images are textured in complex ways. Third, the intensity distributions of different tissues have non-negligible overlaps. Classifiers solely based on voxel intensity values tend to have poor performance in segmenting MRI images. Fourth, due to the limited spatial resolution of MRI, a voxel may contain more than one tissue type, which is known as the partial volume (PV) effect. Voxels located on the borders of various tissues tend to contain a mixture of tissues. Last, abnormal structures of MS lesions make the segmentation task even more difficult, because they cannot be precisely defined. In general, MS lesions are brighter than the surrounding WM in T2-w, PD-w, and FLAIR images and are darker than the surrounding WM in T1-w images. But this definition lacks precision as MS lesions have large overlapping in intensity distribution with healthy tissues. For example, in T1-w images, lesions have large intensity distribution overlapping with GM while the borders of lesions and CSF are not clear in T2-w images.

Manual segmentation of MS lesions by radiologists is time consuming and is prone to large intra- and inter-expert variability. Therefore, it is highly desirable to develop fully automated and reproducible methods for segmenting MRI images. A few methods have been proposed in the literature. A Parzen window density estimation method was used for

*Corresponding author.

the segmentation [3] after performing histogram-based intensity normalization [15]. Tomas-Fernandez and Warfield proposed to segment an MRI image by comparing its intensities on a voxel-by-voxel basis to a group of MRI images of healthy subjects [22]. While voxel intensity information can be used to effectively detect the MS lesions, incorporating some spatial information could improve the segmentation. For example, Souplet et al. proposed to use a Gaussian mixture model together with a priori information of the MNI registered atlas [17]. A constrained Gaussian mixture model was proposed to simultaneously include spatial and intensity features [4]. Additionally, several studies combined intensity information from multiple MRI modalities for MS segmentation. For example, an automated statistical method was proposed to incorporate multiple MRI modalities using logistic regression [21]. A local logistic regression together with inter-modal coupling regression was developed to capture the relationships between imaging modalities [23]. In addition, some recent methods employed a large set of features calculated from image intensities. Morra et al. used an Adaboost approach with more than 10,000 features such as Haar-like filters of various shapes and sizes [14]. Kroon et al. proposed a principal component analysis approach with several hundred features such as intensity derivatives and intensity of the neighbor voxels [12]. While these methods all provide promising ways for MS lesion segmentation, no single method has been shown to have dominant advantages over the rest, probably due to the heterogeneity of lesions and variability in the MRI acquisitions [6]. MS lesion segmentation remains an open problem.

In this study, we proposed a Bayesian framework for segmenting MRI images with MS lesions. Our proposed approach offers a flexible probabilistic model to incorporate not only intensity information from multiple MRI modalities (T1-w, T2-w, PD-w, and FLAIR), but also a large set of texture features while accounting for spatial information from voxel neighborhood through Markov random field. The rest of this paper is organized as follows. We introduce the mathematical formulation of MS segmentation problem in Section 2. Since a voxel is basically a three-dimensional pixel, in this paper we will use voxels and pixels interchangeably. We present our proposed model to integrate intensity information from multiple modalities, spatial information as well as texture features in Section 3. We evaluate our model using both simulated brain data and clinical neuroimaging data of MS patients in Section 4. Finally Section 5 concludes with some discussion and future work.

2. PROBLEM FORMULATION

Segmentation of MRI images with MS lesions is to assign a tissue type to each point in the image, where we are particularly interested in four tissue types, i.e., gray matter (GM), white matter (WM), cerebrospinal fluid (CSF) and MS lesions (MSL). Consider an image of n pixels, $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, where each pixel is d-dimensional vector, i.e., $\mathbf{x}_i \in \mathcal{R}^d$. The d dimensions can represent intensity information from multiple MRI modalities and/or a large set of texture features. Let $\mathcal{K} = \{1, \ldots, K\}$ be the set of K classes. In our segmentation problem, classes can be four tissue types of interest, i.e., GM, WM, CSF and MSL. Image labels are discrete values, denoted as $\mathbf{y} = (y_1, \ldots, y_n)$. For example, y_i can be 0, 1, 2 or 3, representing GM, WM, CSF or MSL. The goal of our MRI image segmentation problem is to assign a label $y_i \in \mathcal{K}$ to each pixel vector $\mathbf{x}_i$.

In our Bayesian framework, the labels $\mathbf{y}$ are usually estimated by maximizing the posterior distribution as follows:

$$p(\mathbf{y}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{y})p(\mathbf{y}). \tag{1}$$

where $p(\mathbf{x}|\mathbf{y})$ is the likelihood function and $p(\mathbf{y})$ is the prior over the image labels. Assuming conditional independence of the features given the labels, i.e., $p(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^{n} p(\mathbf{x}_i|y_i)$, we can write the posterior as below:

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \frac{1}{p(\mathbf{x})} p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \\ &= \frac{1}{p(\mathbf{x})} \prod_{i=1}^{n} p(\mathbf{x}_i|y_i)p(\mathbf{y}) \\ &= g(\mathbf{x}) \prod_{i=1}^{n} \frac{p(y_i|\mathbf{x}_i)}{p(y_i)} p(\mathbf{y}). \end{aligned} \tag{2}$$

where $g(\mathbf{x}) = \frac{\prod_{i=1}^{n} p(\mathbf{x}_i)}{p(\mathbf{x})}$ is a factor not depending on $\mathbf{y}$. Further, in the prior, we assume the classes are equally likely, i.e., $p(y_i = k) = 1/K$ for any $k \in \mathcal{K}$. Thus, the maximum a posteriori criterion is given as follows:

$$\begin{aligned} \hat{\mathbf{y}} &= \arg\max_{\mathbf{y} \in \mathcal{K}^n} \{p(\mathbf{x}|\mathbf{y})p(\mathbf{y})\} \\ &= \arg\max_{\mathbf{y} \in \mathcal{K}^n} \{\sum_{i=1}^{n} (\log p(y_i|\mathbf{x}_i)) + \log p(\mathbf{y})\}. \end{aligned} \tag{3}$$

As discussed earlier, our proposed model can naturally incorporate not only intensity information from multiple MRI modalities, but spatial information from neighborhood voxels, and even a large set of texture features. More specifically, intensity information and texture features are represented by class densities $p(y_i|\mathbf{x}_i)$, which can be learned from training data. Notice that $p(y_i|\mathbf{x}_i)$ gives us a general form for regressing $\mathbf{y}$ on $\mathbf{x}$ and thus many potential models, such as logistic regression and penalized regression, can be used here. On the other hand, the spatial information can be captured by the prior $p(\mathbf{y})$ that encourages neighborhood homogeneity for the belief that neighboring pixels should have higher probability to belong to the same type of tissue. Although mathematically simple, this type of maximum a posteriori estimation presents computational challenges. In this paper, we adopt the graph cuts algorithm proposed by Boykov et al. [2]. Details of our proposed model are discussed in the following section.

3. METHODS

3.1 Multinomial logistic regression

If we only consider intensity values from T1-w, T2-w, and PD-w MRI images, each pixel $\mathbf{x}_i$ is a three dimensional vector, combining intensity information from multiple MRI modalities. We can directly use a multinomial logistic regression. More specifically,

$$p(y_i = k|\mathbf{x}_i, \boldsymbol{\beta}) = \frac{exp(\boldsymbol{\beta}_k^T \mathbf{x}_i)}{1 + \sum_{j=1}^{K-1} exp(\boldsymbol{\beta}_j^T \mathbf{x}_i)}, \tag{4}$$

where $\boldsymbol{\beta}$ are regression coefficients and k=1,..,K-1.

To include texture features, we first apply the Elastic Net model [26] for feature selection and then use the multinomial logistic regression to combine the intensity information with the selected texture information. Due to the heterogeneity in the data, all features are standardized as recommended by Zou and Hastie [26]. The resulting model maximizes the penalized likelihood by incorporating the elastic net penalty.

3.2 Markov random field

In a Markov random field (MRF), the pixel sites $\mathcal{S} = \{1, ..., n\}$ are related to one another via a neighborhood system, defined as $N = \{N_i, i \in \mathcal{S}\}$. A random field $\mathbf{z}$ is a MRF on $\mathcal{S}$ with respect to a neighborhood N if and only if $p(\mathbf{z}) > 0$ and $p(z_i|z_{S-i}) = p(z_i|z_{N_i})$. A MRF can be characterized by a Boltzmann distribution (also called Gibbs distribution), $p(\mathbf{z}) = Z^{-1}exp(-U(\mathbf{z}))$, where Z is a normalizing constant and $U(\mathbf{z})$ is an energy function. One commonly used MRF for capturing spatial information is the Ising model $p(\mathbf{z}) = C(\beta)^{-1}exp\{\sum_i^N \alpha_i(z_i) + \beta \sum_{i\sim j} w_{ij} I(z_i = z_j)\}$, where $C(\beta)$ is a normalizing constant, $\sum_i^N \alpha_i(z_i)$ represents the external field, and $\beta \sum_{i\sim j} w_{ij} I(z_i = z_j)$ captures the internal field, a summation over all neighboring pairs z_i and z_j. Here, $I(\cdot)$ is the identity function. More specifically, in our model, we used a MRF prior as below:

$$p(\mathbf{y}) = \frac{1}{Z} \exp(\mu \sum_{i\sim j} I(y_i = y_j)), \tag{5}$$

where Z is a normalizing constant, and the parameter μ controls the level of smoothness. Notice that under the Equation 5, we have $p(y_i = k) = 1/K$ for any $k \in \mathcal{K}$.

3.3 Maximum a posteriori via graph cuts

Given the posterior densities $p(y_i|\mathbf{x}_i)$ (Equation 4) and prior $p(\mathbf{y})$ (Equation 5), the maximum a posteriori can be written as follows:

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}\in\mathcal{K}^n} \{\sum_{i\in\mathcal{S}} -\log p(y_i|\mathbf{x}_i, \boldsymbol{\beta}) - \mu \sum_{i\sim j} I(y_i = y_j)\}. \tag{6}$$

This is a combinatorial optimization problem involving unary and pairwise interaction terms, which poses computational challenges. We used the Graph Cuts algorithm proposed by Boykov et al. [2]. Note that Z is the constant normalization term and thus is dropped in Equation 6. The tuning parameter μ balances the contribution of intensity information and spatial information to the model estimation and could be selected using cross-validation.

3.4 Texture analysis

Image texture feature can be regarded as the spatial variation of pixel intensities calculated from a predefined neighborhood. Statistical approaches for texture analysis attempt to represent the texture indirectly by statistics that summarize the distributions and relationships between the gray level of an image.

In the first-order statistical texture analysis, texture information is extracted from the image intensity histogram or pixel occurrence probability. The commonly used histogram based texture features include mean, variance, skewness, kurtosis, energy and entropy. These features can also be calculated from the histogram of absolute values of intensity gradients.

In the second-order statistical texture analysis, texture information is calculated from the joint probability distributions of a pair of pixels at certain distances and orientations. The gray-level co-occurrence matrix (GLCM) describes the number of times that a pair of pixels having gray-level values i and j occurs at a distance d apart along a given direction θ from a predefined neighborhood. For computational consideration, an image is often rebinned to N_g gray levels where N_g is a positive integer, such as 8, 16, 32, and 64. For an image of N_g gray levels, the GLCM is a $N_g \times N_g$ matrix. Common values for θ are $0°, 45°, 90°, 135°$. To capture fine texture information, a relatively small distance range of $d = 1, \ldots, 4$ is commonly used.

Haralick et al. [8] proposed a set of statistics that are extracted from the GLCM to quantify the spatial dependence of gray-level values as texture features. Let $p(i, j)$ be the (i, j)th entry in the given GLCM. The three statistics we considered are described as below:

- Energy: sum of squared elements in the GLCM,

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)^2. \tag{7}$$

- Contrast: measures of the intensity contrast between a pixel and its neighbor,

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j). \tag{8}$$

- Homogeneity: measure of the closeness of the distribution of elements in the GLCM to the GLCM diagonal,

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)/(1 + |i - j|). \tag{9}$$

In higher-order statistical texture analysis, a set of texture features based on gray-level run length matrix (GLRLM) is studied by Galloway [5]. The GLRLM matrix element (i, j) describes the number of times that a predefined neighborhood contains a run of length j consisting of pixels having gray level i along a given direction θ. Let N_r be the number of different run lengths. Hence, the GLRLM is a $N_g \times N_r$ matrix. Here $p(i, j)$ is the (i, j)th entry in the given GLRLM. The four texture measures we used are listed as follows:

- Short runs emphasis,

$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{p(i,j)}{j^2}/\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j). \tag{10}$$

- Long runs emphasis,

$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}j^2p(i,j)/\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j). \tag{11}$$

- Gray level nonuniformity,

$$\sum_{i=1}^{N_g}(\sum_{j=1}^{N_r}p(i,j))^2/\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j). \tag{12}$$

- Run length nonuniformity,

$$\sum_{j=1}^{N_r}(\sum_{i=1}^{N_g}p(i,j))^2/\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j). \tag{13}$$

In this study, a large set of texture features generated from the first-order, second-order and higher-order statistical texture analyses are included. We consider four different distance values d ranging from 1 to 4. For each distance value, we extract features (e.g., energy) from all four directions $\theta = 0°, 45°, 90°, 135°$. Texture features are selected by the Elastic Net model and then included in the multinomial logistic regression as input data for predicting MS lesions.

4. RESULTS

4.1 BrainWeb synthetic data

To evaluate segmentation algorithms, synthetic data and real clinical data are often both used. Synthetic images are simulated by a computer which have the advantages that users can fully control all the parameters in the image and the ground truth is known. In MS segmentation, the images most widely used for validation are those from BrainWeb [13]. These images were produced by an MRI simulator, which combines an anatomic model and MRI acquisition physics to generate images with different nonuniformity and noise level. Additionally, it is essential to validate a segmentation algorithm on real clinical images. Lesion regions are manually segmented by radiologists, which is also prone to large intra- and inter-expert variability. These segmented lesion labels can be used to evaluate model prediction. One limitation of using clinical data is the lack of a ground truth for the MS lesion.

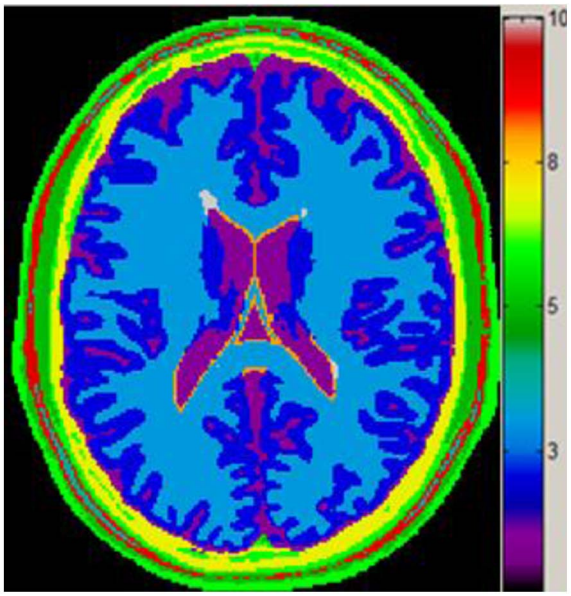

Figure 1. An image of different tissue type labels. Different tissue types are color coded as follows: 0 = Background, 1 = CSF, 2 = Grey Matter, 3 = White Matter, 4 = Fat, 5 = Muscle/Skin, 6 = Skin, 7 = Skull, 8 = Glial Matter, 9 = Connective, 10 = MS Lesion.

We first evaluated our model on MRI images with MS lesions obtained from the BrainWeb. Three coregistered modalities were used including T1-w, T2-w, and PD-w MRI. Voxel size is $1mm \times 1mm \times 1mm$. Figure 1 shows an image of tissue type labels. Each pixel was labeled from 0 to 10, representing one of the eleven possible tissue types. These labels serve as the ground truth to evaluate our model performance. Pixels were colored based on their labels according to the color bar. Here, we are particularly interested in segmenting the images to GM, WM, CSF, and MS lesions, while everything else is considered as background in our model.

Figure 2 shows one T1-w MRI image, its contour plot as well as the intensity histograms of WM, GM, CSF and MS lesions. Clearly the intensity distributions of WM, GM, and CSF have different peak locations though the overlapping of these distributions is not negligible. However, the intensity distribution of MS lesions has a large overlap with the distribution of GM in T1-w MRI images. As a result, classifiers solely based on intensities cannot produce satisfactory segmentation results.

Figure 3 shows one T2-w MRI image and its contour plot, along with the intensity histograms of WM, GM, CSF and MS lesions. Similar observations can be found in Figures 2 and 3. One difference is that the intensity distribution of

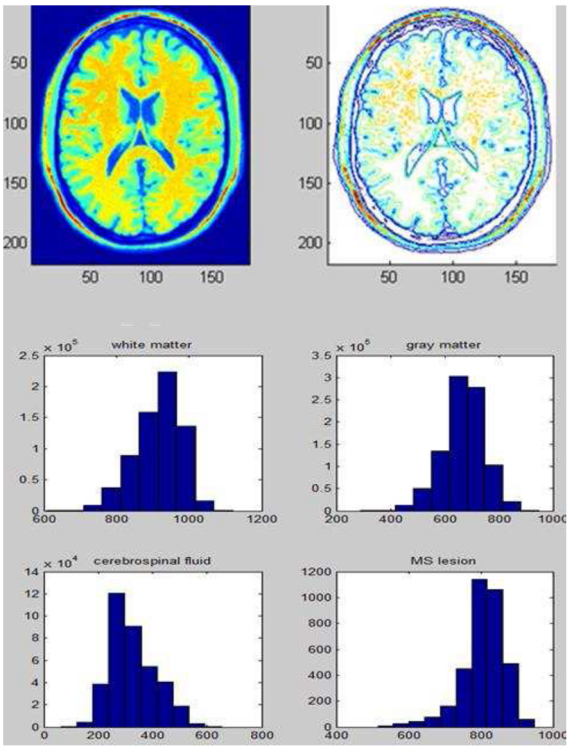

Figure 2. A T1-w MRI image with its contour plot and intensity histograms of WM, GM, CSF and MS lesions.

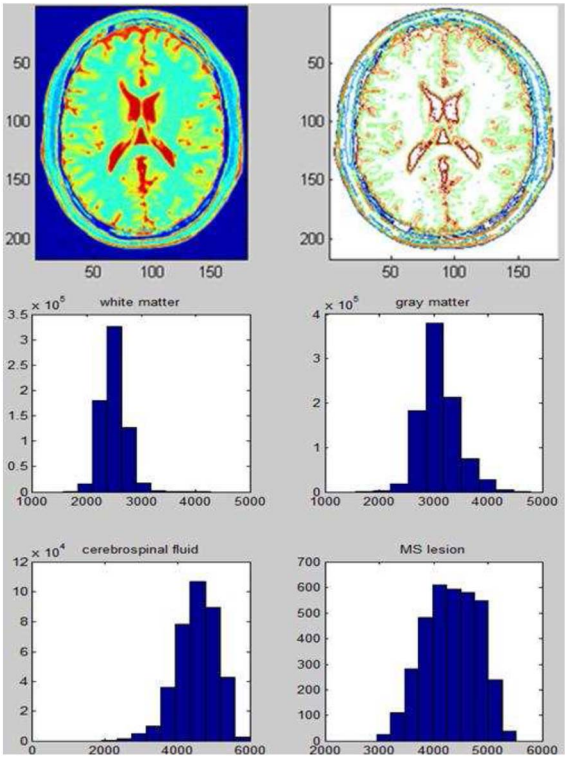

Figure 3. A T2-w MRI image with its contour plot and intensity histograms of WM, GM, CSF and MS lesions.

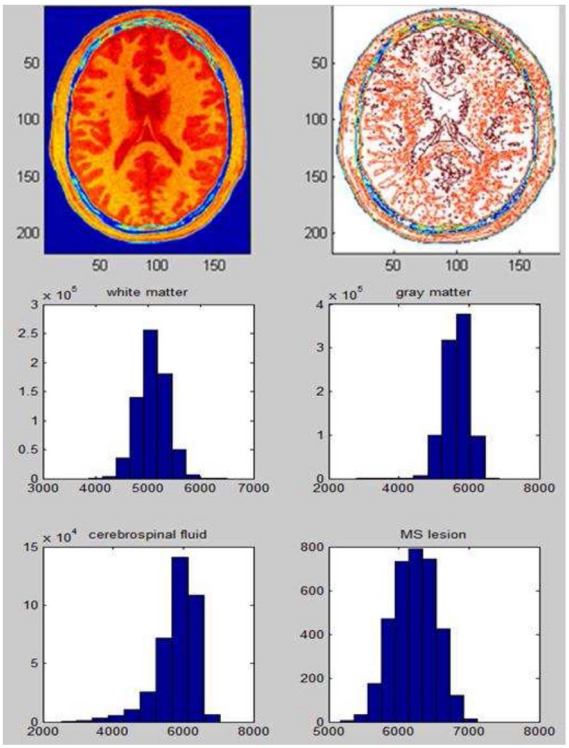

Figure 4. A PD-w MRI image with its contour plot and intensity histograms of WM, GM, CSF and MS lesions.

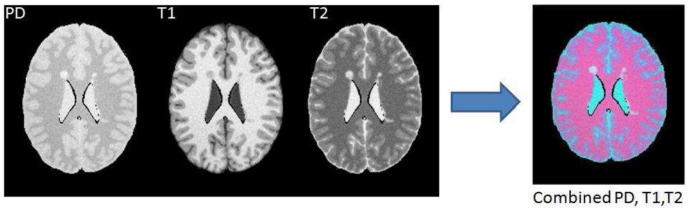

Figure 5. T1-w, T2-w, and PD-w MRI gray images and the combined color image.

MS lesions has a large overlap with the distribution of CSF in T2-w MRI images.

Figure 4 shows one PD-w MRI images, its contour plot, and the intensity histograms of WM, GM, CSF and MS lesions. We can see in general PD-w MRI images do not provide good intensity contrast among WM, GM, CSF and MS lesions.

As discussed earlier, we proposed to combine intensity information from multiple MRI modalities. Figure 5 shows T1-w, T2-w, and PD-w MRI gray images and the combined color image. The combined color image serves as the input of our model. Texture features are not calculated because the simulated BrainWeb data is unlikely to have texture information similar to true MRI images.

Segmentation results on BrainWeb MRI data at 20% intensity non-uniformality level and 3% noise level from our

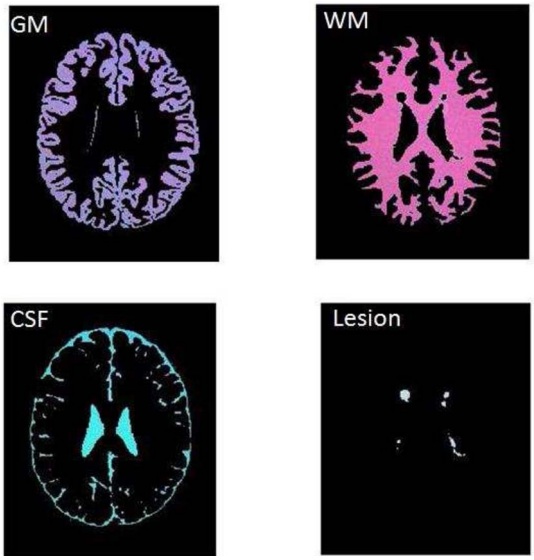

Figure 6. Segmentation results obtained by our proposed model.

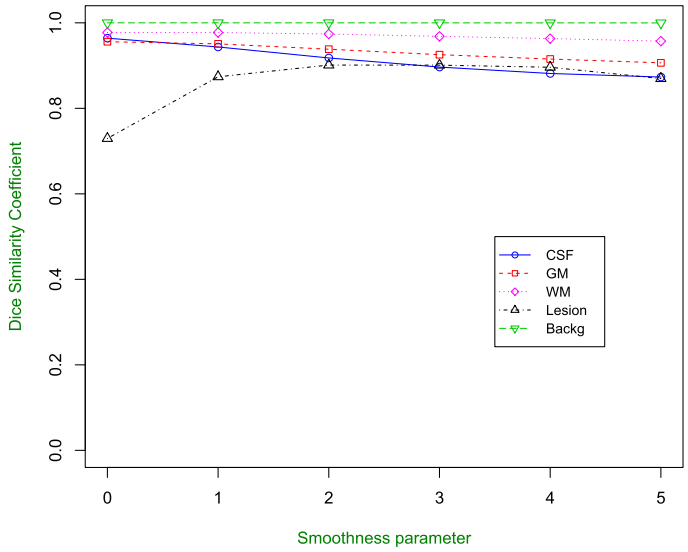

Figure 7. Dice similarity coefficients for different tissue types with different values of the tuning parameter μ.

model are shown in Figure 6. It can be seen that our model performed well for segmenting the images to GM, WM, CSF, and even for lesion detection.

In the literature, the most widely employed measure of segmentation accuracy is the Dice similarity coefficient (DSC).

$$DSC = \frac{2 \times TP}{FP + FN + 2 \times TP},$$

where TP, FP, and FN stand for true positives, false positives, and false negatives respectively. The value of DSC varies between 0 and 1, with 0.7 normally considered as a good segmentation [6]. It should be noted that $DSC = 0.7$ is considered as the clinical stratification for lesion detection [25]. In addition, a comparison between two segmentations given by two different human experts assuming one person's segmentation result is the ground truth is likely to yield a DSC lower than 0.7 [9].

Hence, we summarized the Dice similarity coefficients obtained from our model for different tissue types with different values of the parameter μ in Figure 7. The parameter μ controls the spatial smoothness. With varying level for μ, we can see how this parameter affects classification results. Further, the optimal parameter μ was selected by cross-validation and we summarized the DSC obtained by our model in Table 1. For each time, four out of the five slices from one brain volume simulated by BrainWeb were used as the training set and the leave-out slice was used as the test set. This process was repeated five times, as five-fold cross validation.

In addition, we compared our proposed multinomial logistic regression with Markov random field (MLMR) approach

Table 1. Dice similarity coefficients for different tissue types from five-fold cross validation using five image slices from one simulated brain volume by BrainWeb

Slice Number	CSF	GM	WM	Lesion	Backg
96	0.93	0.94	0.97	0.91	1
97	0.91	0.93	0.97	0.73	0.99
98	0.92	0.94	0.97	0.90	1
99	0.91	0.94	0.98	0.78	0.99
100	0.92	0.93	0.97	0.81	0.99

with two other methods, including the constrained Gaussian mixture model with curve evolution (CGMM+CE) and KVL algorithm [4, 24]. These methods were tested on bias-free simulated MRI data taken from BrainWeb with 3% noise level. Experiments were done on 61 slices (axial slices 60-120) that contain 93% of the lesion burden. Dice similarity coefficients were summarized in Table 2. Results from CGMM+CE and KVL algorithms were reported in Freifeld et al. [4]. Result from MLMR was calculated on the same data set using four-fold cross validation. It can be seen that MLMR achieved higher dice similarity coefficients, compared with the other two methods.

4.2 Real clinical data

Datasets of real MR images used for evaluation are publicly available from the MS Lesion Segmentation Challenge 2008 website [18]. There are 10 cases (subjects) provided by the University of North Carolina at Chapel Hill (UNC) with lesion labels publicly available. Lesion voxels were manually segmented by radiologists. For each case, three MRI modalities are available (T1, T2, and FLAIR). Data are re-sliced to

Table 2. Comparison of dice similarity coefficients across different methods from four-fold cross validation using one simulated brain volume by BrainWeb. Here, KVL, CGMM+CE and MLMR represent the algorithm developed by Van Leemput et al. [24], the constrained Gaussian mixture model with curve evolution, and our proposed multinomial logistic regression with Markov random field approach respectively.

KVL	CGMM+CE	MLMR
0.80	0.79	0.84

be $512 \times 512 \times 512$ with resolution $0.5mm \times 0.5mm \times 0.5mm$. MRI sequences of the data are co-registered.

4.2.1 Preprocessing

The data were preprocessed before performing lesion segmentation, including skull-stripping and bias field correction (Figure 8). The skull-stripping step is to extract the intracranial space from the image. Connected component

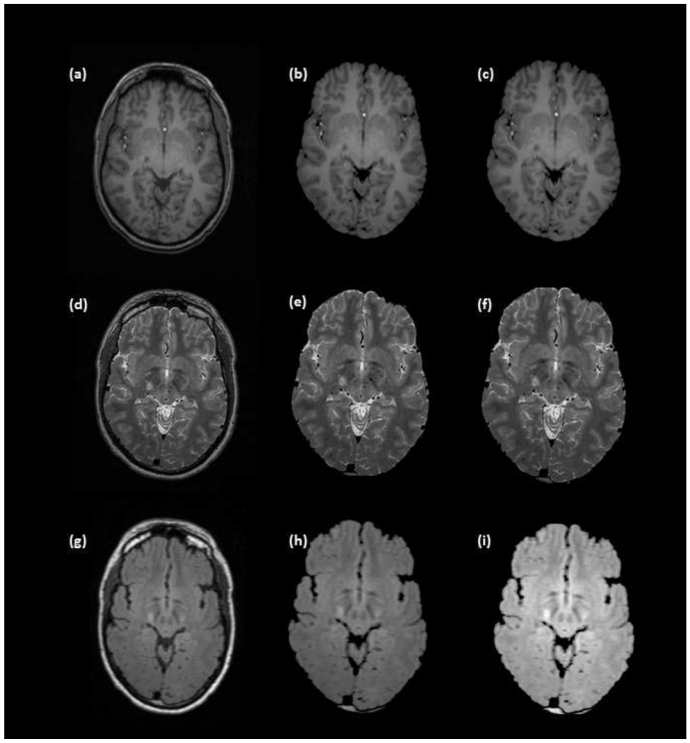

Figure 8. Preprocessing steps: (a) a given T1-w image slice, (b) skull-stripped T1-w image slice, (c) skull-stripped and unbiased T1-w image slice, (d) a given T2-w image slice, (e) skull-stripped T2-w image slice, (f) skull-stripped and unbiased T2-w image slice, (g) a give FLAIR image slice, (h) skull-stripped FLAIR image slice, (i) skull-stripped and unbiased FLAIR image slice.

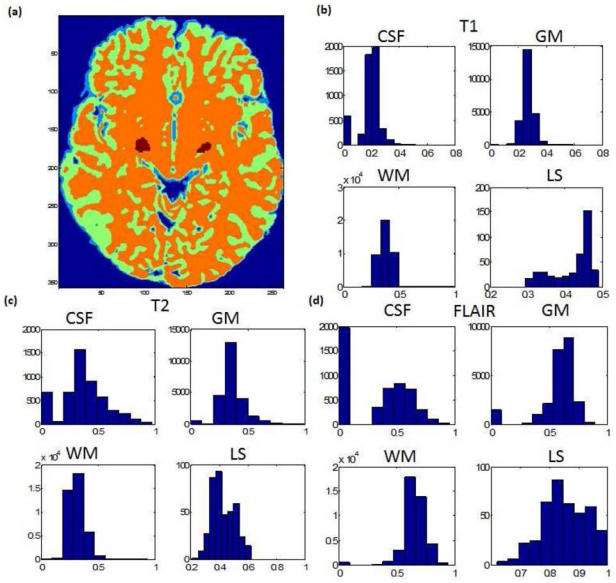

Figure 9. Preprocessing steps: (a) segmentation by fuzzy c-means combined with true lesion labels, (b) intensity histograms for four tissue types (CSF, GM, WM, LS) in T1-w images, (c) intensity histograms for four tissue types in T2-w images,(d) intensity histograms for four tissue types in FLAIR images.

labeling was used to identify the skull. A bias field is a low frequency spatial intensity variation that corrupts MRI images because of the inhomogeneities induced by the radio-frequency coil in MRI machines. Such inhomogeneities make intensity-based classification of MRI images difficult [1]. The bias field correction step is to correct the fact that two voxels belonging to the same tissue type may not have the same intensity. Lighting correction was applied to fit a linear plane to estimate the bias field. Intensity normalization was conducted by converting the units from each imaging modality into z scores based on the methodology developed by Shinohara et al. [16].

Intensity histograms for each tissue type under three MRI modalities are shown in Figure 9. For this data set, only lesion labels are provided. In order to plot the intensity distributions for all tissue types, a fuzzy c-means algorithm [1] was first applied to segment the image into WM, GM, CSF, and background. Lesion labels provided by the MS Lesion Segmentation Challenge 2008 website were then combined with the segmentation result from the fuzzy c-means algorithm (Figure 9). It can be seen that the intensity overlaps of these tissue types are much larger, compared with BrainWeb data (Figures 2, 3).

4.2.2 Lesion segmentation

Our proposed model was applied to the skull-stripped and unbiased image slices from all three modalities. Texture features including mean and variance of intensities, mean

Table 3. Comparison of lesion segmentation performance from different methods using five-fold cross validation. These methods are three variants of our proposed multinomial logistic regression with Markov random field (MLMR) approach.

Methods	Accuracy	Sensitivity	Specificity	DSC Lesion	DSC Others
MLMRF	0.99	0.33	0.99	0.47	0.99
MLMRL	0.98	0.59	0.98	0.25	0.99
MLMRLT	0.99	0.57	0.99	0.50	0.99

and variance of absolute values of intensity gradients, gray-level co-occurrence matrix based features (contrast, energy, and homogeneity from all four directions), and gray-level run length matrix based features (short run emphasis, long run emphasis, gray-level nonuniformity, run length nonuniformity from all four directions) with neighborhood sizes 5, 7, and 9 were calculated for T1-w and FLAIR images. T2-w images are not used for texture calculation because T2-w images do not provide as good tissue type contrast as T1-w and FLAIR images.

Five image slices were taken from UNC dataset case 1. We compared three variants of our proposed model MLMR on their lesion segmentation performance using five-fold cross validation. In the first approach (MLMRF), we applied our model multinomial logistic regression with Markov random field to fully combine T1, T2, and FLAIR intensity information from the whole image during training (Figure 10 part(a)). The second approach (MLMRL) used MLMR to locally combine T1, T2, and FLAIR intensity information from a block of the image containing lesion during training (Figure 10 part(c)). The third variant (MLMRLT) used the same intensity information as MLMRL but with extra texture information. Texture features were selected by the Elastic Net model during training stage. Cross validation was used to evaluate model performance. Segmentation results from the three approaches are shown in Figure 10 parts (d), (e) and (f). Compared with the true lesion region (Figure 10 part(b)), we can see that MLMRLT is a good compromise between MLMRF and MLMRL.

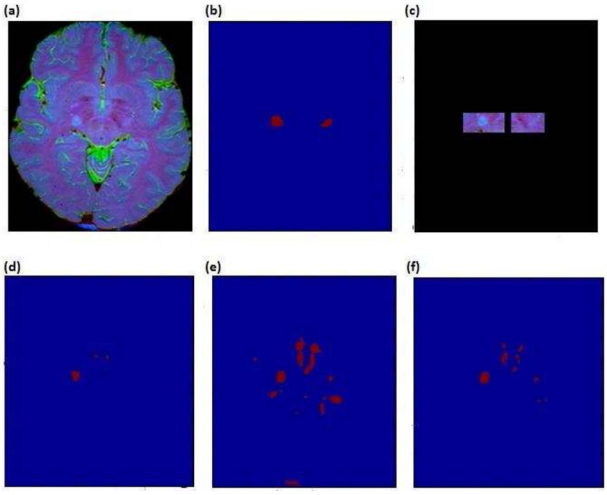

Figure 10. Lesion segmentation: (a) an image slice with combined intensities from T1, T2, and FLAIR modalities, (b) the true lesion region, (c) a block of the image containing lesion,(d) lesion segmentation by MLMRF, (e) lesion segmentation by MLMRL, (f) lesion segmentation by MLMRLT.

In addition, classification accuracy, specificity, sensitivity, and dice similarity coefficients are summarized in Table 3 based on five-fold cross validation. Comparing the first two approaches, we found MLMRF has larger dice similarity coefficients (DSC) for lesion segmentation while MLMRL has higher sensitivity to detect lesion (Table 3). By adding texture features, MLMRLT obtained higher DSC for lesion compared to MLMRL, while achieving higher sensitivity for lesion detection compared to MLMRF (Table 3). In the literature, it is reported that average sensitivity values are 0.42 ± 0.13 for the UNC data sets [22]. Hence, the MLMRLT model performed well for lesion detection.

Based on the above results, we then applied MLMRLT (multinomial logistic regression with Markov random field combining local intensity information and texture features) to all the image slices of the 10 labeled cases provided by UNC for the MS Lesion Segmentation Challenge 2008. For each case (subject), half of the data were used as the training set and the remaining half were used as the test set. As sensitivity and positive predictive value (PPV) were reported for the winner algorithm of the MS Segmentation Challenge 2008 [7, 17], we compared results from MLMRLT and the winner algorithm [17] using these two metrics (Table 4). PPV is defined as the proportion of predicted positive results that are true positives. As shown in Table 4, MLMRLT outperformed the winner algorithm for 8 of 10 cases. For the two cases (UNC04 and UNC09), MLMRLT achieved higher sensitivity but lower PPV than the winner algorithm. The average sensitivity value obtained by MLMRLT was 0.48 ± 0.14, better than the average values (0.42 ± 0.13) reported in the literature [22].

For texture information, features calculated from FLAIR images with neighborhood size 7 were most helpful to improve classification performance. In addition to the intensity information from T1-w, T2-w, and FLAIR images, the useful texture features included the average gradient magnitude, contrast, average long runs emphasis, run length nonuniformity at 90° and 135°. Figure 11 shows the local feature

Table 4. Comparison of our model MLMRLT with the winner algorithm of the MS Segmentation Challenge 2008

Methods	Winner algorithm		MLMRLT	
Patient	Sensitivity	PPV	Sensitivity	PPV
UNC01	0.01	0.01	0.40	0.41
UNC02	0.37	0.39	0.59	0.39
UNC03	0.12	0.16	0.26	0.37
UNC04	0.38	0.54	0.64	0.18
UNC05	0.38	0.08	0.50	0.23
UNC06	0.09	0.09	0.24	0.09
UNC07	0.57	0.18	0.61	0.21
UNC08	0.27	0.20	0.48	0.21
UNC09	0.16	0.43	0.56	0.32
UNC010	0.22	0.28	0.55	0.37

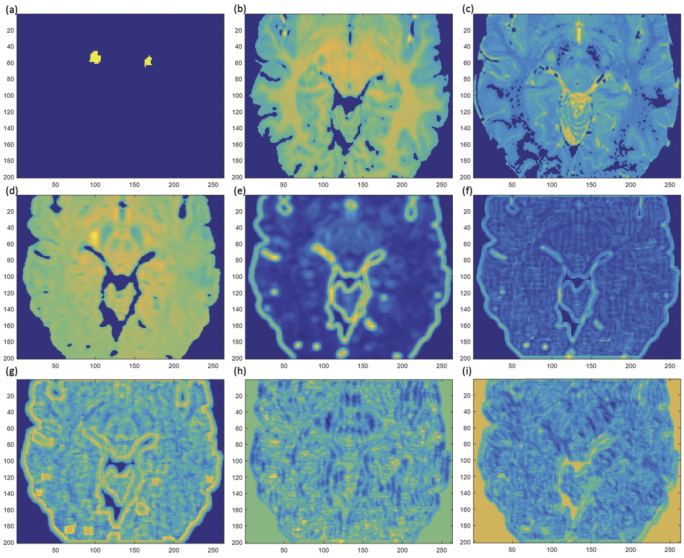

Figure 11. Local feature maps for lesion segmentation: (a) the lesion region segmented by a radiologist, (b) the T1-w image slice, (c) the T2-w image slice,(d) the FLAIR image slice, (e) the map of average gradient magnitude, (f) the map of contrast, (g) the map of average long runs emphasis, (h) the map of run length nonuniformity at 90°*, (i) the map of run length nonuniformity at* 135°*.*

maps. MS lesions are darker than the surrounding WM in T1-w images and the local maps of average gradient magnitude, contrast, and run length nonuniformity (Figure 11 parts (b), (e), (f), (h) and (i)) and are brighter than the surrounding WM in T2-w and FLAIR images and the local map of average long runs emphasis (Figure 11 parts (c), (d), and (g)).

5. DISCUSSION

In this study, we proposed a Bayesian framework for MS lesion segmentation that integrates the intensity information from multiple MRI modalities, the spatial information from neighborhood pixels, and the texture information. Results from both synthetic data and real clinical data show that the overlapping in intensity distribution of these different tissue types is not negligible. For example, the intensity overlaps between MS lesions and GM in T1-w MRI images and between MS lesions and CSF in T2-w MRI images are both substantial. Combining intensity information from different MRI modalities helps to better differentiate lesion regions from other tissue types. One innovation of this paper is to incorporate the inherent spatial information of pixels, which tend to encourage neighborhood homogeneity and reduce noise in classification. The smoothness parameter μ is a tuning parameter that balances the contribution of intensity information and spatial information to the model estimation. Furthermore, we examined a large number of texture features using the clinical data. Texture features derived from FLAIR images with neighborhood size 7 were found to be most helpful. In summary, both synthetic data and clinical data analyses have demonstrated the effectiveness of our proposed model.

In the analysis of clinical data, we observed that the number of lesion pixels is much smaller than the number of non-lesion pixels, which leads to imbalanced two classes. The class imbalance problem poses great challenges for classification models [20, 10]. To deal with the imbalance in the number of lesion and non-lesion pixels, we proposed to use local intensity information from a block of the image containing lesion during the training stage. This approach provides a good compromise between sensitivity and specificity. We compared the lesion segmentation performance of three variants of our proposed multinomial logistic regression with Markov random field approach. MLMRLT obtained higher DSC for lesion compared to MLMRL, while obtaining higher sensitivity for lesion detection compared to MLMRF. For future research, it is worth examining cost-sensitive classifiers for handling the class imbalance problem in lesion detection [11, 19]. As we put more penalization on false negatives than false positives, sensitivity is likely to be improved at the expense of specificity. An alternative approach to handle the class imbalance problem is to use re-sampling based techniques to create an artificially balanced training set, such as random oversampling and undersampling, informed undersampling, and adaptive synthetic sampling [10]. The Ising prior used in our model for capturing spatial information is a non-informative prior. Another interesting extension is to utilize a Markov random field that reflects the imbalance in the number of different tissue types. Additionally, future studies can consider to explore 3-dimensional (3D) texture features derived directly from 3D MRI data instead of 2-dimensional MR image slices.

ACKNOWLEDGMENT

We thank the reviewers for helpful comments that have improved this paper.

Received 3 April 2018

REFERENCES

[1] Ahmed, M., Yamany, S., Mohamed, N., Farag, A., and Moriarty, T. (2002). A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging* **21,** 193–199.

[2] Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20,** 1222–1239.

[3] Datta, S., Sajja, B., He, R., Wolinsky, J., Gupta, R., and Narayana, P. (2006). Segmentation and quantification of black holes in multiple sclerosis. *Neuroimage* **29,** 467–74.

[4] Freifeld, O., Greenspan, H., and Goldberger, J. (2009). Multiple sclerosis lesion detection using constrained gmm and curve evolution. *International Journal of Biomedical Imaging* **2009,** 13.

[5] Galloway, M. (1975). Texture analysis using gray level run lengths. *Computer Graphics Image Processing* **4,** 172–179.

[6] Garcia-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L. (2013). Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical Image Analysis* **17,** 1–18.

[7] Geremia, E., Menze, B. H., Clatz, O., Konukoglu, E., Criminisi, A., and Ayache, N. (2010). Spatial decision forests for MS lesion segmentation in multi-channel MR images. In Jiang, T., Navab, N., Pluim, J. P. W., and Viergever, M. A., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, pages 111–118, Berlin, Heidelberg. Springer Berlin Heidelberg.

[8] Haralick, R., Sharmugam, K., and Dinstein, I. (1973). Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **3,** 610–621.

[9] Harmouche, R., Collins, L., Arnold, D., Francis, S., and Arbel, T. (2006). Bayesian MS lesion classification modeling regional and local spatial information. *Proceedings of the International Conference on Pattern Recognition* **3,** 984–987.

[10] He, H. and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* **21,** 1263–1284.

[11] Hwang, J. P., Park, S., and Kim, E. (2011). A new weighted approach to imbalanced data classification problem via support vector machine with quadratic cost function. *Expert Systems with Applications* **38,** 8580–8585.

[12] Kroon, D., Oort, E., and Slump, K. (2008). Multiple sclerosis detection in multispectral magnetic resonance images with principal components analysis. *MS Lesion Segmentation (MICCAI 2008 Workshop)*.

[13] Kwan, R., Evans, A., and Pike, G. (1999). MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging* **18,** 1085–1097.

[14] Morra, J., Tu, Z., Toga, A., and Thompson, P. (2008). Automatic segmentation of MS lesions using a contextual model for the MICCAI grand challenge. In *MS Lesion Segmentation (MICCAI 2008 Workshop)*.

[15] Nyul, L., Udupa, J., and Zhang, X. (2000). New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging* **19,** 143–150.

[16] Shinohara, R. T., Sweeney, E. M., Goldsmith, J., Shiee, N., Mateen, F. J., Calabresi, P. A., and ... the Alzheimer's Disease Neuroimaging Initiative (2014). Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical* **6,** 9–19.

[17] Souplet, J., Lebrun, C., Ayache, N., and Malandain, G. (2008). An automatic segmentation of T2-FLAIR multiple sclerosis lesions. In *MIDAS Journal - MICCAI 2008 Workshop*.

[18] Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., and Warfield, S. (2008). 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. *MIDAS* pages 1–5.

[19] Sun, Y., Kamel, M. S., Wong, A. K. C., and Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition* **40,** 3358–3378.

[20] Sun, Y., Wong, A. K. C., and Kamel, M. S. (2009). Classification of imbalanced data: a review. *Internaltional Journal of Pattern Recognition and Artificial Intelligence* **23,** 687–719.

[21] Sweeney, E. M., Shinohara, R. T., Shiee, N., Mateen, F. J., Chudgar, A. A., Cuzzocreo, J. L., Calabresi, P. A., Pham, D. L., Reich, D. S., and Crainiceanu, C. M. (2013). OASIS is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *Neuroimage Clinical* **2,** 402–13.

[22] Tomas-Fernandez, X. and Warfield, S. (2011). A new classifier feature space for an improved multiple sclerosis lesion segmentation. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1492–1495.

[23] Valcarcel, A., Linn, K., Vandekar, S., Satterthwaite, T., Muschelli, J., Calabresi, P., Pham, D., Martin, M., and Shinohara, R. (2018). MIMoSA: An automated method for intermodal segmentation analysis of multiple sclerosis brain lesions. *Neuroimaging* in press.

[24] Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999). Automated model-based bias field correction of mr images of the brain. *IEEE Transactions on Medical Imaging* **18,** 885–896.

[25] Zijdenbos, A., Forghani, R., and Evans, A. (2002). Automatic pipeline analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE Transactions on Medical Imaging* **21,** 1280–1291.

[26] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67,** 301–320. MR2137327

Fengqing (Zoe) Zhang
Department of Psychology
Drexel University
3201 Chestnut Street
Philadelphia, PA 19104
USA
E-mail address: fz53@drexel.edu

Ji-Ping Wang
Department of Statistics
Northwestern University
2006 Sheridan Road
Evanston, IL 60208
USA
E-mail address: jzwang@northwestern.edu

Wenxin Jiang
Department of Statistics
Northwestern University
2006 Sheridan Road
Evanston, IL 60208
USA
E-mail address: wjiang@northwestern.edu