Materials and Methods

Imaging

Quantitative analysis

Random forests for machine learning

Machine learning and pattern recognition techniques have seen increased application for various medical image analysis workflows (see, for example, the annual Workshop on Machine Learning in Medical Imaging held in conjunction with the Medical Image Computing and Computer-Aided Intervention international meeting [1]). Popular techniques such as support vector machines and neural networks have been applied successfully to clinically relevant imaging tasks such as supervised image segmentation (e.g., [2]) and diagnostic prediction (e.g., [3, 4]). Facilitating the current employment of such techniques are the number of available imaging data sets [5] and the public availability of data science packages such as SciPy [6] and the R project for statistical computing [7] and their associated add-on toolkits.

Random forests [8] is a popular machine learning technique that has demonstrated significant utitility for supervised segmentation tasks (e.g., normal human brain segmentation) and other computer vision applications (e.g., human gait detection [9]). In the context of neuropathology, random forest-based paradigms have been employed in the delineation of multiple sclerosis lesions [10], stroke lesions [11], and brain tumors [12–15]. Of note, these latter random forest approaches for brain tumor segmentation have performed well in recent international competitions. In response to the lack of objective comparisons between segmentation algorithms, the Multimodal Brain Tumor Segmentation (BRATS) challenge was initiated in 2012 [16] and has continued every year since under the auspices of the International Conference of Medical Image Computing and Computer Assisted Interventions (MICCAI).

Random forests are conceptually simple [8]. They consist of ensembles of decision trees that are built from training data. Once constructed, data to be classified is “pushed” through each decision tree resulting in a single classification “vote” per tree. These votes are then be used for regression or classification of the data. Although decision trees had been extensively studied previously, the success of employing collections of such weak learners for boosting machine learning performance (e.g., AdaBoost [17, 18]) influenced the similarly sytled conglomeration of decision trees into “forests” with randomized node optimization [19, 20]. Finally, Breiman [8] improved accuracy by random sampling of the training data (i.e., “bagging”) resulting in the current random forest framework.

Crucial to these supervised segmentation approaches are the creation and selection of "features" as input in conjunction with the ground-truth for model construction. For the targeted application in this work (i.e., white matter hyperintensities), regression/classification are performed at the voxelwise level. In other words, each voxel within the region of interest is sent through the ensemble of decision trees and receives a set of classification votes from each tree permitting a regression or classification solution. Since this procedure is performed at the voxelwise level, intensity information alone is insufficient for good segmentation performance since it lacks spatial context. For example, as pointed out in [21], higher intensities can be found at the periventricular caps in normal subjects which often confounds automated lesion detection algorithms. Other potential confounds include MR signal inhomogeneity and noise. Therefore, even though machine learning and pattern recognition techniques are extremely powerful and have significant potential, just as crucial to outcome is the creative construction and deployment of salient feature images which we detail below.

Feature images for WMH segmentation

Supervised methodologies are uniquely characterized, in part, by the feature images that are used to identify the regions of interest. In Table 1, we provide a list and basic categorization of the feature images used for the initial (i.e., Stage 1—more on the use of multiple random forest stages below) segmentation of the white matter hyperintensities. In addition Figure 1 provides a representation of a set of feature images for a single subject analyzed in this work.

As mentioned previously, input for each subject comprises FLAIR, T1-, and T2-weighted acquisitions. The FLAIR and T2 images are rigidly registered to the T1 image using the open-source Advanced Normalization Tools (ANTs) [22]. The aligned images are then preprocessed using the denoising algorithm of [24] followed by N4 bias correction [23] which are then normalized to the intensity range $[0, 1]$. Although we could have used the intensity standardization algorithm of [26], we found that a simple linear rescaling produced better results.

The T1 image is then processed via the ANTs brain extraction and tissue segmentation protocols described in [25] in order to produce a mask for the brain parenchyma and provide probabilistic estimates of the cerebrospinal fluid (csf), gray matter, white matter, deep gray matter, brain stem, and cerebellum. Segmentation is performed using the ANTs Atropos tool [27] and multi-model optimal symmetric shape/intensity templates [15] created from the public MMRR data set [28].

To model the intensity information the first set of images simply includes the preprocessed and normalized intensity FLAIR, T1, and T2 images. We also calculate a set of neighborhood statistics (mean, standard deviation, and skewness) feature images using a radius of one voxel. For each of the normalized images, we calculate the

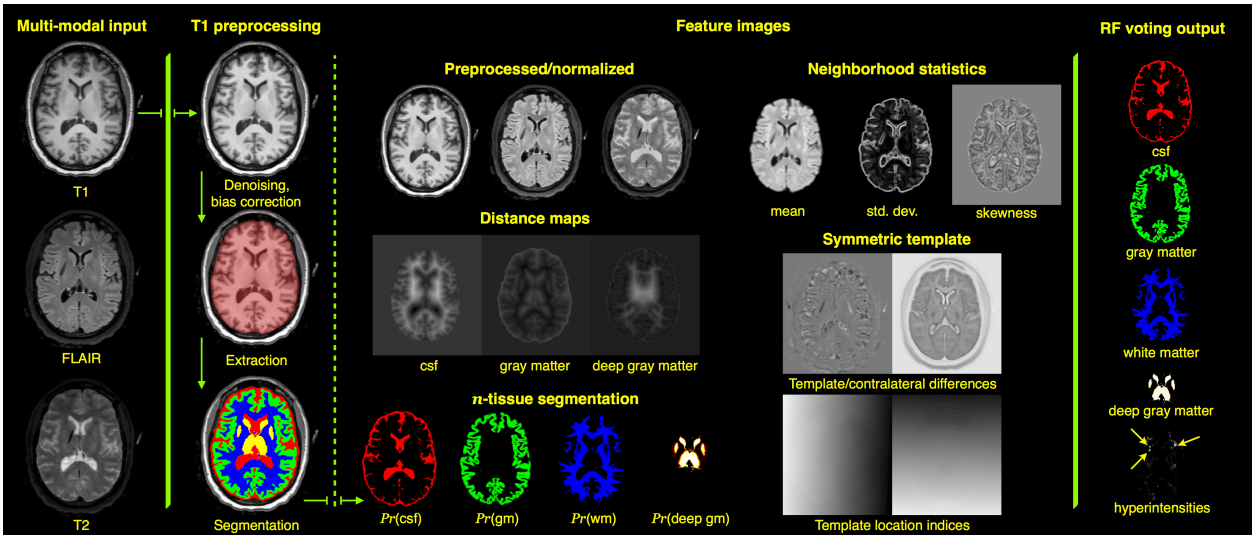

Figure 1: Representation of Stage 1 feature images for subject 01C1019. The FLAIR, T1-, and T2-weighted images are rigidly pre-aligned [22] to the space of the T1 image. The three images are then preprocessed (N4 bias correction [23] and adaptive denoising [24]) followed by application of standard ANTs brain extraction and n-tissue segmentation protocols using the MMRR symmetric template corresponding and priors [25] 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 which gives a probabilistic classification of tissue type. Not shown are the probability and voting images for the brain stem and cerebellum.

Feature type	Image source
Intensities	
normalized/preprocessed	FLAIR, T1, and T2
Symmetric template	
template difference	FLAIR, T1, and T2
contralateral difference	FLAIR, T1, and T2
location indices	FLAIR, T1, and T2
Segmentation probabilities	
Pr(cerebrospinal fluid)	T1
Pr(gray matter)	T1
Pr(white matter)	T1
Pr(deep gray matter)	T1
Pr(brain stem)	T1
Pr(cerebellum)	T1
Distance maps	
cerebrospinal fluid	T1 brain segmentation
gray matter	T1 brain segmentation
deep gray matter	T1 brain segmentation
whole brain	T1 brain segmentation
Neighborhood statistics	
mean	FLAIR, T1, and T2
standard deviation	FLAIR, T1, and T2
skewness	FLAIR, T1, and T2

Table 1: List of feature images used for Stage 1 of the proposed white matter hyperintensity segmentation framework.

difference in intensities with the corresponding warped template component. Previous success in the international brain tumor segmentation competition [16] was based on an important set of intensity features that were created from multi-modal templates mentioned previously [15]. We employ the same strategy here. For example, the template difference feature image for the FLAIR image, S_{FLAIR} is calculated as:

$$S_{FLAIR} - T_{FLAIR}\left(\phi_b^{-1}\right)$$

where $\phi_b : S \underset{b}{\leftrightarrow\!\!\!\sim} T$ is the transform which maps from the individual subject space to the template space and T_{FLAIR} is the FLAIR template component. Also, to take advantage of the bilateral symmetry of the normal brain (in terms of both shape and intensity), and the fact that the presence of WMH violates that assumption, we use the symmetric templates to compute the contralateral intensity differences as an additional intensity feature. For the FLAIR component, this contralateral difference image is calculated from

$$S_{FLAIR} - S_{FLAIR}\left(\phi_b^{-1}\left(\phi_R\left(\phi_b\right)\right)\right)$$

where ϕ_R denotes a horizontal reflection perpendicular to the mid-sagittal plane of the symmetric template.

The segmentation probability images described above are used as feature images to provide a spatial context for the random forest model prediction step. Additional spatial contextual feature images include the distance maps [29] based on the csf, gray matter, and deep gray matter images. These latter images are intended to help distinguish white matter hyperintensities from false positives induced by the partial voluming at the gray/white matter interface. A third set of images are based on the voxel location within the space of the template. The T1 image of the subject is registered to the T1 template component using a B-spline variant [30] of the well-known ANTs Symmetric Normalization (SyN) algorithm [31]. Since the inverse transform is also derived as part of the registration process, we can warp the voxel index locations back to the space of the individual subject. Note that this is similar in motivation to the work of [32]. However, this previous work lacks the normalization to the standard coordinate system provided by the template to dramatically improve spatial specificity across all subjects.

Stacked/cascaded/concatenated random forests for improved segmentation performance

It has been previously observed that

Code and data availability

Evaluation protocol

Results

To calculate this quantity for a single feature from a single random forest model, the decrease in prediction accuracy produced by omitting the specified feature is calculated during the out-of-bag phase of model creation.

During the out-of-bag error calculation stage of the random forest model creation, the decrease in prediction accuracy with the omission of a single feature or variable is tracked and averaged. Those features which have

Figure 2: Average `MeanDecreaseAccuracy` plots generated from the creation of all 24 random forest models for both Stage 1 and Stage 2 during the leave-one-out evaluation. These plots are useful in providing a quantitative assessment of the predictive importance of each feature. The error bars provide the 95^{th} percentile (i.e., $1.96 \times \sigma$) and illustrate the stability of the feature importance across the leave-one-out models.

the greatest decrease in mean accuracy are considered to be the most discriminative. In this work, we do not use these measurements for feature pruning.

References

1. Available at http://mlmi2015.web.unc.edu

2. Bauer, S., Nolte, L.-P., and Reyes, M. "**Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization**" *Med Image Comput Comput Assist Interv* 14, no. Pt 3 (2011): 354–61.

3. Tong, T., Wolz, R., Gao, Q., Guerrero, R., Hajnal, J. V., Rueckert, D., and Alzheimer's Disease Neuroimaging Initiative. "**Multiple Instance Learning for Classification of Dementia in Brain MRI**" *Med Image Anal* 18, no. 5 (2014): 808–18. doi:10.1016/j.media.2014.04.006

4. Liu, X., Tosun, D., Weiner, M. W., Schuff, N., and Alzheimer's Disease Neuroimaging Initiative. "**Locally Linear Embedding (LLE) for MRI Based Alzheimer's Disease Classification**" *Neuroimage* 83, (2013): 148–57. doi:10.1016/j.neuroimage.2013.06.033

5. Van Horn, J. D. and Toga, A. W. "**Human Neuroimaging as a 'Big Data' Science**" *Brain Imaging Behav* 8, no. 2 (2014): 323–31. doi:10.1007/s11682-013-9255-y

6. Jones, E., Oliphant, T., Peterson, P., and others. "**SciPy: Open Source Scientific Tools for Python**" (2001–2001--): Available at http://www.scipy.org/

7. R Core Team. "**R: A Language and Environment for Statistical Computing**" (2016):

8. Breiman, L. "**Random Forests**" *Machine learning* (2001): 5–32.

9. Viola, P., Jones, M., and Snow, D. "**Detecting Pedestrians Using Patterns of Motion and Appearance**" *International Journal of Computer Vision* 63, (2005): 153–161.

10. Geremia, E., Clatz, O., Menze, B. H., Konukoglu, E., Criminisi, A., and Ayache, N. "**Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel Magnetic Resonance Images**" *Neuroimage* 57, no. 2 (2011): 378–90. doi:10.1016/j.neuroimage.2011.03.080

11. Pustina, D., Coslett, H. B., Turkeltaub, P. E., Tustison, N., Schwartz, M. F., and Avants, B. "**Automated Segmentation of Chronic Stroke Lesions Using LINDA: Lesion Identification with Neighborhood Data Analysis**" *Hum Brain Mapp* (2016): doi:10.1002/hbm.23110

12. Geremia, E., Menze, B. H., and Ayache, N. "**Spatial Decision Forests for Glioma Segmentation in Multi-Channel MR Images**" *Proceedings of mICCAI-bRATS 2012* (2012):

13. Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.-P., and Reyes, M. "**Segmentation of Brain Tumor Images Based on Integrated Hierarchical Classification and Regularization**" *Proceedings of*

mICCAI-bRATS 2012 (2012): 10–13.

14. Zikic, D., Glocker, B., Konukoglu, E., Shotton, J., Criminisi, A., Ye, D. H., Demiralp, C., Thomas, O. M., Das, T., Jena, R., and Price, S. J. "**Context-Sensitive Classification Forests for Segmentation of Brain Tumor Tissues**" *Proceedings of mICCAI-bRATS 2012* (2012): 1–9.

15. Tustison, N. J., Shrinidhi, K. L., Wintermark, M., Durst, C. R., Kandel, B. M., Gee, J. C., Grossman, M. C., and Avants, B. B. "**Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with** $ANTsR$" *Neuroinformatics* (2014): doi:10.1007/s12021-014-9245-2

16. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., Corso, J. J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C. R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K. M., Jena, R., John, N. M., Konukoglu, E., Lashkari, D., Mariz, J. A., Meier, R., Pereira, S., Precup, D., Price, S. J., Raviv, T. R., Reza, S. M. S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C. A., Sousa, N., Subbanna, N. K., Szekely, G., Taylor, T. J., Thomas, O. M., Tustison, N. J., Unal, G., Vasseur, F., Wintermark, M., Ye, D. H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., and Van Leemput, K. "**The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)**" *IEEE Trans Med Imaging* 34, no. 10 (2015): 1993–2024. doi:10.1109/TMI.2014.2377694

17. Schapire, R. "**The Strength of Weak Learnability**" *Machine Learning* 5, (1990): 197–227.

18. Freund, Y. and Schapire, R. "**A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting**" *Journal of Computer and System Sciences* 55, (1997): 119–139.

19. Ho, T. K. "**Random Decision Forests**" *Document analysis and recognition, 1995., proceedings of the third international conference on* 1, (1995): 278–282 vol.1. doi:10.1109/ICDAR.1995.598994

20. Amit, Y. and Geman, D. "**Shape Quantization and Recognition with Randomized Trees**" *Neural Computation* 9, (1997): 1545–1588.

21. Neema, M., Guss, Z. D., Stankiewicz, J. M., Arora, A., Healy, B. C., and Bakshi, R. "**Normal Findings on Brain Fluid-Attenuated Inversion Recovery MR Images at 3T**" *AJNR Am J Neuroradiol* 30, no. 5 (2009): 911–6. doi:10.3174/ajnr.A1514

22. Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., and Gee, J. C. "**The Insight ToolKit Image**

Registration Framework" *Front Neuroinform* 8, (2014): 44. doi:10.3389/fninf.2014.00044

23. Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. "**N4ITK: Improved N3 Bias Correction**" *IEEE Trans Med Imaging* 29, no. 6 (2010): 1310–20. doi:10.1109/TMI.2010.2046908

24. Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L., and Robles, M. "**Adaptive Non-Local Means Denoising of MR Images with Spatially Varying Noise Levels**" *J Magn Reson Imaging* 31, no. 1 (2010): 192–203. doi:10.1002/jmri.22003

25. Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., Kandel, B. M., Strien, N. van, Stone, J. R., Gee, J. C., and Avants, B. B. "**Large-Scale Evaluation of ANTs and FreeSurfer Cortical Thickness Measurements**" *Neuroimage* 99, (2014): 166–79. doi:10.1016/j.neuroimage.2014.05.044

26. Nyúl, L. G., Udupa, J. K., and Zhang, X. "**New Variants of a Method of MRI Scale Standardization**" *IEEE Trans Med Imaging* 19, no. 2 (2000): 143–50. doi:10.1109/42.836373

27. Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., and Gee, J. C. "**An Open Source Multivariate Framework for n-Tissue Segmentation with Evaluation on Public Data**" *Neuroinformatics* 9, no. 4 (2011): 381–400. doi:10.1007/s12021-011-9109-y

28. Landman, B. A., Huang, A. J., Gifford, A., Vikram, D. S., Lim, I. A. L., Farrell, J. A. D., Bogovic, J. A., Hua, J., Chen, M., Jarso, S., Smith, S. A., Joel, S., Mori, S., Pekar, J. J., Barker, P. B., Prince, J. L., and Zijl, P. C. M. van. "**Multi-Parametric Neuroimaging Reproducibility: A 3-T Resource Study**" *Neuroimage* 54, no. 4 (2011): 2854–66. doi:10.1016/j.neuroimage.2010.11.047

29. Maurer, C. R., Rensheng, Q., and Raghavan, V. "**A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions**" *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25, no. 2 (2003): 265–270. doi:10.1109/TPAMI.2003.1177156

30. Tustison, N. J. and Avants, B. B. "**Explicit B-Spline Regularization in Diffeomorphic Image Registration**" *Front Neuroinform* 7, (2013): 39. doi:10.3389/fninf.2013.00039

31. Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. "**A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration**" *Neuroimage* 54, no. 3 (2011): 2033–44. doi:10.1016/j.neuroimage.2010.09.025

32. Anbeek, P., Vincken, K. L., Osch, M. J. P. van, Bisschops, R. H. C., and Grond, J. van der. "**Probabilistic Segmentation of White Matter Lesions in MR Imaging**" *Neuroimage* 21, no. 3 (2004): 1037–44. doi:10.1016/j.neuroimage.2003.10.012