Abstract

White matter hyperintensities (WMHs) are foci of abnormal signal intensity in white matter regions seen with magnetic resonance imaging (MRI). WMHs are associated with normal aging and have shown prognostic value in neurological conditions such as traumatic brain injury (TBI). The impracticality of manually quantifying these lesions limits their clinical utility and motivates the utilization of machine learning techniques for automated segmentation workflows. Herein, we develop a concatenated random forest framework with image features for segmenting WMHs in a TBI cohort. The framework is built upon the Advanced Normalization Tools (ANTs) and ANTsR toolkits. MR (3D FLAIR, T2-, and T1-weighted) images from 24 service members and veterans scanned in the Chronic Effects of Neurotrauma Consortium's (CENC) observational study were acquired. Manual annotations were employed for both training and evaluation using a leave-one-out strategy. Performance measures include sensitivity, positive predictive value, F_1 score, and relative volume difference. Final average results were: sensitivity = 0.68 $\pm$ 0.38, positive predictive value = 0.51 $\pm$ 0.40, F_1 = 0.52 $\pm$ 0.36, relative volume difference = 43 $\pm$ 26%. In addition, three lesion size ranges are selected to illustrate the variation in performance with lesion size. Paired with correlative outcome data, supervised learning methods may allow for identification of imaging features predictive of diagnosis and prognosis in individual TBI patients.

Introduction

White matter hyperintensities in TBI

White matter hyperintensities (WMHs) are foci of abnormally increased signal intensity seen within white matter regions within the cerebrum and brainstem on fluid attenuation inversion recovery (FLAIR) magnetic resonance imaging (MRI) sequences. WMHs are a frequent finding following traumatic brain injury (TBI) and have been correlated with functional outcome and injury severity in both pediatric [1, 2] and adult [3–6] populations.

Further research involving WMHs has shown that regional distribution and volume of WMHs have been shown to possess prognostic value in the TBI patient [2, 6–8]. Specifically, lesion volume in corpus callosum correlates with functional scores in the acute phase following injury, while lesion volume in frontal lobes correlates with scores at 1 year following injury [6]. Further, volume of FLAIR lesions within the corpus callosum, brainstem, and thalamus in patients with severe TBI correlates with Glasgow Outcome-Extended (GOS-E) scores—a numeric groupwise assessment used to classify "outcome" in TBI patients where "outcome" refers to the spectrum of possible prognoses from death to disability to recovery [4]. Additionally, in patients who are comatose following severe TBI the regional distribution of FLAIR lesions within the pons, midbrain, hypothalamus, basal forebrain, parietal, temporal, occipital lobes, and insula along with the observation of grasping or chewing behavior are associated with poor outcome [7].

Despite the above findings which demonstrate that WMHs have potential prognostic value, they are not routinely employed as a diagnostic measure in clinical practice. Performing a comprehensive manual counting of number and distribution of lesions in the clinical setting is simply not practical. The development of such lesion quantification approaches may allow for the practical inclusion of this type of information within routine radiological practice. In this work, we present an automated framework for quantification of WMHs in multi-modal MRI using the random forest machine learning technique.

Random forests for WMH segmentation

The random forests framework [9] is a popular machine learning technique that has demonstrated significant utility for supervised segmentation tasks (e.g., normal human brain segmentation [10]) and other computer vision applications [11]. Random forest-based paradigms have been successfully employed in the delineation of other neuropathologies [12–17] for both single and multi-modal acquisition protocols.

Random forests are conceptually straightforward [9]. The basic component of the random forest paradigm is the "decision tree" often represented by a flowchart or graph where internal nodes represent "tests", or

decisions, and the edges represent the outcome of those tests. The final, or end, nodes represent the various classifications produced by traversal through the decision tree. For the proposed application, individual voxels (and their corresponding feature values) are introduced at the root of a particular decision tree and traverse the edges and internal nodes ultimately ending up at one of the classification nodes according to the tests at each internal node. A single random forest model will consist of many such trees (often refered to as an "ensemble"). Although decision trees had been extensively studied, the success of employing collections of such weak learners for boosting machine learning performance (e.g., AdaBoost) [18, 19] influenced the similarly styled conglomeration of decision trees into "forests" with randomized node optimization [20, 21]. Finally, Breiman [9] improved accuracy by random sampling of training data (i.e., "bagging") resulting in the current random forest technique applied here. As voxels and their feature values are "pushed" through each decision tree in the forest, votes for each label are accumulated and converted to probability values for all classification possiblilities at each voxel location.

In this work, we develop a concatenated random forest framework with a feature image set (both spatial and intensity-based) for segmenting WMHs in a large TBI cohort. The entire framework is built on the well-known open-source Advanced Normalization Tools (ANTs)[1] and ANTsR[2] toolkits. Further motivating this research is the availability of several large publicly available imaging data sets that permits testing reproducibility of this automated routine for WMH segmentation and quantification.

Materials and Methods

Imaging

MR images utilized for this initial report were acquired from a single scanner involved in the Chronic Effects of Neurotrauma Consortium's (CENC) observational study (see Walker et al., this issue). Briefly, participants were Operation Iraqi Freedom/Operation Enduring Freedom (OIF/OEF) era Service Members and Veterans between the ages of 18-60 years with prior combat exposure and deployment(s). The feature images were derived from MR acquisitions of 26 subjects aged 39.6 ± 8.1 years (range 28–58 years). Within this cohort, 24 (92%) were considered positive for TBI based upon the potential concussive events (PCE) interview process described in detail in Walker et al., this issue. All lesions were isolated in the white matter of the cerebrum. Table 1 provides a descriptive statistical summary of the variation in lesion load across the selected cohort.

Images were acquired on a Philips 3.0T Ingenia system with an 8-channel SENSE head coil (Philips Medical Systems, Best, Netherlands). 3D FLAIR sequences were acquired with a turbo spin echo inversion recovery

[1] https://github.com/stnava/ANTs

[2] https://github.com/stnava/ANTsR

sequence with the following parameters: repetition time (TR) = 4800 ms, echo time (TE) = 325 ms, inversion time (TI) = 1650 ms; 170 sagittal slices with a 1.2 mm slice thickness, 256 $\times$ 256 acquisition matrix, and 256 $\times$ 256 mm FOV. 3D T1-weighted sequences were acquired with a fast field echo (FFE) sequence with the following parameters: TR = 6.8 ms, TE = 3.2 ms, echo train length (ETL) = 240; flip angle = 9°, 170 sagittal slices with a 1.2 mm slice thickness, 256x240 acquisition matrix, and 256 $\times$ 256 mm FOV. In addition, 3D T2-weighted images were acquired with a turbo spin echo sequence with the following parameters: TR = 2500 ms, TE = 245 ms, ET: = 133; 170 sagittal slices with a 1.2 mm slice thickness, 256 $\times$ 256 acquisition matrix, and 256 x 256 mm FOV.

The first author (J. R. S.) performed the manual WMH tracings for all 26 subjects. J. R. S. is a radiologist certified by the American Board of Radiology, with a certificate of advanced qualification in vascular and interventional radiology, over 18 years of research experience in TBI, and 6 years of clinical imaging experience. All multi-modal MR dicom image slices were converted to the nifti file format.[3] All nifti image volumes for each subject were rigidly aligned to the T1 image of that subject using the ANTs software [22]. The normalized MRI volumes were then provided to J. R. S. who traced each lesion using the ITK-SNAP tool [23] which has multi-image overlay capabilities for visualizing all modalities in all three canonical views.

Quantitative analysis

Figure 1 provides a graphical overview of the proposed workflow. The major components include offline generation of symmetric multimodal templates, the creation of feature images from the training data which are then employed for statistical prediction using a concatenated random forest framework. This framework involves the use of two random forest models where the outcome (i.e., the tissue probability estimates) of the first RF model application, which we denote as "Stage1", is used as input (along with the original set of feature images) to a refinement RF model segmentation which we denote as "Stage 2". Once these offline steps are performed, a new (i.e., unsegmented) subject can then be processed using the proposed pipeline.

Symmetric multi-modal templates

Following [24] and [17], optimally derived templates serve for both brain tissue segmentation (for the derivation of tissue prior probability maps) and generation of asymmetry feature images. For this work we use the multi-modal data available from the public MMRR data set [25]. We used all 46 multi-modal acquisitions of that study to produce a multi-modal template according to the procedure described in [26] which results in a mean (in terms of both shape and intensity) multivariate template representing the entire cohort. Mid-canonical slices of the FLAIR, T1, and T2 components are illustrated in Figure 2.

[3] http://nifti.nimh.nih.gov/nifti-1

Feature images for WMH segmentation

Crucial to these supervised segmentation approaches are the creation and selection of "features" as input (i.e., feature images constructed from the training data) in conjunction with expertly identified structures of interest (i.e., WMHs) for model construction. For the targeted application in this work, tissue classification is 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 thus permitting a regression or classification solution. Since this procedure is performed at the voxelwise level, intensity information alone is insufficient for good segmentation performance due to the lack of spatial context. For example, as pointed out in [27], 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.

Supervised methodologies are uniquely characterized, in part, by the feature images that are used to identify the regions of interest. In Table 2, 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 WMHs. In addition Figure 3 provides a representation of a set of feature images for a single subject analyzed in this work. Note that in this work we categorize the brain parenchyma with seven labels:

- cerebrospinal fluid (label 1),
- gray matter (label 2),
- white matter (label 3),
- deep gray matter (label 4),
- brain stem (label 5),
- cerebellum (label 6), and
- white matter hyperintensities (label 7).

As mentioned previously, input for each subject comprises FLAIR, T1-, and T2-weighted acquisitions. The T1 and T2 images are rigidly registered to the FLAIR image using the open-source Advanced Normalization Tools (ANTs) [22]. The aligned images are then preprocessed using the denoising algorithm of [28] followed by N4 bias correction [29] which are then normalized to the intensity range $[0, 1]$. Although we could have used an alternative intensity standardization algorithm (e.g., [30]), we found that a simple linear rescaling produced better results similar to previous work[17].

The T1 image is then processed via the ANTs brain extraction and normal tissue segmentation pipelines [24]. The result is a mask delineating the brain parenchyma and probabilistic estimates of the CSF, gray matter, white matter, deep gray matter, brain stem, and cerebellum [31]. These provide the expertly annotated labels for the first six tissue labels given above. Tissue prior probability maps for segmentation are from multi-model optimal symmetric shape/intensity templates [17] created from the public MMRR data set [25] (cf Figure 2).

Feature values include the preprocessed FLAIR, T1, and T2 image voxel intensities. We also calculate a set of neighborhood statistics (mean, standard deviation, and skewness) feature images using a Manhattan radius of one voxel given the typical size of individual WMHs. For each of the preprocessed images, we calculate the difference in intensities with the corresponding warped template component. Previous success in the international brain tumor segmentation competition [32] was based on an important set of intensity features that were created from multi-modal templates mentioned previously [17] and listed in Table 2. We employ the same strategy here.

To take advantage of the gross bilateral symmetry of the normal brain (in terms of both shape and intensity), and the fact that WMHs do not generally manifest symmetrically across hemispheres, we use the symmetric templates to compute the contralateral intensity differences as an additional intensity feature.

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 [33] 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. Similar feature images were used in [34] although, unlike the proposed framework, this previous work lacks normalization to the standard coordinate system provided by the template to dramatically improve spatial specificity across all subjects. To generate these images, the T1 image of each subject is registered to the T1 template component using a B-spline variant [35] of the well-known ANTs Symmetric Normalization (SyN) algorithm [36]. Using the derived transforms, the template coordinate images are warped back to the space of the individual subject.

Stacked (concatenated) random forests for improved segmentation performance

In previous brain tumor segmentation work [17], it was demonstrated that a concatenated supervised approach, whereby the prediction output from the first random forest model serves as partial input for a second random forest model, can significantly improve segmentation performance. We do the same thing for the work described here where we employ two stacked random forests (or two "stages"). The Stage 1 feature images of

the training data (as described previously) are used to construct the Stage 1 model. The training data Stage 1 features are then used to produce the voxelwise "voting maps" (i.e., the classification count of each decision tree for each tissue label) via the Stage 1 random forest model. All the Stage 1 features plus the Stage 1 voting maps are used as input to the Stage 2 model. In addition, we use the Stage 1 voting maps as tissue priors (i.e., probabilistic estimates of the tissue spatial locations) for a second application of the 6-tissue segmentation algorithm with an additional Markov Random Field spatial prior (MAP-MRF) [31]. In order to maximize the spatial information for the n-tissue segmentation process following the voxelwise RF classification of Stage 1, we use all three aligned preprocessed images for multivariate segmentation during the second stage. The resulting seven posterior probability images constitute a third additional feature image set for Stage 2.

Implementation

As pointed out in a recent comprehensive lesion segmentation review [37], although the number of algorithms reported in the literature is quite extensive, there were only four publicly available segmentation algorithms at the time of writing this article. In contrast to the current work, none are based on supervised learning. As we did for our brain tumor segmentation algorithm [17], all of the code described in this work is publicly available through the open-source ANTs/ANTsR toolkits. Through ANTsR (an add-on toolkit which, in part, bridges ANTs and the R statistical project) we use the *randomForest* package [38] using the default settings with 2000 trees per model and 500 randomly selected samples per label per image. Note that we saw little variation in performance when these parameters were changed (i.e. up to 1000 random samples and as little as 1000 trees) which is consistent with our previous experience.

In addition, similar to our previous offering,[4] we plan on creating a self-encapsulated example to showcase the proposed methodology which will also be available on github.[5] The fact that the data will also be made available through the Federal Interagency Traumatic Brain Injury Research (FITBIR) repository along with the manual labelings will facilitate reproducibility on the part of the reader as well as any interest in extending the proposed framework to other data sets.

Evaluation protocol overview

In order to evaluate the protocol described, we performed a leave-one-out evaluation using the data acquired from the 24 subjects described above. Initial processing included the creation of all Stage 1 feature images for all subjects. The initial brain segmentation of each T1 image and the manual white matter hyperintensity tracings were combined to provide the truth labels for the training data. The "truth" labels are the seven

[4] https://github.com/ntustison/ANTsAndArboles

[5] https://github.com/ntustison/WatchMeHyperventilate

anatomical regions given above.

The leave-one-out procedure is as follows:

- Create Stage 1 feature images for all 24 subjects.
- For each of the 24 subjects:
  - sequester the current subject and corresponding feature images.
  - construct the Stage 1 random forest model from the remaining 23 subjects.
  - apply the Stage 1 random forest model to the feature images of the 23 training subjects.
  - the previous step produces the Stage 1 voting maps for all seven labels.
  - for each of the 23 subjects, perform a Bayesian-based segmentation with an MRF spatial prior using the seven voting maps as additional tissue priors.
  - construct the Stage 2 random forest model from all the Stage 1 feature images, seven voting maps, and seven posterior probability maps from the previous step.
  - send the sequestered subject through the random forest models for both stages.
  - compare the final results with the manually-defined white matter hyperintensity regions.

Several measures have been employed in the literature for evaluating automated white matter lesion segmentation involving such quantities as lesion load, voxel-based overlap measures (such as the Dice similarity coefficient), and lesion-based measures.[37]. For this work, due to the relatively small-size distribution of the lesion load in our data cohort (see Table 1), we used four lesion-based measures: sensitivity, positive predictive value (i.e., precision), F_1 score, and relative volume difference. The first three quantities are based on the number of false positives (TN, false negatives (FN), and true positives (TP) in terms of identified lesions. It should be noted that the number of true negatives is not readily incorporated into measures of accuracy as the quantity of "true negatives" (i.e., correctly identified normal brain tissue) would severely skew accuracy assessments. The F_1 score is an assessment of accuracy which takes into account both the sensitivity,

$$sensitivity = \frac{TP}{TP + FN},$$

and the positive predictive value (PPV),

$$PPV = \frac{TP}{TP + FP},$$

such that F_1 is given by

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}.$$

The relative volume difference is calculated for each of the true positive lesions using the manual and predicted lesion volumes:

$$Relative\ volume\ difference = \frac{V_{manual} - V_{predicted}}{V_{manual}}.$$

To illustrate how the performance of our framework varies with lesion size, we calculated the above measures based on evenly split quantiles of the manual estimates of lesion volumes into 3 groups. These three size ranges (in terms of the number of voxels) are:

- Quantile 1: $[1 - 12)$,
- Quantile 2: $[12 - 28)$, and
- Quantile 3: $[28 - 551]$.

These quantiles are used to showcase the performance (cf Figure 5).

Results

White matter hyperintensity segmentation evaluation

In Figure 5 we provide the segmentation evaluations derived from the leave-one-out evaluation of the previously described TBI data over the three lesion volume ranges. These performance measures include sensitivity, positive predictive value, F_1 score, and relative volume difference. The three lesion size ranges over which these measures are computed are meant to illustrate the variation in performance with lesion size.

Smaller lesions (< 12 voxels) are more difficult to identify which is why the sensitivity for this range is more varied compared with the largest set of lesions (> 28 voxels). The first three measures are based on the identification of entire lesions. The relative volume difference provides a direct assessment of the accuracy of the volumetric estimate when comparing the manually identified lesions versus the automatically predicted lesions.

We averaged these measures over all lesions and all subjects which resulted in the following values:

- Stage 1
  - sensitivity = 0.70 $\pm$ 0.34
  - PPV = 0.42 $\pm$ 0.36
  - F_1 = 0.47 $\pm$ 0.36

  - relative volume difference = 43 $\pm$ 38%
- Stage 2
  - sensitivity = 0.68 $\pm$ 0.38
  - PPV = 0.51 $\pm$ 0.40
  - F_1 = 0.52 $\pm$ 0.36
  - relative volume difference = 43 $\pm$ 26%

Ranking feature importance

After performing the leave-one-out evaluation, we calculated the *MeanDecreaseAccuracy* feature values for each of the 24 subjects $\times$ 2 models per subject $= 48$ total models. This measure (per feature, per model) is calculated during the out-of-bag phase of the random forest model construction and quantifies the decrease in prediction accuracy from omitting the specified feature. In other words, this quantity helps determine the importance of a particular feature and, although we save such efforts for future work, this information provides us with guidance for future feature pruning and/or additions.

The resulting rankings for both Stages are given in Figures 6 and 7 where the values for the separate stages are averaged over the entire corresponding model set. In addition, we track the variance for each feature over all models to illustrate the stability of the chosen features during the evaluation. This latter information is illustrated as horizontal errors bars providing the 95^{th} percentile Note that the reader can cross reference Table 1 for identifying corresponding feature types and names.

Discussion

In evaluating the segmentation results, there are limited methodologies described in the literature for automatic segmentation of WMHs despite the numerous clinical studies [39, 40] exploring the connections between WMHs and TBI. However, we can extrapolate from similar application areas where such methodologies have a much longer history of development such as white matter lesion segmentation in multiple sclerosis (MS). In a recent literature review [37], 47 papers involving automated segmentation of white matter lesions in MS. The most widely-used data for algorithmic evaluation consists of two cohorts from two different sites used in the MS Lesion segmentation challenge associated with the international MICCAI 2008 conference. For comparison, this challenge training data set has a mean lesion load of 204 ($\pm$ 752) mm^3 per lesion (compared with our mean lesion load of 33.8 ($\pm$ 55.4) voxels $\times$ 1.2 mm^3 per voxel = 40.56 ($\pm$ 66.48) mm^3 per lesion) and the resolution is almost twice what is used in this study (i.e., 0.5 $\times$ 0.5 $\times$ 0.5). Note that our relative volume difference numbers are comparable to the relative volume difference *between raters* of 68 percent (cf Figure 5 (d)) [41].

Regarding the feature rankings, it is interesting to note some of the other top performing features for Stage 1. The contralateral difference FLAIR image is highly discriminative over the set of evaluation random forest models (see Figure 8). This accords with the known clinical relevance of FLAIR images for identifying white matter hyperintensities and the fact that such pathology does not typically manifest symmetrically in both hemispheres. Interestingly, the posterior maps for the deep gray matter are extremely important for accurate white matter hyperintensity segmentation. Perhaps the spatial specification of deep gray matter aids in the removal of false positives. Inspection of the bottom of the plots demonstrates the lack of discriminating features associated with the T1 image which is also well-known in the clinical literature.

As described earlier, for Stage 2, we used the output random forest voting maps from Stage 1 as both features themselves and as priors for input to a Bayesian-based segmentation with an additional MRF spatial prior. In Figure 7, the voting maps are labeled as "*RFStage1VotingMaps*" where the final numeral is associated with the brain parenchymal labeling given previously. Similarly, the additional RF prior segmentation feature probability maps are labeled as "*RFBrainSegmentationPosteriors*". The Stage 2 feature importance plot follows similar trends as that for Stage 1 with the T1 images not contributing much to the identification of white matter hyperintensity voxels. The initial voting maps from Stage 1 are extremely important with the top 3 being the estimated locations of the 1) gray matter, 2) white matter, and 3) white matter hyperintensities. Since these tissue type can be conflated based on intensity alone it is intuitive that such features would be important.

Conclusions

The current communications describes a supervised statistical learning methodology for identifying WMHs within multimodal MR brain imaging. This effort utilized information acquired from the manual segmentation of WMHs from FLAIR images to help build two-stage ensembles of decision trees for the automated identification of these lesions. Although only a single expert was used to produce the manual labelings, our intent is to further refine the proposed paradigm by crowdsourcing with feedback from other experts who interact with both the data and methodology. Also, we recognize that only a single site was used for evaluating the proposed framework. However, we are currently processing other site data with the models developed for this work and the results look promising since the developed features are site-agnostic.

As far as we know, this is the first report utilizing a novel random forest approach to identify WMHs in a cohort of TBI patients. TBI WMHs tend to be more difficult to segment than MS lesions as the former tend to be smaller with an overall smaller lesion load. Also, enhancement protocols with the former tend to be less successful than with the latter. As mentioned previously, the work in MS lesion segmentation is extensive with a handful of techniques being publicly available.

Two major meta-analyses of WMHs have been published covering the periods prior to [39] and after 2010 [40]. Debette & Markus [39] found that the presence of WMHs was related to subsequent cognitive decline, a higher risk of developing dementia, stroke, and of mortality. Lesion volume at baseline was also predictive of cognitive decline. Kloppenborg et al. [40] reviewed 23 cross-sectional studies reporting MRI and concurrent neuropsychological results in patients with heterogeneous diagnoses but without previously diagnosed cognitive impairment. This review found that WMHs were associated with cognitive deficit (effect size of -0.10, 95% CI: -0.13 to -0.08) after controlling for age.

Despite the potential clinical significance of WMHs these lesions receive little attention in current clinical workflows. When reported in a standard neuroradiologist interpretation, they are typically handled as incidental findings and are assigned little clinical significance. This likely reflects the impracticality of performing a detailed assessment of number, volume, and distribution within a qualitative neuroradiologist interpretation as well as the lack of correlative information on how the presence and distribution of these lesions may inform a diagnosis and prognosis in the appropriate clinical setting. To date, automated or semi-automated tools for the detection of WMHs have lacked the specificity and efficiency for the mining of large-scale datasets to generate highly granular data on whether these lesions possess any true diagnostic or prognostic value in the setting of a specific disease process. The present communication describes a supervised statistical learning tool that is appropriate for the application to such large-scale datasets.

Acknowledgements

The authors wish to acknowledge all other members of the CENC Neuroimaging Steering Committee and CENC leadership (Drs. David X. Cifu, Ramon Diaz-Arrastia, and Rick Williams) for their support. We also gratefully acknowledge the assistance of Tracy Nolen, Chris Siege and Kevin Wilson. We would also like to thank the study participants and their family members. This project was jointly supported by the Department of Defense (W81XWH-13-2-0095), the U.S. Department of Veterans Affairs (I01 CX001135 and I01 RX 002174), as well as USUHS Grant HU 0001-08-0001.

Declaration of Interest/Disclaimer

The authors report no financial disclosures or conflicts of interest. The views expressed here are those of the authors and do not necessarily reflect the official policy of position of the Department of the Navy, Department of Defense, nor the U.S. Government. This work was prepared as a part of official duties; Title 17 USC §105 provides that Copyright protection under this title is not available for any work of the U.S. Government. Title 17 USC §101 defines a US Government work as a work prepared by a military service member of employee of the US Government as part of that person's official duties.

References

1. Bigler, E. D., Abildskov, T. J., Petrie, J., Farrer, T. J., Dennis, M., Simic, N., Taylor, H. G., Rubin, K. H., Vannatta, K., Gerhardt, C. A., Stancin, T., and Owen Yeates, K. "**Heterogeneity of Brain Lesions in Pediatric Traumatic Brain Injury**" *Neuropsychology* 27, no. 4 (2013): 438–51. doi:10.1037/a0032837

2. Smitherman, E., Hernandez, A., Stavinoha, P. L., Huang, R., Kernie, S. G., Diaz-Arrastia, R., and Miles, D. K. "**Predicting Outcome After Pediatric Traumatic Brain Injury by Early Magnetic Resonance Imaging Lesion Location and Volume**" *J Neurotrauma* 33, no. 1 (2016): 35–48. doi:10.1089/neu.2014.3801

3. Marquez de la Plata, C., Ardelean, A., Koovakkattu, D., Srinivasan, P., Miller, A., Phuong, V., Harper, C., Moore, C., Whittemore, A., Madden, C., Diaz-Arrastia, R., and Devous, M., Sr. "**Magnetic Resonance Imaging of Diffuse Axonal Injury: Quantitative Assessment of White Matter Lesion Volume**" *J Neurotrauma* 24, no. 4 (2007): 591–8. doi:10.1089/neu.2006.0214

4. Moen, K. G., Brezova, V., Skandsen, T., Håberg, A. K., Folvik, M., and Vik, A. "**Traumatic Axonal Injury: The Prognostic Value of Lesion Load in Corpus Callosum, Brain Stem, and Thalamus in Different Magnetic Resonance Imaging Sequences**" *J Neurotrauma* 31, no. 17 (2014): 1486–96. doi:10.1089/neu.2013.3258

5. Ding, K., Marquez de la Plata, C., Wang, J. Y., Mumphrey, M., Moore, C., Harper, C., Madden, C. J., McColl, R., Whittemore, A., Devous, M. D., and Diaz-Arrastia, R. "**Cerebral Atrophy After Traumatic White Matter Injury: Correlation with Acute Neuroimaging and Outcome**" *J Neurotrauma* 25, no. 12 (2008): 1433–40. doi:10.1089/neu.2008.0683

6. Pierallini, A., Pantano, P., Fantozzi, L. M., Bonamini, M., Vichi, R., Zylberman, R., Pisarri, F., Colonnese, C., and Bozzao, L. "**Correlation Between MRI Findings and Long-Term Outcome in Patients with Severe Brain Trauma**" *Neuroradiology* 42, no. 12 (2000): 860–7.

7. Weiss, N., Galanaud, D., Carpentier, A., Tezenas de Montcel, S., Naccache, L., Coriat, P., and Puybasset, L. "**A Combined Clinical and MRI Approach for Outcome Assessment of Traumatic Head Injured Comatose Patients**" *J Neurol* 255, no. 2 (2008): 217–23. doi:10.1007/s00415-008-0658-4

8. Levin, H. S., Williams, D., Crofford, M. J., High, W. M., Jr, Eisenberg, H. M., Amparo, E. G., Guinto, F. C., Jr, Kalisky, Z., Handel, S. F., and Goldman, A. M. "**Relationship of Depth of Brain Lesions to Consciousness and Outcome After Closed Head Injury**" *J Neurosurg* 69, no. 6 (1988): 861–6.

doi:10.3171/jns.1988.69.6.0861

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

10. Yi, Z., Criminisi, A., Shotton, J., and Blake, A. "**Discriminative, Semantic Segmentation of Brain Tissue in MR Images**" *Med Image Comput Comput Assist Interv* 12, no. Pt 2 (2009): 558–65.

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

12. 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

13. 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

14. 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):

15. 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.

16. 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.

17. 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* 13, no. 2 (2015): 209–25. doi:10.1007/s12021-014-9245-2

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

19. 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.

20. 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

21. Amit, Y. and Geman, D. "**Shape Quantization and Recognition with Randomized Trees**" *Neural*

Computation 9, (1997): 1545–1588.

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. Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., and Gerig, G. "**User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability**" *Neuroimage* 31, no. 3 (2006): 1116–28. doi:10.1016/j.neuroimage.2006.01.015

24. 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

25. 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

26. Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., and Gee, J. C. "**The Optimal Template Effect in Hippocampus Studies of Diseased Populations**" *Neuroimage* 49, no. 3 (2010): 2457–66. doi:10.1016/j.neuroimage.2009.09.062

27. 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

28. 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

29. 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

30. 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

31. 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

32. 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

33. 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

34. 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

35. 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

36. 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

37. García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L. "**Review of Automatic Segmentation Methods of Multiple Sclerosis White Matter Lesions on Conventional Magnetic Resonance Imaging**" *Med Image Anal* 17, no. 1 (2013): 1–18. doi:10.1016/j.media.2012.09.004

38. Liaw, A. and Wiener, M. "**Classification and Regression by randomForest**" *R News* 2/3, (2002): 18–22.

39. Debette, S. and Markus, H. S. "**The Clinical Importance of White Matter Hyperintensities on Brain Magnetic Resonance Imaging: Systematic Review and Meta-Analysis**" *BMJ* 341, (2010): c3666.

40. Kloppenborg, R. P., Nederkoorn, P. J., Geerlings, M. I., and Berg, E. van den. "**Presence and Pro-**

gression of White Matter Hyperintensities and Cognition: A Meta-Analysis" *Neurology* 82, no. 23 (2014): 2127–38. doi:10.1212/WNL.0000000000000505

41. Ginneken, B. V., Heimann, T., and Styner, M. "**3D Segmentation in the Clinic: A Grand Challenge**" *3D segmentation in the clinic: A grand challenge* (2007): 7–15.