# Figure Captions

**Figure 1:** Representation of Stage 1 feature images for subject 01C1019. The FLAIR, T1-, and T2-weighted images are rigidly pre-aligned [32] to the space of the T1 image. The three modality images are then preprocessed (N4 bias correction [34] and adaptive denoising [33]) followed by application of standard ANTs brain extraction and -tissue segmentation protocols using the MMRR symmetric template and corresponding priors36 applied to the T1 image. The feature images are then generated for voxelwise input to the RF model which results in the voting maps illustrated on the right. This gives a probabilistic classification of tissue type. Not shown are the probability and voting images for the brain stem and cerebellum.

**Figure 2:** Sample FLAIR acquisition image slices showing both manual and random forest segmentations for both stages obtained during the leave-one-out evaluation. Manual segmentations were performed by one of the authors and provided the ground truth WMH labels for training the random forest models.

**Figure 3:** Canonical views of the mutlivariate, bilaterally symmetric template constructed from the MMRR data set [39] (only shown are the FLAIR, T1, and T2 modalities--- the components relevant for this work). Template construction is detailed in [25]. These images are important for specific intensity-based features.

**Figure 4:** Average *MeanDecreaseAccuracy* plots generated from the creation of all 24 random forest models for Stage 1 during the leave-one-out evaluation. These plots are useful in providing a quantitative assessment of the predictive importance of each feature. Features are ranked in descending order of importance. The horizontal error bars provide the percentile and illustrate the stability of the feature importance across the leave-one-out models. At this initial stage only 31 feature images are used.

**Figure 5:** Average *MeanDecreaseAccuracy* plots generated from the creation of all 24 random forest models for Stage 2 during the leave-one-out evaluation. These plots are useful in providing a quantitative assessment of the predictive importance of each feature. Features are ranked in descending order of importance. The horizontal error bars provide the percentile and illustrate the stability of the feature importance across the leave-one-out models. We augment the 31 feature images from the first stage by adding an additional seven voting maps and 7 segmentation posteriors from application of the Bayesian-based segmentation for a total of 45 images for the second stage.

**Figure 6:** (a) FLAIR image slice illustrating WMHs which have been manually delineated. The region around the WMHs is enlarged (b) in the original FLAIR and the (c) contralateral FLAIR difference image.

**Figure 7:** Voxelwise comparison with manual delineation of white matter hyperintensities. On the left are the calculated Dice values over all white matter hyperintensities. Note the improvement in the Dice metric from the employment of the Stage 2 component of the processing pipeline. (Right) Similar results can be seen by comparing the total lesion load volume between manual and automated detection strategies. Although some outliers are found after the Stage 2 processing in a couple subjects, the number of outliers caused by false positives is decreased significantly with the second stage processing.