Author/s:
Mitra, J;Bourgeat, P;Fripp, J;Ghose, S;Rose, S;Salvado, O;Connelly, A;Campbell, B;Palmer, S;Sharma, G;Christensen, S;Carey, L

Title:
Lesion segmentation from multimodal MRI using random forest following ischemic stroke

Date:
2014-01-01

Citation:
Mitra, J., Bourgeat, P., Fripp, J., Ghose, S., Rose, S., Salvado, O., Connelly, A., Campbell, B., Palmer, S., Sharma, G., Christensen, S. & Carey, L. (2014). Lesion segmentation from multimodal MRI using random forest following ischemic stroke. Neuroimage, 98, pp.324-335. https://doi.org/10.1016/j.neuroimage.2014.04.056.

Persistent Link:
https://hdl.handle.net/11343/43940

Lesion segmentation from multimodal MRI using random forest following ischemic stroke

Jhimli Mitra, Pierrick Bourgeat, Jurgen Fripp, Soumya Ghose, Stephen Rose, Olivier Salvado, Alan Connelly, Bruce Campbell, Susan Palmer, Gagan Sharma, Soren Christensen, Leeanne Carey

PII: S1053-8119(14)00336-X
DOI: doi: 10.1016/j.neuroimage.2014.04.056
Reference: YNIMG 11312

To appear in: *NeuroImage*

Accepted date: 21 April 2014

Lesion segmentation from multimodal MRI using random forest following ischemic stroke[☆]

Jhimli Mitra[a,*], Pierrick Bourgeat[a], Jurgen Fripp[a], Soumya Ghose[a], Stephen Rose[a], Olivier Salvado[a], Alan Connelly[b], Bruce Campbell[c], Susan Palmer[b], Gagan Sharma[d], Soren Christensen[d], Leeanne Carey[b,e]

[a] *CSIRO Preventative Health Flagship, CSIRO Computational Informatics, The Australian e-Health Research Centre, Herston, QLD, Australia*
[b] *The Florey Institute of Neuroscience and Mental Health, Parkville, VIC, Australia*
[c] *Department of Medicine and Neurology, Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Parkville, VIC, Australia*
[d] *Department of Radiology, Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Parkville, VIC, Australia*
[e] *Department of Occupational Therapy, La Trobe University, Bundoora, VIC, Australia*

Abstract

Understanding structure-function relationships in the brain after stroke is reliant not only on the accurate anatomical delineation of the focal ischemic lesion, but also on previous infarcts, remote changes and the presence of white matter hyperintensities. The robust definition of primary stroke boundaries and secondary brain lesions will have significant impact on investigation of brain-behavior relationships and lesion volume correlations with clinical measures after stroke. Here we present an automated approach to identify chronic ischemic infarcts in addition to other white matter pathologies, that may be

[☆] The START Research Team (www.START.csiro.au)

[*] Corresponding author

Email addresses: jhimli.mitra@csiro.au (Jhimli Mitra), pierrick.bourgeat@csiro.au (Pierrick Bourgeat), jurgen.fripp@csiro.au (Jurgen Fripp), soumya.ghose@csiro.au (Soumya Ghose), stephen.rose@csiro.au (Stephen Rose), olivier.salvado@csiro.au (Olivier Salvado), spalmer1@gmail.com (Susan Palmer), Leeanne.Carey@florey.edu.au (Leeanne Carey)

used to aid the development of post-stroke management strategies. Our approach uses Bayesian-Markov Random Field (MRF) classification to segment probable lesion volumes present on fluid attenuated inversion recovery (FLAIR) MRI. Thereafter, a random forest classification of the information from multimodal (T1-weighted, T2-weighted, FLAIR, and apparent diffusion coefficient (ADC)) MRI images and other context-aware features (within the probable lesion areas) was used to extract areas with high likelihood of being classified as lesions. The final segmentation of the lesion was obtained by thresholding the random forest probabilistic maps. The accuracy of the automated lesion delineation method was assessed in a total of 36 patients (24 male, 12 female, mean age: 64.57±14.23 yrs) at 3 months after stroke onset and compared with manually segmented lesion volumes by an expert. Accuracy assessment of the automated lesion identification method was performed using the commonly used evaluation metrics. The mean sensitivity of segmentation was measured to be 0.53±0.13 with a mean positive predictive value of 0.75±0.18. The mean lesion volume difference was observed to be 32.32%±21.643% with a high Pearson's correlation of r =0.76 ($p < 0.0001$). The lesion overlap accuracy was measured in terms of Dice similarity coefficient with a mean of 0.60±0.12, while the contour accuracy was observed with a mean surface distance of 3.06mm±3.17mm. The results signify that our method was successful in identifying most of the lesion areas in FLAIR with a low false positive rate.

Keywords: Chronic stroke, ischemic infarct, white matter lesions, secondary lesions, FLAIR MRI, lesion likelihood, Markov random field, random forest

1. Introduction

The ischemic stroke lesion changes over time and secondary and remote changes may occur in response to this injury. These dynamic changes are reflected in MRI tissue contrasts, especially between acute (less than 7 days) and chronic ($\geq$ 3 months) stages (Carey et al., 2013b). It is therefore unlikely that a single MR parameter can fully characterize the complexity of the tissue changes post-stroke (Baird and Warach, 1998; Welch et al., 1995; Knight et al., 1994; Vannier et al., 1985). In clinical practice, diffusion weighted images (DWI), T1-weighted (T1W), T2-weighted (T2W) and fluid attenuated inversion recovery (FLAIR) images are often acquired to monitor the progression of stroke. In the acute stage, hyperintense signal observed on DWI provides important information about the anatomical location and extent of the infarcted territory (Rivers et al., 2007; Carey et al., 2013b). In the more chronic phase, T2W and FLAIR images are normally used to delineate the final lesion volume (Xavier et al., 2003). Chronic ischemic lesions appear as hyperintense regions in FLAIR with some heterogeneity within the lesion volume due to ongoing gliosis and demyelination (Cramer et al., 2006; Clark et al., 1993). In some patients, infarct delineation may be complicated by the presence of remote, asymptomatic lesions in addition to the primary lesion of interest. In fact silent brain infarcts and white matter (WM) lesions are common in healthy elderly people (Vermeer et al., 2003; Norrving, 2008) with an estimated 7%-28% of the patients with known stroke history and aged over 65 years having evidence of silent lesions (Vernooij et al., 2007; Bernick et al., 2001). Further, white matter changes may be observed in late stages post-stroke (Lindenberg and Seitz, 2012). Therefore, patients with clinically

first-ever stroke may also have multiple lesions either in one or both hemispheres. Estimation of the ischemic lesion, secondary lesions, and remote changes are likely important to appreciate the full impact of the brain injury on functional outcome and recovery after the primary stroke (Carey et al., 2013b).

White matter lesions or hyperintensities (WMH) are observed as hyperintense regions in FLAIR when compared to the surrounding WM tissue. WM lesions are associated with a number of neurological disorders, including ageing (Debette and Markus, 2010; Cavalieri et al., 2010; Yamauchi et al., 2002; Park et al., 2010; Launer, 2004; O'Sullivan, 2008; Silbert et al., 2008; Teodorczuk et al., 2010). Despite the probable differences in the underlying tissue pathologies of ischemic strokes and WM lesions, the tissue textures appear similar in T1W, T2W and FLAIR images; i.e. they are hyperintense in T2W and FLAIR and hypointense in T1W images. Adding to the complexity of delineation is the location of the infarct, for instance when the ischemic lesion occurs within deep WM territories that are anatomically coincident with the presence of WM hyperintensities. Differentiating between the two types of lesions is extremely difficult using FLAIR, and even more so in the absence of acute/sub-acute DWI images (Campbell et al., 2012). A number of examples highlighting the complex tissue contrast associated with chronic stroke are given in Figure B.1. In these cases, the primary stroke lesion exhibits both hyperintense (inflammation) and hypointense (tissue atrophy) along with the presence of other WM pathologies on FLAIR images.

Robust automated techniques that can segment the chronic ischemic lesions, WM and other secondary lesions from normal brain tissues offer a

streamlined approach to investigate important structural-functional relationships following stroke (Carey et al., 2013a). A variety of automated and semi-automated methods have been proposed for ischemic lesion detection at the acute and sub-acute stages of stroke. Segmentation of intensity histograms followed by region growing and decision trees (Heinonen et al., 1998) was proposed by Dastidar et al. (2000) with extensive manual intervention. Clustering based methods like ISODATA (Iterative Self-organizing Data), mean-shift procedure, and fuzzy C-means have been proposed by Jacobs et al. (2001b,a); Hevia-Montiel et al. (2007); Shen et al. (2010); Seghier et al. (2008); Wilke et al. (2011) who either considered edge-confidence maps or prior tissue probabilities to drive the automated segmentation. Markov random field (MRF)-based automated segmentation from multimodal MRI volumes was suggested by Kabir et al. (2007). Similarly several automated methods have been proposed to segment WM lesions (Jack et al., 2001; Mohamed et al., 2001; Wei et al., 2002). Pattern classification methods involving either K-Nearest Neighbor (KNN), Support Vector Machines (SVM) and Ada Boost have been proposed by Anbeek et al. (2004b,a) and Lao et al. (2008). Kruggel et al. (2008) proposed a texture based segmentation of WM lesions using multi-sort co-occurrence matrices (Kovalev et al., 2001) and principal component analysis. Both the methods of (Anbeek et al., 2004b,a) and (Kruggel et al., 2008) provided probabilistic segmentation of the lesions. Yang et al. (2010) proposed an automated segmentation method using the method proposed by van Leemput et al. (2001) combined with random field theory. A multi-level Expectation-Maximization (EM) algorithm was proposed by Wang et al. (2012) and a fuzzy inference-based system aided with

prior tissue probabilities was proposed by Admiraal-Behloul et al. (2005) to segment WM lesion loads. Recently, Ong et al. (2012) proposed an outlier detection approach based on adaptive trimmed means algorithm and box-whisker plot on intensity histograms.

In general, the methods show good accuracy in terms of volume correlation and visual agreement. However, most focus on delineation of either the primary stroke lesion or the WM lesions, despite evidence and clinical significance of secondary abnormalities. A common reported difficulty associated with lesion segmentation was the challenge to reduce false positives from the FLAIR and T1W MRI due to heterogeneity in tissue contrasts. For ischemic lesion detection, most of the available methods utilize DWI images to help define the anatomical location of the initial infarcted territory acutely post-stroke. These hyperintense regions easily target and isolate the ischemic stroke from other brain pathologies, they are however not always available, especially in the late stages of recovery.

In this paper we investigate the challenges of intensity heterogeneities within the lesion areas, and aim to develop a method to segment stroke with minimized detection of WMH due to normal-ageing relying more on statistical concepts of randomness and information theory that allow the amount of heuristic decisions to be minimized without requiring acute DWI MRI. As outlined in Figure B.2, we have employed the methods of Expectation Maximization (EM) likelihood estimation, Bayesian-Markov random field (MRF) segmentation and random forest classification to segment the lesion from the multimodal MRI (FLAIR, T1W, T2W) and apparent diffusion coefficient (ADC) maps obtained from the DWI of the chronic stage. Since the primary

lesions are best identified in FLAIR with hyperintense regions, we apply the initial Bayesian-MRF classification on the FLAIR images. Having the probable lesion areas identified from the Bayesian-MRF classification, we refine the segmentation using random forest on the multimodal data with spatial and context-rich features. Random forest classification has been successfully used by Geremia et al. (2011) for multiple sclerosis (MS) lesion segmentation and by Zikic et al. (2012) for identifying regions of high-grade glioma.

The key areas of novelty in our method are:

- the multilevel splitting and merging of Gaussian intensity classes using Bayesian/MRF classification to extract probable lesion areas,
- probabilistic classification using random forest of the probable lesion areas, and

In this study, we developed our segmentation framework using MRI data acquired from 36 chronic stroke patients and compared the accuracy of the technique against lesion volumes manually defined by a stroke expert. As our technique is also based, in part on random forest classification as in (Geremia et al., 2011), we further assessed the performance of our method using the data available from the MS segmentation lesion challenge 2008 (www.ia.unc.edu/MSseg/).

2. Materials and Methods

2.1. Patients

MRI data from 36 patients was used to develop our automated lesion segmentation strategy. Thirty-four patients had radiological evidence of a

primary ischemic infarct consistent with their clinical history and presentation, while two other patients did not have evidence of a primary stroke, as determined by an expert neurologist, at 3 months after stroke. Most patients (n=30) contained additional WM lesions, whilst 7 subjects also exhibited evidence of a previous lesion. The patient demographics are summarized in Table B.1. The precise anatomical locations of the stroke for some patient are given within the supplementary material (Figure 8).

Acquisition and manual segmentation

MRI scans namely, T1W, T2W, FLAIR and DWI were acquired at 3 months post-stroke using a 3T MR (Magnetom Trio; Siemens, Erlangen, Germany) scanner. The T1W images were acquired with MPRAGE sequence with the following parameters; TE=2.55ms, TR=1900ms, flip angle=9°, voxel resolution=1mm × 1mm ×1mm. The T2W FLAIR images were acquired with the following parameters; TE=388ms, TR=6000ms, T1=2100ms, flip angle=120° with voxel resolution=1mm × 0.5mm × 0.5mm. The ADC maps are generated from DWI images acquired using a spin echo diffusion-weighted echoplanar imaging (EPI) sequence with b-value=1000 employing 25 diffusion encoding directions with isotropic voxel resolution of 2.5mm. Lesion volumes were manually segmented by an expert neurologist on multiplanar reprojections of the FLAIR scans with voxel resolution of 3mm×0.5mm×0.5mm using MINC (Medical Image NetCDF) Tools by the McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University for display. The segmentation was fully manual for the primary stroke and secondary lesions, while WM lesions were segmented based on threshold.

Data pre-processing

The T1W images were rigidly registered to the (Montreal Neurological Institute) MNI atlas (1mm isotropic), and all other modalities were rigidly registered to the T1W in MNI space. All analyses were performed in the MNI space. The method of van Leemput et al. (1999) was used to obtain the segmented gray matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) masks. The GM and WM masks were obtained from the T1W scans, while the CSF mask was derived from the T2W MRI scans. The method also provided bias-corrected images when applied to T1W and FLAIR. Magnetic bias correction is particularly important for the Bayesian-MRF based multi-level segmentation to extract the probable lesion areas while bias-corrected intensities are not required for random forest training since context-aware features are of more importance; this will be discussed in more details in the subsequent sections. Skull stripping was performed by applying the GM, WM and, CSF segmented masks to remove the remaining part of the brain. The probabilistic GM and WM estimates for each patient are obtained by non-rigidly aligning (Modat et al., 2010) the GM/WM masks of other patients to the target patient (leave-one-out manner); and then averaging the aligned masks. This method ensures that the mean errors in the segmentation masks due to atrophies are minimized.

2.2. Automated segmentation

Our proposed approach comprises the following steps:

- Hierarchical segmentation of the probable lesion class from the FLAIR MRI

- random forest training and probabilistic classification on context-aware features of multimodal MRI; GM, WM, and lesion likelihood maps,

Figure B.2 shows the schematic diagram of our method.

2.2.1. Hierarchical segmentation of probable lesions from FLAIR

The Bayesian-MRF segmentation involves a two-level classification. The initial classification made an assumption of 5 Gaussian classes as background, GM, WM, CSF and an extra lesion class from the FLAIR intensities. A second level Bayesian-MRF segmentation of the FLAIR intensities of the extra-lesion class into 8-classes further reduced the intra-class variabilities. The 3-classes with highest means were then combined to form one probable binary lesion class that approximated the stroke and WM lesions more appropriately.

Bayesian inference and Markov Random Field (MRF) segmentation. Given $\mathcal{F} = \{f_s\}_{s \in S}$ a set of image data where f_s denotes the gray value at pixel s. A segmentation problem is to find the labeling $\hat{\omega}$, $\omega \in \Omega$, where Ω is the set of all possible discrete labellings which maximizes the posterior probability $P(\omega|\mathcal{F})$. The MRF on each of the pixel s is defined by the clique potentials and by a neighborhood-system $\mathcal{G}$. Let C denote a clique of $\mathcal{G}$, and $\mathcal{C}$ the set of all cliques. The restriction of the label ω to the site of a given clique C is denoted by ω_C. The clique potentials are given by $E_C(\omega_C)$ for every C in $\mathcal{C}$ and every $\omega \in \Omega$. Following the definition of Bayes theorem (Bayes and Price, 1763) and assumption of Gaussian ($\mathcal{N}(\mu_\lambda; \sigma_\lambda)$) discrete class λ, the

global labeling which we are trying to find is given by:

$$\hat{\omega} = \min_{\omega \in \Omega} \left(\sum_{s \in S} \left(\log(\sqrt{2\pi\sigma_{\omega_s}}) + \frac{(f_s - \mu_{\omega_s})^2}{2\sigma_{\omega_s}^2} \right) + \sum_{C \in \mathcal{C}} E_C(\omega_C) \right) \quad (1)$$

The estimation of $\hat{\omega}$ was done through the energy minimization using the Iterated Conditional Modes (ICM) algorithm (Besag, 1986). The initial estimates of the ICM method were obtained from the maximum aposteriori (MAP) estimate of the Bayes theorem considering uniform priors for all classes. The final probable binary lesion class after the two-level Bayesian-MRF segmentation was further approximated with EM algorithm (Dempster et al., 1977) into a lesion and non-lesion class and the likelihood of the lesion class was used for random forest training and classification. A detailed step-wise formulation of the Bayesian-MRF framework was shown in (Fischl et al., 2002).

2.2.2. Random forest training and probabilistic prediction

Our problem of lesion detection may be formalized as a binary classification of the probable lesion areas into non-lesion and lesion regions. A supervised method employing random decision forest was used to address our problem. Random decision forest use decision trees which are discriminative classifiers but are not known to suffer from over-fitting (Breiman et al., 1984). A random forest grows an ensemble of independent decision trees built on random subsets of the training data and further randomizing the feature set at each node (Yi et al., 2009). A standard classification forest was used in this work where the training data consisted of a feature set of labeled voxels $v_k \in \mathbb{V}$ with associated label $Y(v_k)$. The forest was composed of T trees with t indexing each tree. During training, the full feature set $\mathbb{V}$ was pushed

through each of the trees. Each node p received a random partition of feature set $\mathbb{V}_p$. An optimal test threshold was found by exhaustive search over the randomly sampled feature space and uniformly discretized threshold space to split the feature set $\mathbb{V}_p$ into the left and right child nodes respectively. The test threshold was optimized by maximizing the information gain. Trees were grown to a maximum depth D which can be full-grown trees or until the information gain is below a minimal value. At the end of the training process, each leaf node l_t contained a class predictor as $P_{l_t}(Y(v) = c) = \frac{\left|\mathbb{V}_{l_t}^c\right|}{\left|\mathbb{V}_{l_t}\right|}$, which was the fraction of points of each class $c \in \mathcal{C}$ in the $\mathbb{V}_{l_t}$.

For a new test volume each voxel of feature v_k was propagated through all trees T by successive application of the learned split functions. After reaching a leaf node l_t in all trees $t \in [1 \dots T]$, the posteriors $P_{l_t}(Y(v) = c)$ of belonging to a class $c \in \mathbb{C}$ were gathered in order to compute the final posterior probability $P(Y(v) = c) = \frac{1}{T}\sum_{t=1}^{T} P_{l_t}(Y(v) = c)$, which was the mean over all the trees in the forest. This probability may be thresholded in a number of ways to achieve a binary segmentation. In this work, we preferred to keep a probabilistic classification and therefore a posterior probability map P_c was generated by applying the same prediction procedure to all the voxels.

Features in random forest. In our work, local features, spatial features and context-rich features were derived from the input MRI channels (T1W, T2W, FLAIR, ADC map), probabilistic estimates of the GM (P_{GM}) and WM (P_{WM}) and the likelihood of the probable lesion class (P_{les}) (obtained from the EM estimate of the probable binary lesion class). Therefore, an input

channel comprises of

$$I_j = (I_{FLAIR}, I_{T1}, I_{T2}, I_{ADC}, P_{GM}, P_{WM}, P_{les})$$

The intensities across each channel were normalized. Let us consider $x \in \mathcal{X}$ as a spatial point, to be assigned a class (lesion/background). $N_j^{\mathcal{D}}(x)$ denotes an x-centered and axis-aligned 3D cuboid with edge lengths $\mathcal{D} = (\mathcal{D}_u, \mathcal{D}_v, \mathcal{D}_w)$, $v \in \mathbb{R}^3$ as an offset vector and $d = (u, v, w)$ as the spatial location of x. The local and spatial features were defined as

$$f(x, I)_j = I_j, \quad \forall j \tag{2}$$

$$f(x)_d = d. \tag{3}$$

The three other context-rich features were constructed similar to Zikic et al. (2012) and are described in Appendix A.

2.3. Quantitative evaluation metrics

Our proposed method and its variations to segment ischemic infarcts were evaluated using the standard voxel-based and surface-based evaluation metrics and are detailed in Appendix B. The evaluation criteria for the MS Lesion Segmentation Challenge (2008) however, were based on the number of lesions as described in Styner et al. (2008).

2.4. Statistical analysis

We have compared the proposed method that uses random forest probabilities (acronym: RF) with two variations which are intuitive considering the same post-processing steps were applied for each variant. The method variants are as follows:

- The Bayesian-MRF two-level segmentation (acronym: MAP-MRF)
- the likelihood posterior probabilities from EM, thresholded above ν (acronym: LIK). This would justify the use of random forest over Bayesian-MRF segmentation.

In order to compare the performance of the proposed method with other proposed variations, we need to measure the statistical significance of the null hypothesis; i.e., that the means of the quantitative evaluations for the compared methods are not different. Therefore, we also need to verify the normality and homogeneity (homoscedasticity) of the variances of the data (Sokal and Rohlf, 1995). The composite normality of the data is tested using the Lilliefor's test (Lilliefors, 1967). If the null hypothesis is rejected then common data transformations are applied. Subsequently Bartlett's test (Bartlett, 1937) was used to test the null hypothesis that the distributions have equal variances. If the distributions are homoscedastic, then a two-tailed Student's t-test (Gosset, 1908) is performed to test the equality of the mean ranks of the pair-wise compared methods. Otherwise, if the distributions show heteroscedasticity, then a two-tailed Welch's t-test (Welch, 1947) is performed to test the same null hypothesis.

3. Experiments and results

3.1. Parameters and implementation

During the random forest training process, randomly-sampled training samples were used to train the forest to avoid over-fitting (Criminisi et al., 2011). The edge length $\mathcal{D}$ for extracting context-rich features was fixed at

3mm and the offset υ was chosen as 8mm empirically. The random forest was trained with 100 trees to a maximum depth of 25 restricting the number of samples per leaf node to be 2 samples and information gain of 10^{-6}. The validation for random forest was done using a 4-fold validation method. All our implementations were done in C++/OpenCV and ITK platform on a 6-core CPU of 3.2GHz with 23.5GB of memory. The Bayesian-MRF classification required 6min−7min on an average for each patient, the random forest training on a single-core CPU (non-parallel implementation) required an average of 3 hours to train in each fold of the validation method, while each testing required 7min−10min (including feature extraction time). The threshold ν for binary classification was experimentally found to be 0.2. The post-processing steps had been consistent across all the compared methods and involved removal of false positive lesions below ϵ=10 voxels and a morphological closing operation with a circular mask of radius 2 to include hypointense areas within chronic strokes.

3.2. Results on patient cohort

The proposed method has been evaluated on 36 patient datasets. Figure B.3 shows the likelihood and random forest probabilistic segmentations. We observe that the probabilities from the likelihood (Figure 3(b)) show noisy probabilistic regions, while random forest probabilities exhibit high confidence in stroke or secondary lesion areas (Figure 3(c)). This justifies that context-aware features of random forest help to improve the probabilities.

Table B.2 shows the quantitative results for the 36 patients using the standard evaluation metrics. The maximum DSC value achieved in our method is 0.82. The lesion volume correlation (Pearson's correlation) of the segmented

lesion by the proposed method when compared against the ground truth is r =0.76 ($p < 0.0001$) with a relative volume difference of 32.32%±21.64% and a low mean SMAD of 3mm.

An example showing the effects of processing a patient case (patient 7) with the steps described in Sections 2.2.1 and 2.2.2 respectively, is given in Figure B.4. It is observed that the two-level Bayesian and MRF segmentation clearly segmented the hyperintense lesions as shown in Figures 4(b) and 4(c). Splitting the lesion output of the initial 5-class estimation further into 8-Gaussian classes clearly aided to model the heterogeneities in the lesion class itself. Although, to reduce areas of false positives in GM regions, not more than 3 classes with high means could be considered to approximate the probable lesion class (Figure 4(d)). The likelihood of the probable lesion class from the EM estimation showed high probabilities in lesion areas and lower probabilities in non-lesion areas. However, the likelihood estimation being noisy (Figure 4(e)), the probabilistic information was further used in the random forest training to provide probabilities of lesion areas with high confidence. The qualitative evaluation of the segmented lesion region (Figure 4(f)) showed a good agreement with the expert ground truth (Figure 4(g)).

Table B.3 shows the statistical significance (p-values with 95% confidence) of the measured mean DSC, PPV, TPR, SMAD and VD values of the other methods when compared with the proposed method. Analyzing the p-values of Table B.3, we observed that the means of most of the evaluation metrics of the proposed method RF and other methods (MAP-MRF and LIK) were significantly different ($p < 0.05$). Therefore, from the analysis we may likely infer that our proposed method of random forest significantly improved the

lesion approximations over two-level Bayesian-MRF and lesion likelihood.

In Figure B.5, the results of the proposed method and other compared methods can be seen for patient 20. It is observed from Figure 5(d) and Figure 5(e) that both the methods MAP-MRF and LIK either under-estimated or over-estimated the lesion areas compared to the ground truth (Figure 5(g)). The proposed method RF (Figure 5(f)) showed good approximations of the lesion area. It is also observed in Figure 5(b) that the WM lesion in the right hemisphere below the periventricular region, showed lower probabilities, but the same area has higher confidence in the RF probability map (Figure 5(b)), therefore random forest better estimated the actual lesion area.

Figure B.6 shows some qualitative results of lesion segmentation using our proposed method for a critical analysis of the quantitatively evaluated measures in Table B.2. Patient 25 shows inferior evaluation measures with respect to the quantitative evaluation metrics DSC, PPV, TPR and SMAD, while a low VD as observed in Table B.2. Consistent with this, in, row 1, column 3 of Figure B.6 we observe false positives when compared to the ground truth in row 1 column 4. This is likely due to the intensity heterogeneities as observed in the FLAIR and T1W images of the first 2 columns. It is worth noting, that even if the DSC, TPR and PPV show average performance values, the VD may be quite low since the false positive voxels may add up with the true positive voxels to reduce the difference between the segmented and manual ground truth volumes. Results of patients 23 and 16 in rows 2 and 3 respectively show that segmented lesion areas were under-estimated by the automatic method compared to the ground truth. This was because the lesions contained large necrotic hypointense regions in FLAIR which could

not be approximated even after morphological post-processing. This fact was also evident in the quantitative evaluation scores of PPV and TPR for the respective patients (Table B.2), where a high PPV and a low TPR indicated that the lesions were not completely approximated but less false positives were observed. For patient case 2 in row 4, an over-estimation of ground truth was observed (column 4) which resulted in a low TPR value even if the DSC and PPV values were acceptable. This kind of over-estimation was sometimes observed since the ground truth accurately marked by an expert on low-resolution images was upsampled to isotropic voxels.

3.3. Results on MS lesion segmentation challenge (2008) data

To evaluate the performance of our method in detecting MS lesions which appear as hyperintensities on T2W and FLAIR images and around the periventricular areas, we used the publicly available dataset of the MS lesion segmentation challenge 2008. The detailed results and comparison with some of the best performing algorithms in the challenge are given in the supplementary file in Tables 1 to 5 respectively.

4. Discussions

In this paper, we have proposed an automated segmentation method to identify ischemic, WM and other secondary lesions in chronic stroke from multimodal brain images using random forest after initial screening of probable lesion areas from the FLAIR images. The proposed method aimed to segment lesion areas from the FLAIR images obtained at 3-months post-stroke using a Bayesian and MRF classification. The assumption of 5 Gaussian classes and modeling FLAIR lesion intensities as an extra class has been

applied previously by Jacobs et al. (2001a); Seghier et al. (2008), providing support for our choice of the number of Gaussian classes in the initial stages of our approach. There are however, some key differences between our method and some existing methods detecting stroke and WMH. Bayesian-based tissue segmentation methods generally rely on tissue priors and likelihoods to estimate the posterior distribution. In comparison, in our case, we could only estimate the likelihoods of tissue classes using the EM method, but spatial priors especially for stroke lesions could not be derived. The method of Wang et al. (2012) attempted to separate WMH and cortical infarcts (CI) using multilevel EM approach. However, their method applied morphological operations at each stage of their processing pipeline and made an assumption that strokes appear only along the cortical regions and extend to WM regions. These assumptions reduced the suitability of their method to be applied to our stroke cohort since we have repeated evidence in our dataset that strokes also appear in the WM regions and along the periventricular area with similar appearances as WM hyperintensities. MS lesions are usually present bilaterally, though not exactly symmetrically with periventricular WM being often extensively affected Love (2006). Therefore, probably with the assumption of such bilateral evidences, random forest used by Geremia et al. (2011) to segment MS lesions used symmetric constraints for feature extraction during the training phase. Also a sub-sampling of the images was done by (Geremia et al., 2011) along the axial direction. Both these assumptions made by (Geremia et al., 2011) were inappropriate in our case, since stroke lesions are not associated with any specific location in the brain and sub-sampling the images would result in missing stroke lesions in

the cerebellum regions for our patient cohort.

The primary difference between our method and other methods that use random-forest for tumor or MS lesion segmentation was the use of the multilevel Bayesian-MRF to estimate the initial lesion class prior to the random forest. Most methods Geremia et al. (2011); Zikic et al. (2012) used only random forest as the segmentation algorithm i.e., during the training phase, data were sampled only from the lesion regions available from the ground truth, with some offset around the regions to maintain a balance between the positive and negative classes; and during testing, either the whole brain or randomly sampled regions were passed into the forest. On the contrary, the use of MRF before random forest classification allowed informative sampling of the lesion class both during training and testing phases. All voxels with FLAIR hyperintensities similar to the ischemic lesions were localized with MRF. We observed that no suspicious lesions were exempted using the Bayesian-MRF framework (as seen in Figure 5(d)) and the number of samples sent for subsequent classification with the trained random forest was significantly reduced. Initial Gaussian modeling and sampling from probable lesion voxels also reduced the chance of mis-classifications in the testing phase. The effect of having a Bayesian-MRF stage prior to random forest training was evident in the results of the experiments performed on MS lesion segmentation challenge (2008) data (in supplementary material). We observed that a higher number of actual lesions were identified by our method when compared to (Geremia et al., 2011) shown in Table 4 and Table 5 respectively with comparable false positive lesions; although not exceeding the overall segmentation accuracy of the latter.

In our experiments with the stroke cohort, it was clear that the multilevel Bayesian-MRF based segmentation was not enough and therefore, random forest was used to refine the segmentation to reduce the false positives as shown in Figure 5(d) and Figure 5(f). Although the choice of the number of classes in the second stage of Bayesian-MRF segmentation was based on the experimental data, a robust method can be adopted by spectral clustering of the FLAIR intensity histograms of the lesion class obtained from the first level of segmentation. The quantitative evaluation of our proposed method of performing segmentation showed that the results were similar to those of (Seghier et al., 2008) (mean DSC of 0.64 ± 0.1 with maximum of 0.81 for 10 cases) to segment any type of brain lesion. We do however acknowledge that comparison of quantitative results is only meaningful when the same datasets are used to evaluate different methods or a method is publicly available to test on different data.

We have seen that a probabilistic segmentation is more meaningful when characterizing lesions of heterogeneous intensities as shown by Anbeek et al. (2004b). Nevertheless, a quantitative measure of lesion volumes is necessary to evaluate the presence of lesions in clinical analysis and therefore a trade-off between a threshold value and good lesion approximation is required. We have also seen that lesion-likelihood maps used as a channel in random forest provided better lesion approximation of lesion volumes rather than lesion likelihood maps being thresholded or only Bayesian-MRF segmentation.

Previous studies have shown that brain atrophy is more closely aligned with cognitive status and disability than the volume of typical ischemic lesions alone (O'Sullivan et al., 2004; Peters et al., 2006). Thus the measure of

tissue loss should also be considered when assessing important structure-function relationships after stroke. This is a limitation of our proposed method where the tissue loss specifically due to stroke (hypointense tissue contrast in FLAIR) could not be quantified. We have observed that large necrotic regions, including areas of tissue loss within the stroke territory, could not be separately segmented using our proposed method since the initial screening of lesion areas relies on FLAIR hyperintensities. A similar problem was observed for lacunar strokes. We are continuing our research to address this limitation and may be tackled in one of the following manners: either the DWI at the acute stage can be used to isolate the ischemic lesion areas, and for the remaining lesions the FLAIR hyperintensities may be used; or separate ground truth for the necrotic and peri-necrotic stroke lesion areas could be employed to train the random forest for a multi-class stroke lesion segmentation as performed by Zikic et al. (2012) for high-grade glioma. It was suggested by Alexander et al. (2010) that clear boundaries between the necrotic, peri-necrotic and hyperintense ischemic lesion areas are difficult to identify. Although the quantification of tissue loss due to ischemic stroke is necessary for correlative studies focused on investigating post-stroke behavorial-functional associations, it may not be as critical to separate the cause of the tissue loss in the chronic phase of recovery. The assessment of tissue loss utilizing measures of CSF difference between the ipsilesional and contralesional hemispheres is part of future studies.

Age related WMH that appear around the periventricular regions exhibit similar hyperintensities as ischemic lesions on T2/FLAIR, and therefore are indistinguishable from stroke using our method. As discussed before in Sec-

tion 1, the separation between chronic-stage ischemic stroke and age-related WMH is only clinically possible in presence of acute-stage DWI that clearly characterize the stroke lesions. Also, estimation of the ischemic lesion, secondary lesions remote changes and white matter changes are likely important to appreciate the full impact of the brain injury on functional outcome and recovery after the primary stroke Carey et al. (2013b). If however, distinguishing between various lesion types is required in the absence of DWI, improved multimodal texture features using Gabor filters and local binary patterns may be employed and investigation of such methods is within the scope of our future studies. Finally, our method was developed based on a small patient cohort and therefore needs to be validated on prospective stroke data.

5. Conclusions

In this paper, we presented an automated method to segment ischemic, WM and other secondary lesions, that involves strategic combination of state-of-the-art methods including Bayesian and MRF segmentation; and random forest classification. We have shown how the different stages of the method aid in segmenting the lesion areas out of the heterogeneous intensities of the multichannel MRI. Incremental improvements have been observed through the subsequent stages and qualitative results have justified our strategy. The proposed method when compared to some of the most intuitive variations, demonstrated high metrics of similarity, positive prediction and true positive values relative to the ground truth. Overall these metrics were shown to be significantly different for our method than the remaining methods. The

incorporation of context-rich features of multichannel MRI in random forest over intensity-based Bayesian and MRF classification on single modality FLAIR image has been qualitatively and statistically justified. Given the complexity of stroke pathology and considerable overlap with existing WM disease, we believe the described method show promise for robust, automated segmentation of lesion volumes.

Future work could include incorporation of more discriminative features in random forest and developing a multi-class classifier to segment the ischemic, WM and other secondary lesions as separate classes that would also be meaningful for the clinical evaluation of post-stroke recovery.

Acknowledgments

We would like to acknowledge the Stroke Imaging Prevention and Treatment (START) program of research which is supported in part by the CSIRO of Australia through the Preventative Health Flagship Cluster, the National Health and Medical Research Council of Australia, and a Victorian Government Operational Infrastructure Support Grant. In particular, we wish to acknowledge the stroke patients, radiologists and START researchers who contributed to the data collected for this study. LC is supported by an Australian Research Council Future Fellowship [number FT0992299]. The funding sources had no role in conduct of the study or writing of the report.

Appendix A. Context-rich features in random forest

The context-rich features were defined as in (Zikic et al., 2012). These were derivative-like features across different channels, offsets and neighbor-

hoods. The first feature was the intensity difference between x in channel I_{j1} and an offset at point $x + \upsilon$ in a channel I_{j2}, where $I_{j1} = I_{j2}$ was allowed.

$$f(x, I)_{j1,j2,\upsilon} = I_{j1}(x) - I_{j2}(x + \upsilon). \tag{A.1}$$

The second feature measured the difference between the intensity means of a cuboid around x in I_{j1} and that around an offset point $x + \upsilon$ in I_{j2} was given by

$$f(x, I)_{j1,j2,\mathcal{D}1,\mathcal{D}2,\upsilon} = \mu(N_{j1}^{\mathcal{D}1}(x)) - \mu(N_{j2}^{\mathcal{D}2}(x + \upsilon)), \quad I_{j1} \neq I_{j2}. \tag{A.2}$$

Finally, the third context-rich feature assumed that lesions usually appear as partial volumes of GM and WM areas with bordering CSF regions, except for pure WM lesions, therefore the feature measured the difference of intensity along a 3D line for each channel, and was given as

$$f(x, I)_{j,\upsilon} = \max_{\lambda}(I_j(x + \lambda\upsilon)) - \min_{\lambda}(I_j(x + \lambda\upsilon)), \quad \lambda \in [0, 1]. \tag{A.3}$$

Appendix B. Quantitative evaluation metrics

The Dice similarity index (DSC) that measures the volumetric overlap between the ground truth and the segmented lesion with the automated methods is given by

$$\text{DSC} = \frac{2 \times TP}{FP + 2 \times TP + FN} \tag{B.1}$$

where, TP =true positive, TN =true negative, FP =false positive and FN =false negative. The positive predictive value or the precision rate (PPV) denotes the proportion of true positives i.e. a high PPV would indicate that a patient identified with a lesion does actually have the lesion and

is given by

$$\text{PPV} = \frac{TP}{TP + FP} \tag{B.2}$$

The true positive rate (TPR), also called as the recall rate or sensitivity, measures the proportion of actual positives which are correctly identified as such and is measured by

$$\text{TPR} = \frac{TP}{TP + FN} \tag{B.3}$$

A high PPV and a high TPR would mean that the detected lesions are actually lesions with less FP and most of the actual lesion is detected. While a high PPV and moderate TPR would mean that the detected lesions are actually lesions with less FP, although not all lesion areas are properly detected. The false positive rate (FPR), also called as the precision or 1-sensitivity, measures the proportion of false positives which are incorrectly identified as lesions and is measured by

$$\text{FPR} = \frac{FP}{FP + TN} \tag{B.4}$$

A high TPR and low FPR is an ideal situation when lesions are detected correctly.

The symmetric mean absolute surface distance (SMAD) attempts to estimate the error between the surfaces of the ground truth lesion and the segmented lesion and is given by the mean of the surface distances from two surfaces in both directions. SMAD can be interpreted as

$$\text{SMAD} = \frac{1}{n_s + n_g}\left(\sum_{i=1}^{n_s} |d_i^{sg}| + \sum_{j=1}^{n_g} \left|d_j^{gs}\right|\right) \tag{B.5}$$

where d_i^{sg} and d_j^{gs} are the Euclidean distances between surface voxels of S (segmented) and G (ground truth) and vice-versa and n_s and n_g are the number of surface voxels for each surface. A low SMAD would mean that the surface of the approximated lesion is in good agreement with the ground truth surface. Each of these metrics quantifies the ability of each method to represent the ground truth, and forms the basis for comparison between methods.

The relative volume difference is measured as

$$\mathrm{VD} = (v_s - v_g)/v_g \tag{B.6}$$

where, v_s and v_g are the segmented and ground truth lesion volumes. A low VD signifies that the segmented lesion volume is more in agreement with the ground truth lesion volume. Each of these metrics quantifies the ability of each method to represent the ground truth, and forms the basis for comparison between methods.

References

Admiraal-Behloul, F., van den Heuvel, D.M.J., van Osch, H.O.M.J.P., van der Grond, J., van Buchem, M.A., Reiber, J.H.C., 2005. Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage 28, 607–617.

Alexander, L.D., Black, S.E., Gao, F., Szilagyi, G., Danells, C.J., McIlroy, W.E., 2010. Correlating lesion size and location to deficits after ischemic stroke: the influence of accounting for altered peri-necrotic tissue and incidental silent infarcts. Behavorial and Brain Functions 6, 6.

Anbeek, P., Vincken, K.L., van Osch, M.J., Bisschops, R.H., van der Grond, J., 2004a. Automatic segmentation of different-sized white matter lesions by voxel probability estimation. Medical Image Analysis 8, 205–215.

Anbeek, P., Vincken, K.L., van Osch, M.J., Bisschops, R.H., van der Grond, J., 2004b. Probabilistic segmentation of white matter lesions in MR imaging. NeuroImage 21, 1037–1044.

Baird, A.E., Warach, S., 1998. Magnetic resonance imaging of acute stroke. Journal Cerebral Blood Flow and Metabolism 18, 583–609.

Bartlett, M.S., 1937. Properties of sufficiency and statistical tests, in: Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences, pp. 268–282.

Bayes, T., Price, R., 1763. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a

letter to John Canton, M.A. and F.R.S. Philosophical Transactions of the Royal Society of London 53, 370–418.

Bernick, C., Kuller, L., Dulberg, C., Jr, W.T.L., Manolio, T., Beauchamp, N., Price, T., 2001. Silent MRI infarcts and the risk of future stroke: the cardiovascular health study. Neurology 57, 1222–1229.

Besag, J., 1986. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society, Series B 48, 259–302.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth and Brooks, Monterey, CA.

Campbell, B.C., Tu, H.T., Christensen, S., Desmond, P.M., Levi, C.R., Bladin, C.F., Hjort, N., Ashkanian, M., Sølling, C., Donnan, G.A., Davis, S.M., Ostergaard, L., Parsons, M.W., 2012. Assessing response to stroke thrombolysis: validation of 24-hour multimodal magnetic resonance imaging. Arch Neurology 69, 46–50.

Carey, L., Crewther, S., Salvado, O., Linden, T., Tse, T., Connelly, A., Howells, D., Ma, H., Churilov, L., Davis, S., Donnan, G., 2013a. START (STroke imAging pRevention and Treatment): A longitudinal stroke cohort study: Clinical Trials Protocol. International Journal of Stroke In Press.

Carey, L.M., Seitz, R.J., Parsons, M., Levi, C., Farquharson, S., Tournier, J.D., Palmer, S., Connelly, A., 2013b. Beyond the lesion - Neuroimaging foundations for poststroke recovery. Future Neurology 8, 507–524.

Cavalieri, M., Enzinger, C., Petrovic, K., Pluta-Fuerst, A., Homayoon, N., Schmidt, H., Fazekas, F., Schmidt, R., 2010. Vascular dementia and Alzheimer's disease − are we in a dead-end road? Neurodegenerative Diseases 7, 122–126.

Clark, R.K., Lee, E.V., Fish, C.J., White, R.F., Price, W.J., Jonak, Z.L., Feuerstein, G.Z., Barone, F.C., 1993. Development of tissue damage, inflammation and resolution following stroke: an immunohistochemical and quantitative planimetric study. Brain Research Bulletin 31, 565–572.

Cramer, S.C., Shah, R., Juranek, J., Crafton, K.R., Le, V., 2006. Activity in the peri-infarct rim in relation to recovery from stroke. Stroke 37, 111–115.

Criminisi, A., Shotton, J., Konukoglu, E., 2011. Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. TechReport MSR-TR-2011-114. Microsoft Research.

Dastidar, P., Heinonen, T., Ahonen, J.P., Jehkonen, M., Molnár, G., 2000. Volumetric measurements of right cerebral hemisphere infarction: use of a semiautomatic MRI segmentation technique. Computers in Biology and Medicine 30, 41–54.

Debette, S., Markus, H.S., 2010. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. British Medical Journal 341, 1–9.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, 1–38.

Fischl, B., Salat, D., Busa, E., Albert, M., Dietrich, M., Haselgrove, C., van der Kouwe, A., Killany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N., 2011. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. NeuroImage 57, 378–390.

Gosset, W.S., 1908. The probable error of a mean. Biometrika 6, 1–25.

Heinonen, T., Dastidar, P., Kauppinen, P., Malmivuo, J., Eskola, H., 1998. Semiautomatic tool for segmentation and volumetric analysis of medical images. Medical and Biological Engineering and Computing 36, 291–296.

Hevia-Montiel, N., Jiménez-Alaniz, J.R., Medina-Bañuelos, V., Yáñez-Suárez, O., Rosso, C., Samson, Y., Baillet, S., 2007. Robust nonparametric segmentation of infarct lesion from diffusion-weighted MR images, in: Proc. of IEEE EMBS, pp. 2102–2105.

Jack, C.R., O'Brien, P.C., Rettman, D.W., Shiung, M.M., Yuecheng, X., Muthupillai, R., Manduca, A., Avula, R., Erickson, B.J., 2001. FLAIR histogram segmentation for measurement of leukoaraiosis volume. Journal of Magnetic Resonance Imaging 14, 668–676.

Jacobs, M.A., Mitsias, P., Soltanian-Zadeh, H., Santhakumar, S., Ghanei, A., Hammond, R., Peck, D.J., Chopp, M., Patel, S., 2001a. Multiparametric

MRI tissue characterization in clinical stroke with correlation to clinical outcome : Part 2. Stroke 30, 950–957.

Jacobs, M.A., Zhang, Z.G., Knight, R.A., Soltanian-Zadeh, H., Goussev, A.V., Peck, D.J., Chopp, M., 2001b. A model for multiparametric MRI tissue characterization in experimental cerebral ischemia with histological validation in rat : Part 1. Stroke 32, 943–949.

Kabir, Y., Dojat, M., Scherrer, B., Forbes, F., Garbay, C., 2007. Multimodal MRI segmentation of ischemic stroke lesions, in: Proc. of IEEE EMBS, pp. 1595–1598.

Knight, R.A., Dereski, M.O., Helpern, J.A., Ordidge, R.J., Chopp, M., 1994. Magnetic resonance imaging assessment of evolving focal cerebral ischemia: comparison with histopathology in rats. Stroke 25, 1252–1262.

Kovalev, V.A., Kruggel, F., Gertz, H.J., von Cramon, D.Y., 2001. 3D texture analysis of MRI brain datasets. IEEE Transactions on Medical Imaging 20, 424–433.

Kruggel, F., Paul, J.S., Gertz, H.J., 2008. Texture-based segmentation of diffuse lesions of the brains white matter. NeuroImage 39, 987–996.

Lao, Z., Shen, D., Liu, D., Jawad, A.F., Melhem, E.R., Launer, L.J., Bryan, R.N., Davatzikos, C., 2008. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. Academic Radiology 15, 300–313.

Launer, L.J., 2004. Epidemiology of white matter lesions. Topics in Magnetic Resonance Imaging 15, 365–367.

van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Transactions in Medical Imaging 20, 677–688.

van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based tissue classification of MR images of the brain. IEEE Transactions in Medical Imaging 18, 897–908.

Lilliefors, H.W., 1967. On the KolmogorovSmirnov test for normality with mean and variance unknown. Journal of American Statistical Association 62, 399–402.

Lindenberg, R., Seitz, R.J., 2012. Impact of white matter damage after stroke, in: Bright, P. (Ed.), Neuroimaging-Methods. InTech. chapter 12. Available from: https://www.intechopen.com/books/neuroimaging-methods/impact-of-white-matter-damage-after-stroke.

Love, S., 2006. Demyelinating diseases. Journal of Clinical Pathology 59, 1151–1159.

Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine 98, 278–284.

Mohamed, F.B., Vinitski, S., Gonzalez, C.F., Faro, S.H., Lublin, F.A., Knobler, R., Gutierrez, J.E., 2001. Increased differentiation of intracranial white matter lesions by multispectral 3D-tissue segmentation: preliminary results. Magnetic Resonance Imaging 19, 207–218.

Norrving, B., 2008. Leucoaraiosis and silent subcortical infarcts. Reviews in Neurology (Paris) 164, 801–804.

Ong, K.H., Ramachandram, D., Mandava, R., Shuaib, I.L., 2012. Automatic white matter lesion segmentation using an adaptive outlier detection method. Magnetic Resonance Imaging 30, 807–823.

O'Sullivan, M., 2008. Leukoaraiosis. Practical Neurology 8, 26–38.

O'Sullivan, M., Morris, R.G., Huckstep, B., Jones, D.K., Williams, S.C., Markus, H.S., 2004. Diffusion tensor MRI correlates with executive dysfunction in patients with ischaemic leukoaraiosis. Journal of Neurology, Neurosurgery and Psychiatry 75, 441–447.

Park, M.H., Min, J.Y., Kwon, D.Y., Lee, S.H., Na, H.R., Cho, S.T., Na, D.L., 2010. Vascular risk factors and the effect of white matter lesions on extrapyramidal signs in Alzheimer's disease. International Psychogeriatry 29, 1–8.

Peters, N., Holtmannspotter, M., Opherk, C., Gschwendtner, A., Herzog, J., Samann, P., Dichgans, M., 2006. Brain volume changes in CADASIL: a serial MRI study in pure subcortical ischemic vascular disease. Neurology 66, 1517–1522.

Rivers, C.S., Wardlaw, J.M., Armitage, P.A., Bastin, M.E., Hand, P.J., Dennis, M.S., 2007. Acute ischemic stroke lesion measurement on diffusion-weighted imaging-important considerations in designing acute stroke trials with magnetic resonance imaging. Journal of Stroke and Cerebrovascular Diseases 16, 64–70.

Seghier, M.L., Ramlackhansingh, A., Crinion, J., Leff, A.P., Price, C.J., 2008. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. NeuroImage 41, 1253–1266.

Shen, S., Szameitat, A.J., Sterr, A., 2010. An improved lesion detection approach based on similarity measurement between fuzzy intensity segmentation and spatial probability maps. Magnetic Resonance Imaging 28, 245–254.

Silbert, L.C., Nelson, C., Howieson, D.B., Moore, M.M., Kaye, J.A., 2008. Impact of white matter hyperintensity volume progression on rate of cognitive and motor decline. Neurology 71, 108–113.

Sokal, R.R., Rohlf, F.J., 1995. Biometry: The principles and practice of statistics in biological research. 3rd ed., W. H. Freeman, New York.

Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., Warfield, S., 2008. 3D segmentation in the clinic: a grand challenge II: MS lesion segmentation, pp. 1–5.

Teodorczuk, A., Firbank, M.J., Pantoni, L., Poggesi, A., Erkinjuntti, T., Wallin, A., Wahlund, L.O., Scheltens, P., Waldemar, G., Schrotter, G., Ferro, J.M., Chabriat, H., Bazner, H., Visser, M., Inzitari, D., O'Brien, J.T., 2010. Relationship between baseline white-matter changes and development of late-life depressive symptoms: 3-year results from the LADIS study. Psychological Medicine 40, 603–610.

Vannier, M.W., Butterfield, R.L., Rickman, D.L., Jordan, D.M., Murphy,

W.A., Biondetti, P.R., 1985. Multispectral magnetic resonance image analysis. Radiology 154, 221–224.

Vermeer, S.E., Hollander, M., van Dijk, E.J., Hofman, A., Koudstaal, P.J., Breteler, M.M., 2003. Silent brain infarcts and white matter lesions increase stroke risk in the general population : The rotterdam scan study. Stroke 34, 1126–1129.

Vernooij, M.W., Ikram, M.A., Tanghe, H.L., Vincent, A.J., Hofman, A., Krestin, G.P., Niessen, W.J., Breteler, M.M., van der Lugt, A., 2007. Incidental findings on brain MRI in the general population. New England Journal of Medicine 357, 1821–1828.

Wang, Y., Catindig, J.A., Hilal, S., Soon, H.W., Ting, E., Wong, T.Y., Venketasubramanian, N., Chen, C., Qiu, A., 2012. Multi-stage segmentation of white matter hyperintensity, cortical and lacunar infarcts. NeuroImage 60, 2379–2388.

Wei, X., Warfield, S.K., Zou, K.H., Wu, Y., Li, X., Guimond, A., Mugler-III, J.P., Benson, R.R., Wolfson, L., Weiner, H.L., Guttmann, C.R., 2002. Quantitative analysis of MRI signal abnormalities of brain white matter with high reproducibility and accuracy. Journal of Magnetic Resonance Imaging 15, 203–209.

Welch, B.L., 1947. The generalization of student's problem when several different population variances are involved. Biometrika 34(1-2), 28–35.

Welch, K.M.A., Windham, J.P., Knight, R.A., Nagesh, V., Hugg, J.W., Jacobs, M.A., Peck, D., Booker, P., Dereski, M.O., Levine, S.R., 1995. A

model to predict the histopathology of human stroke using diffusion and T2-weighted magnetic resonance imaging. Stroke 26, 1983–1989.

Wilke, M., de Haan, B., Juenger, H., Karnath, H.O., 2011. Manual, semi-automated, and automated delineation of chronic brain lesions: A comparison of methods. NeuroImage 56, 2038–2046.

Xavier, A.R., Qureshi, A.I., Kirmani, J.F., Yahia, A.M., Bakshi, R., 2003. Neuroimaging of stroke:a review. Southern Medical Journal 96, 367–379.

Yamauchi, H., Fukuda, H., Oyanagi, C., 2002. Significance of white matter high intensity lesions as a predictor of stroke from arteriolosclerosis. Journal of Neurology, Neurosurgery, and Psychiatry 72, 576–582.

Yang, F., Shan, Z.Y., Kruggel, F., 2010. White matter lesion segmentation based on feature joint occurrence probability and χ^2 random field theory from magnetic resonance (MR) images. Pattern Recognition Letters 31, 781–790.

Yi, Z., Criminisi, A., Shotton, J., Blake, A., 2009. Discriminative, semantic segmentation of brain tissue in MR images, in: Yang, G.Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (Eds.), Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer Berlin Heidelberg. volume 5762 of *Lecture Notes in Computer Science*, pp. 558–565.

Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O.M., Das, T., Jena, R., Price, S.J., 2012. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR,

in: Medical Image Computing and Computer-Assisted Intervention (MICCAI). volume 7512 of *Lecture Notes in Computer Science*, pp. 369–376.

Table B.1: Demographic data for stroke patients.

Age	Mean ± Std. Dev.: 64.57±14.23 yrs; Range: 46-87 yrs & 1 patient: 27 yrs
Gender	24 males & 12 females
Primary stroke	34 cases
WM lesions	30 cases
Secondary lesions	10 cases
Lesion volume	Mean ± Std. Dev.: 19.85±15.88 mL; Range: 0.7 mL-69.33 mL

Table B.2: Quantitative results for lesion detection using the proposed method RF. DSC (Dice similarity coefficient), PPV (positive predictive value), and TPR (true positive rate) are in fractions within the range 0 to 1, the SMAD (symmetric mean absolute surface distance) is in mm, and VD in %. μ is the average and σ is the standard deviation.

Patient#	**DSC**	**PPV**	**TPR**	**SMAD**	**VD(%)**
1	0.65	0.86	0.52	1.35	39.29
2	0.59	0.90	0.44	2.99	51.04
3	0.19	0.16	0.25	10.45	54.28
4	0.75	0.82	0.69	0.77	16.12
5	0.69	0.74	0.64	1.31	12.74
6	0.68	0.85	0.57	0.89	32.43
7	0.82	0.97	0.70	1.04	27.42
8	0.70	0.90	0.57	1.12	36.32
9	0.68	0.98	0.52	1.63	47.34
10	0.77	0.88	0.68	0.88	22.43
11	0.51	0.40	0.68	4.13	69.83
12	0.68	0.70	0.67	1.13	4.01
13	0.53	0.64	0.45	3.13	30.05
14	0.64	0.66	0.63	1.47	5.08
15	0.57	0.62	0.53	2.40	13.51
16	0.37	0.91	0.24	4.27	74.15
17	0.44	0.83	0.30	3.21	63.66
18	0.51	0.97	0.34	2.64	64.68
19	0.55	0.70	0.45	1.78	35.62
20	0.70	0.80	0.62	1.03	22.14
21	0.51	0.66	0.42	10.95	36.54
22	0.61	0.61	0.61	1.51	0.50
23	0.60	0.93	0.44	1.96	52.46
24	0.50	0.43	0.59	9.09	37.46
25	0.54	0.55	0.53	6.77	3.67
26	0.57	0.60	0.54	3.09	9.63
27	0.53	0.79	0.40	1.85	49.35
28	0.49	0.95	0.33	3.05	65.12
29	0.52	0.83	0.38	2.39	54.44
30	0.60	0.60	0.59	2.29	1.04
31	0.52	0.63	0.45	13.77	29.30
32	0.73	0.79	0.69	1.81	13.00
33	0.62	0.84	0.49	1.06	42.01
34	0.72	0.73	0.72	0.76	2.03
35	0.73	0.80	0.68	0.70	15.93
36	0.74	0.89	0.63	1.52	29.06
μ	**0.60**	**0.75**	**0.53**	**3.06**	**32.32**
σ	**0.12**	**0.18**	**0.13**	**3.17**	**21.64**

Table B.3: Statistical analysis of the quantitative results for lesion detection of the proposed method RF and the compared methods MAP-MRF and LIK respectively. The mean±standard deviation values of the evaluation metrics for the compared methods are shown. The DSC, PPV and TPR values range from 0 to 1, while the SMAD value is in mm and VD in %. The p-value shows the statistical significance of the equality of the means (with 95% confidence interval) of the methods when compared to the proposed method and t-type denotes the statistical test type.

Metric	**MAP-MRF**	**LIK**
DSC	0.39±0.22	0.42±0.18
p-value	<0.0001	<0.0001
t-type	Welch	Welch
PPV	0.39±0.29	0.62±0.3
p-value	<0.0001	0.05
t-type	Welch	Welch
TPR	0.55±0.13	0.38±0.12
p-value	<0.0001	<0.0001
t-type	Student	Student
SMAD	7.71±6.18	6.09±5.92
p-value	0.0002	0.0091
t-type	Welch	Welch
VD	446.7±971.4	155.7±389.9
p-value	0.0150	0.0663
t-type	Welch	Welch

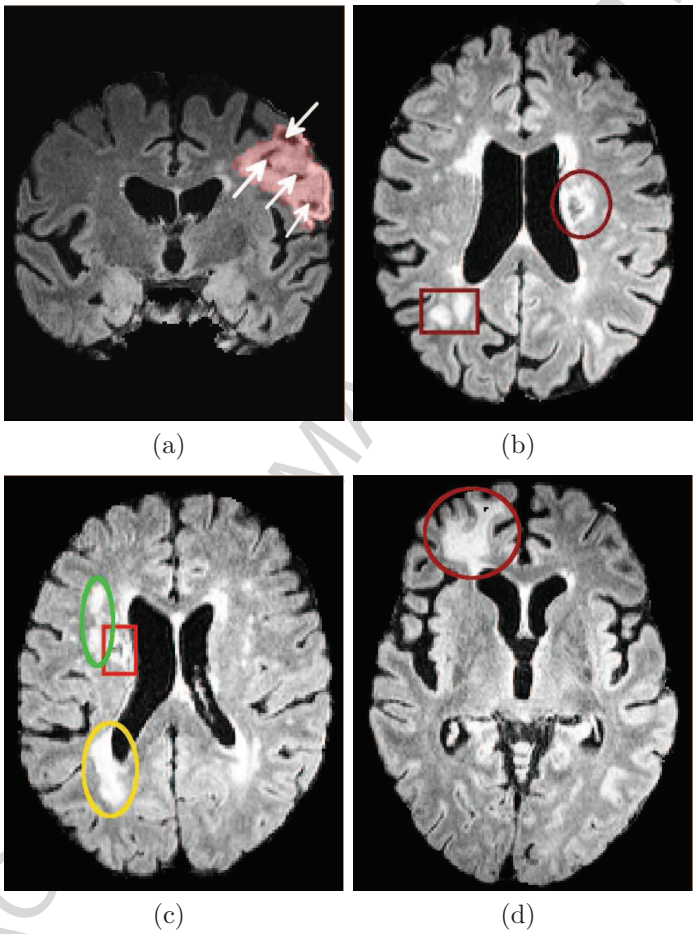


Figure B.1: Stroke lesions at the chronic stage with associated white matter hyperintensities on FLAIR showing similar tissue contrasts. (a) Heterogenous intensities within the chronic stage stroke lesion due to gliosis. The light red mask denotes the entire stroke area as marked by an expert, while the arrows show possible necrosis or atrophy with CSF involvement; (b) anomalous FLAIR hyperintensities. The circular region in the image depicts a stroke region while the square area contains other white matter hyperintensities; (c) the square area delineates a stroke, while the oval areas are WM lesions. The stroke in the red square has a tendency to fuse with the WM lesion in the green oval region. The yellow oval region indicates large WM lesion load around the periventricular region; (d) large secondary/pre-existing stroke lesion in the frontal lobe which is not a primary stroke but has similar intensities.

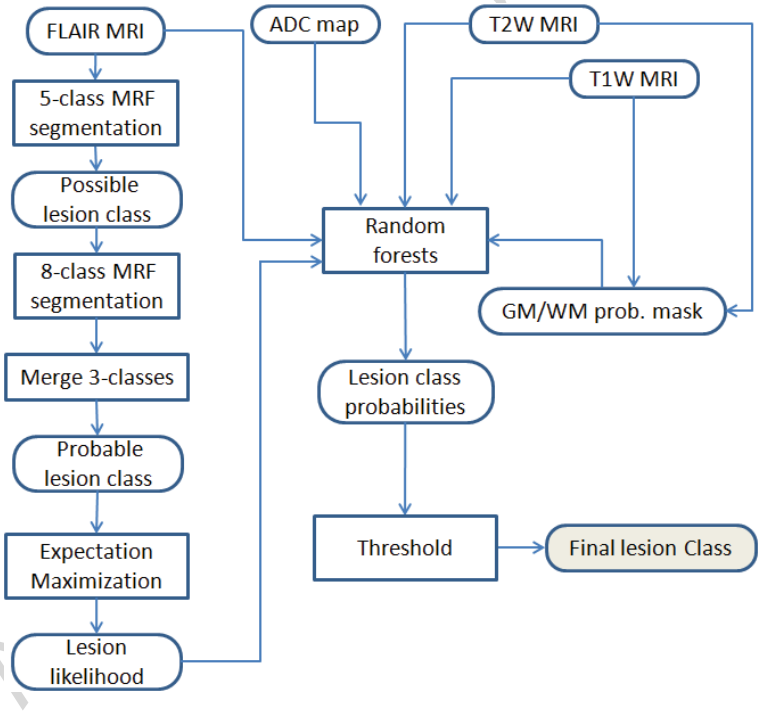
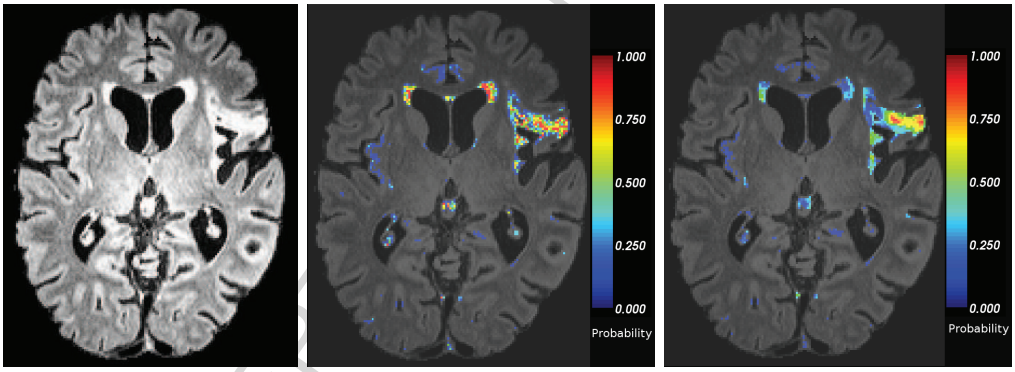


Figure B.2: A schematic diagram of the proposed method for lesion segmentation.



(a) (b) (c)

Figure B.3: Likelihood and random forest probabilities of probable lesion areas. (a) Original FLAIR image; (b) The likelihood obtained from the 8-class EM of the probable lesion area shows heterogeneity in the probabilities; (c) probabilities obtained from random forest are smoother.

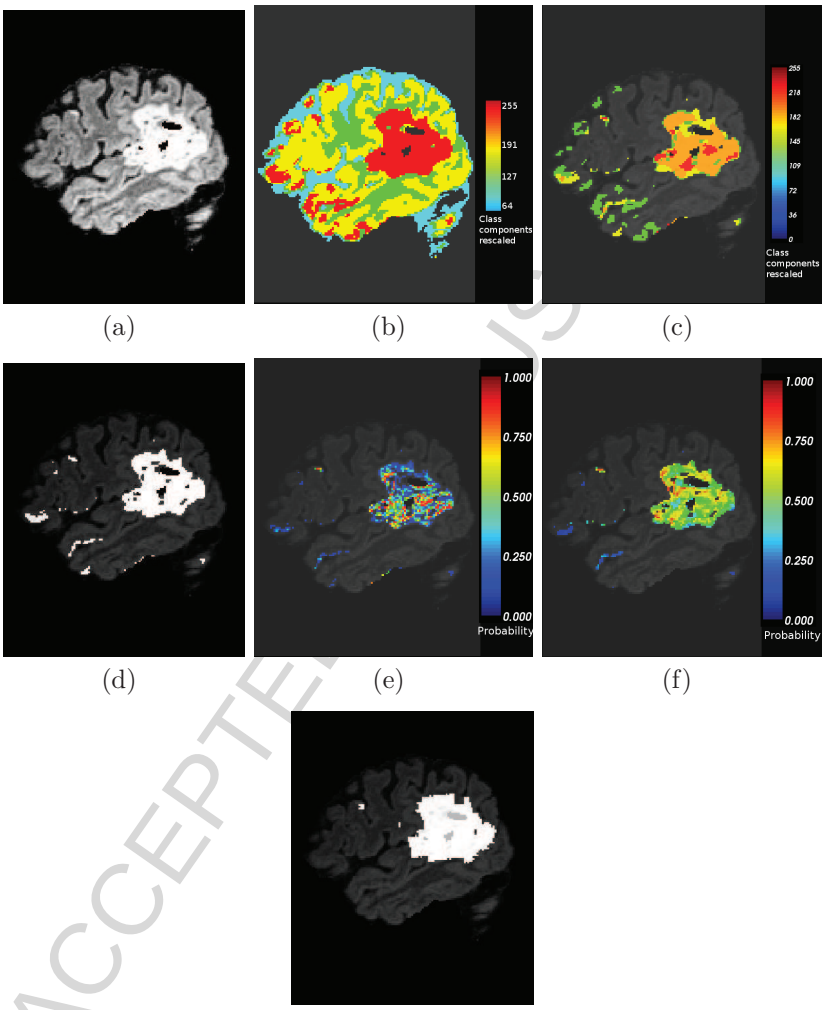


Figure B.4: Stages of processing applied on a patient case showing sagittal view of the brain. (a) Original FLAIR image showing hyperintense lesion; (b) a 5-component Bayesian and MRF segmentation in the FLAIR, showing the class in 'red' as the possible lesion class but also contains many GM areas; (c) a 8-component Bayesian and MRF segmentation on the possible lesion class of (b). This image shows that the lesion area is well-approximated by the 3 classes with high means, i.e., 'red', 'orange' and 'yellow' classes. Although some 'green' areas are also encompassed in the lesion but we avoid those areas to reduce false positives; (d) this shows the binary class of the combined 3 classes from (c) that is used to train the random forest and is the probable lesion area; (e) this is the likelihood of the probable lesion that maximizes the probabilities in the lesion area when split into lesion and non-lesion classes by EM; (f) this image shows the random forest probabilities overlaid; and (g) is the ground truth segmentation of the lesion made by the expert.

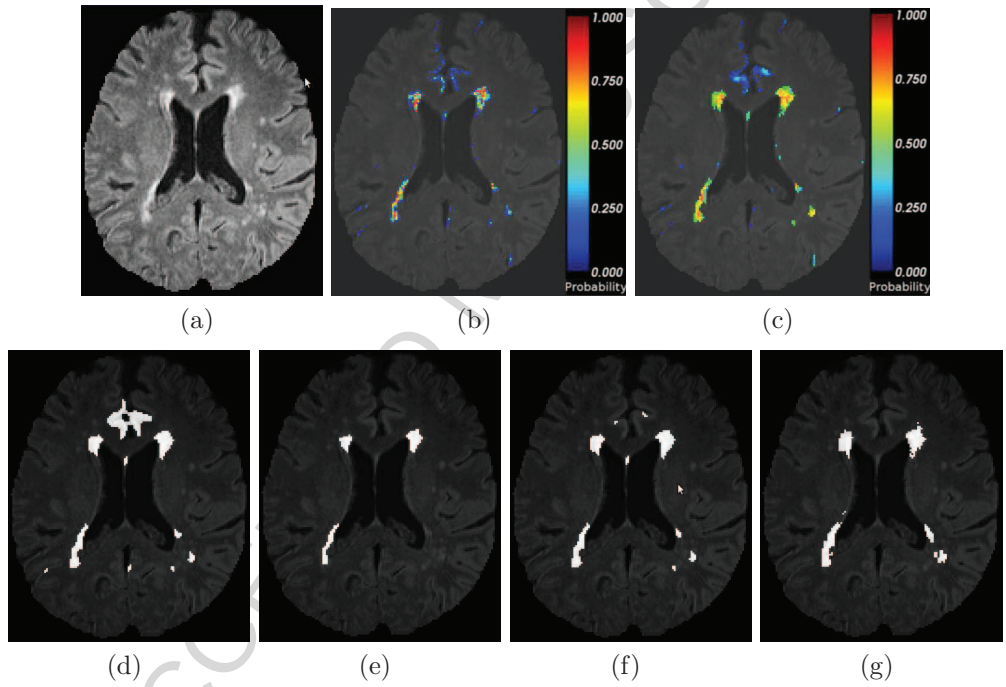


Figure B.5: Results of the proposed and compared methods overlaid on the FLAIR image for patient 20. (a) Original FLAIR image showing hyperintense lesion; (b) probable lesion likelihood obtained from EM estimation overlaid; (c) probabilities from random forest overlaid; (d) segmented lesion from the 3-class components of the second level MRF, i.e., the method MAP-MRF; (e) segmented lesion from the method LIK; (f) segmented lesion from method RF; and (g) is the ground truth segmentation of the lesion made by the expert.

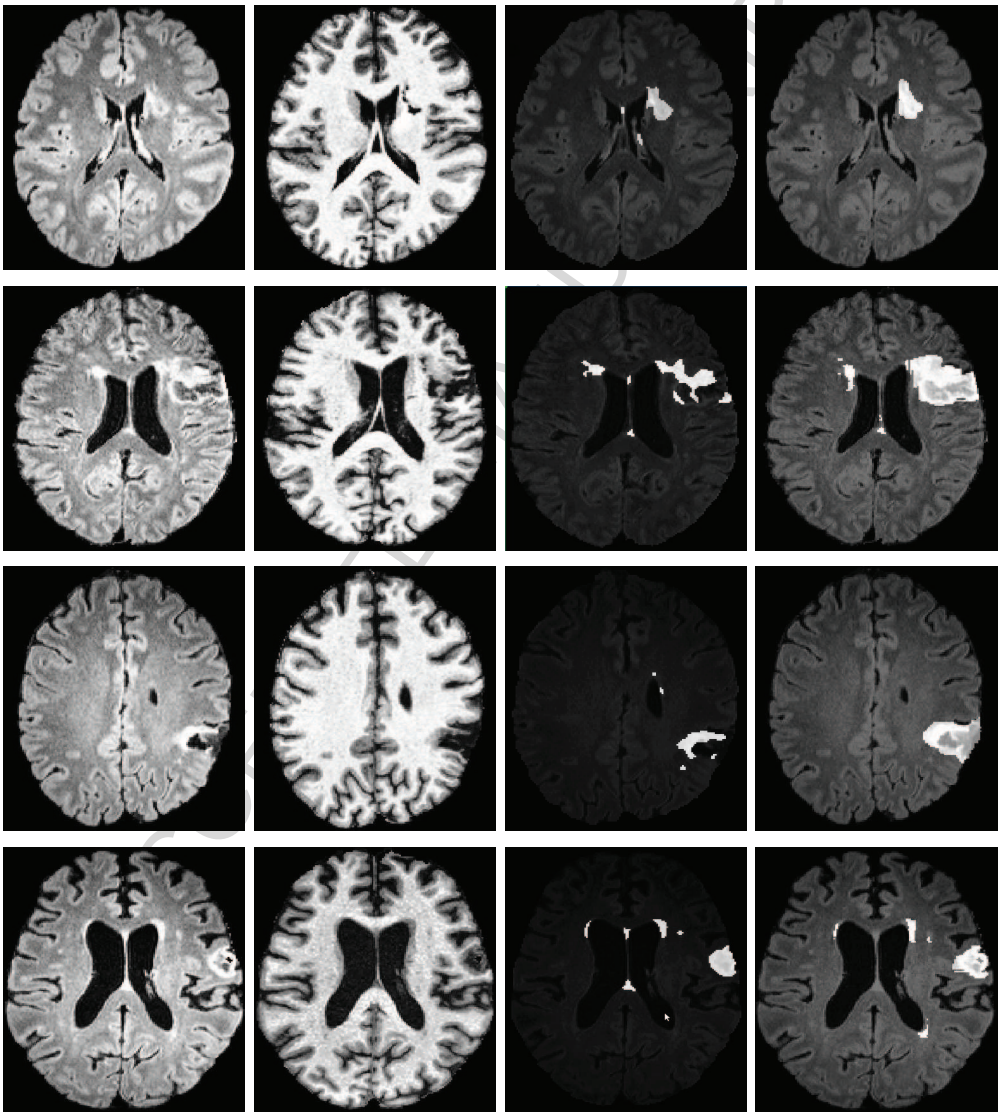
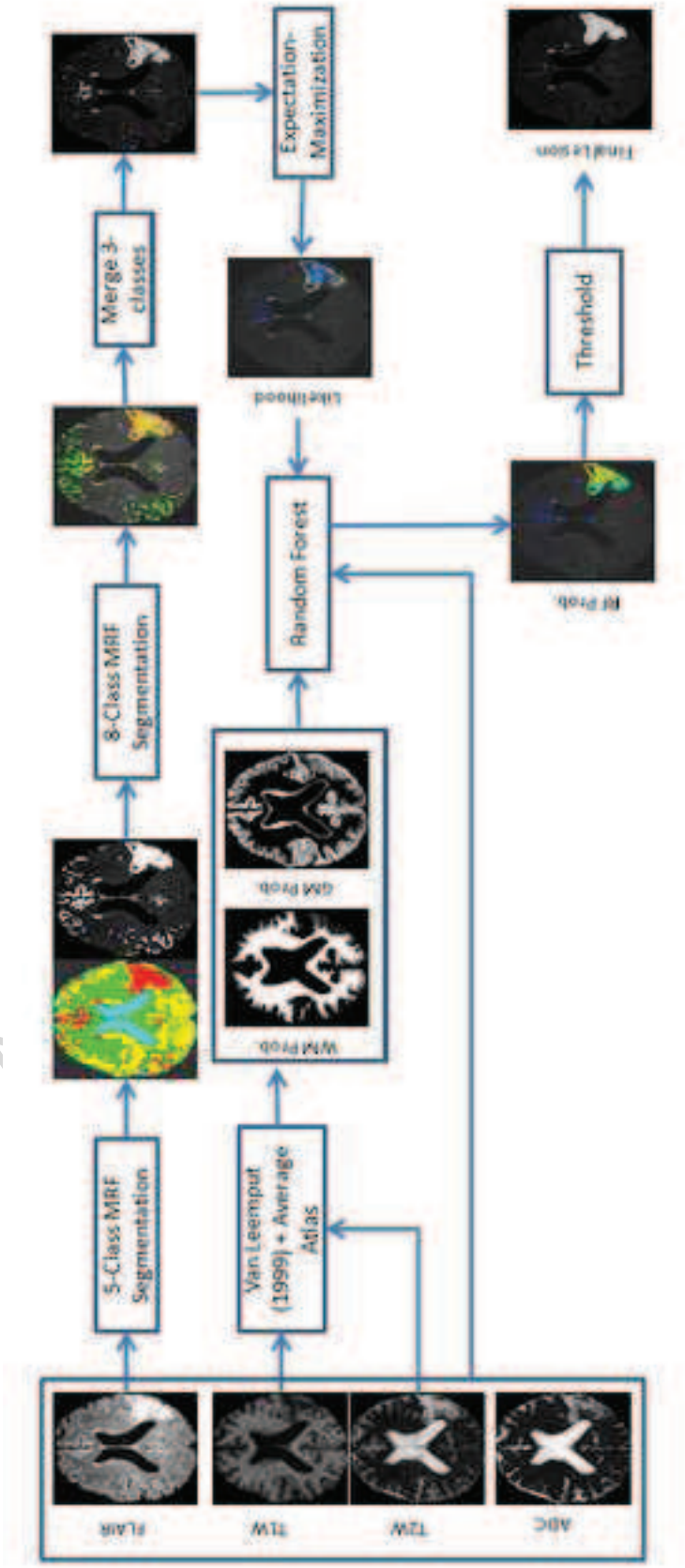


Figure B.6: Quantitative lesion segmentation evaluation. The rows belong to patients 25, 23, 16 and 2 from top to bottom respectively. Columns 1 and 2 are the FLAIR and T1W images, column 3 shows the results of the lesion segmentation overlaid on the FLAIR and column 4 shows the ground truth segmentations overlaid on the FLAIR images.

Graphical Abstract



Highlights

- Automated segmentation of chronic ischemic infarct and secondary lesions impacting post-stroke depression.
- Hierarchical segmentation of the probable lesion class from FLAIR MRI in Bayesian-MRF framework.
- Random-forest training and probabilistic classification on probable lesion class.
- Context-rich features for random-forest include multimodal MRI; GM, WM, and lesion likelihoods from Expectation-Maximization.