



An improved algorithm of white matter hyperintensity detection in elderly adults



T Ding^{*,a}, AD Cohen^a, EE O'Connor^b, HT Karim^a, A Crainiceanu^c, J Muschelli^d, O Lopez^a, WE Klunk^a, HJ Aizenstein^a, R Krafty^a, CM Crainiceanu^d, DL Tudorascu^a

^a University of Pittsburgh, Pittsburgh

^b University of Maryland, Baltimore

^c United States Naval Academy, United States

^d Johns Hopkins University, United States

ARTICLE INFO

Keywords:

Brain
WMH segmentation
Alzheimer's disease
OASIS
MRI
Statistical methods neuroimaging

ABSTRACT

Automated segmentation of the aging brain raises significant challenges because of the prevalence, extent, and heterogeneity of white matter hyperintensities. White matter hyperintensities can be frequently identified in magnetic resonance imaging (MRI) scans of older individuals and among those who have Alzheimer's disease. We propose OASIS-AD, a method for automatic segmentation of white matter hyperintensities in older adults using structural brain MRIs. OASIS-AD is an approach evolved from OASIS, which was developed for automatic lesion segmentation in multiple sclerosis. OASIS-AD is a major refinement of OASIS that takes into account the specific challenges raised by white matter hyperintensities in Alzheimer's disease. In particular, OASIS-AD combines three processing steps: 1) using an eroding procedure on the skull stripped mask; 2) adding a nearest neighbor feature construction approach; and 3) applying a Gaussian filter to refine segmentation results, creating a novel process for WMH detection in aging population. We show that OASIS-AD performs better than existing automatic white matter hyperintensity segmentation approaches.

1. Introduction

Structural magnetic resonance imaging (sMRI) is widely used in clinical practice to diagnose and manage neurodegenerative diseases, such as Alzheimer's disease (AD), that primarily affects older adults. However, using automatic MRI processing techniques for brain images of older adults can be challenging due to white matter hyperintensities (WMHs), atrophy, and other aging pathologies. We focus on WMHs in this article. WMHs are areas in the white matter of the brain that appear hyperintense on a T2-weighted-Fluid-Attenuated Inversion Recovery (T2-FLAIR) scan and appear hypointense on a T1-weighted scan as compared to normal appearing white matter. Because the size, shape, and positioning of WMH regions are highly heterogeneous in the brain of older adults, using segmentation approaches developed for the brain of healthy young adults could lead to substantial tissue miss-classification. For example, segmentation of brain imaging data into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) is a crucial processing step in brain imaging studies. Existing automatic segmentation approaches were developed for the brain images of healthy young adults, which generally do not contain WMHs. When

WMHs are present, automatic segmentation tools that use T1-weighted images may incorrectly classify WMHs as GM since both appear hypointense. Therefore, large WMH areas could artificially increase the estimated GM volumes in cross-sectional studies and could lead to severe underestimation of GM loss in longitudinal studies. This can be a major problem in studies that use GM volume as a general marker of brain atrophy. Moreover, WMHs are more prevalent in older adults and women (Van Den Heuvel et al., 2004), which may lead to differential tissue classification performance in specific subgroups.

WMHs appear in a variety of studies both in individuals who are clinically symptomatic or asymptomatic. In particular, WMHs are pervasive in studies of aging, AD, bipolar disorder (Pillai et al., 2002), and stroke (Wong et al., 2002). WMH segmentation is crucial for correct tissue classification as well as for estimating the WMH volume directly, as this is often used as a marker of cerebrovascular diseases. In this article we focus on WMH segmentation in the aging brain in general and aging brain affected by AD in particular.

A review of existing WMH segmentation methods is provided in Caligiuri et al. (2015). The methods are divided into three categories: (1) supervised learning algorithms using manually-labeled tracings of

* Corresponding author.

E-mail address: dlt30@pitt.edu (T. Ding).

<https://doi.org/10.1016/j.nicl.2019.102151>

Received 14 October 2019; Received in revised form 10 December 2019; Accepted 26 December 2019

Available online 27 December 2019

2213-1582/ © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

WMHs; (2) unsupervised learning algorithms using unlabeled manual tracings; and (3) semi-automated algorithms with various degrees of user intervention. Supervised classification algorithms include: k-nearest neighbors (kNN), non-parametric classification using the k closest training samples in the feature space (Anbeek et al., 2004), support vector machines (SVM) (Lao et al., 2008), Bayesian methods that combine multivariate signal intensity and spatial information (Herskovits et al., 2008), artificial neural networks (ANN) using multi-sequence images (Dyrby et al., 2008), Gaussian mixture models (Simões et al., 2013), logistic regression of multi-sequence images (Sweeney et al., 2013), adaptive intensity threshold search (Yoo et al., 2014), and deep learning approaches: (Kuijf et al., 2016; Moeskops et al., 2018; Raidou et al., 2014), convolutional neural networks (Ghafoorian et al., 2017). Moeskops (Moeskops et al., 2018) and colleagues evaluated different deep learning algorithms and their evaluating metrics were higher than previously reported in the literature using Visual Analytics methods (Kuijf et al., 2016; Raidou et al., 2014) on the same data.

Unsupervised classification algorithms include: a two-level fuzzy inference system based on proton density (PD) and T2-FLAIR images (Admiraal-Behloul et al., 2005), a fuzzy connected algorithm combined with image registration (Wu et al., 2006), and a geostatistical fuzzy c-means clustering algorithm (Anitha et al., 2012). Semi-automated algorithms include: region growing using adaptive thresholding (Itti et al., 2001), bispectral fuzzy class means (Sheline et al., 2008), and semi-automatic peak identification on the 2D histogram of T1 and T2 intensities (Sheline et al., 2008). Caligiuri et al. (2015) concluded that a good WMH segmentation method should include a comprehensive image preprocessing pipeline based on multi-sequence data that takes into account spatial information about lesions and corrects for false positives.

We propose OASIS-AD, an automatic supervised approach based on logistic regression and careful development of the feature space. OASIS-AD is an approach evolved from OASIS (Sweeney et al., 2013) (Automated Statistical Inference for Segmentation), which was developed for automatic lesion segmentation in multiple sclerosis (MS). OASIS-AD is a major refinement of OASIS that takes into account the specific challenges raised by WMH in AD. In particular, OASIS-AD combines three processing steps: 1) using an eroding procedure on the skull stripped mask; 2) adding a nearest neighbor feature construction approach; and 3) applying a Gaussian filter to refine segmentation results, thus creating a novel WMH detection procedure, OASIS-AD. We show that OASIS-AD performs better than existing WMH segmentation approaches using gold standard segmentations.

2. Materials and methods

In this section we introduce OASIS-AD, an automated method of WMH segmentation with application to AD patients and older adults. OASIS-AD has three main components: (a) development of a binary brain tissue mask; (b) normalization of MRI intensities and creation of smoothed volumes; and (c) two-step modeling of WMH regions. The first step of modeling consists of training a richly parameterized logistic regression model using the data preprocessed in the (a) and (b) components of OASIS-AD. The second step consists of refining the voxel-level probability map generated in the first step to shrink WMH regions, and smoothing the probability map to reduce the false positive rate. A flowchart of OASIS-AD is presented in Fig. 1.

2.1. Study participants

We have randomly selected a sample of 20 older individuals from our ongoing Normal Aging study previously described in Karim et al. (2019); Nadkarni et al. (2019). The selected sample included 20 cognitively normal study participants at the time of scanning. The average age in our sample is 81.2 (SD = 7.15), with an average

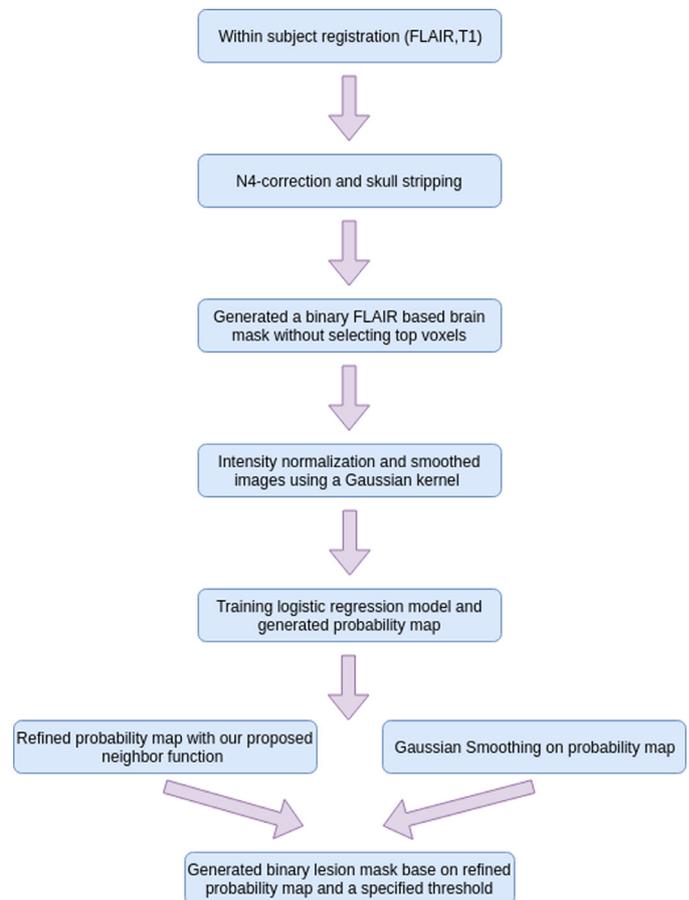


Fig. 1. OASIS-AD procedure.

education equal to 14.2 years (SD = 2.44), 70% of the sample are females, 85% white and 15% african american. In the next sections we describe the OASIS-AD steps in greater detail.

2.2. Image preprocessing

The image preprocessing used fs^lr” (Muschelli et al., 2015) in Neuroconductor” (Muschelli et al., 2018), a comprehensive R” environment for imaging processing tools. The fs^lr” package wraps the FMRIB Software Library (FSL 5.0) (<https://fsl.fmrib.ox.ac.uk/fsl>) into the R” language. The preprocessing steps were applied in the following order:

1. Perform within-subject coregistration of the T1-weighted image to the T2-FLAIR image.
2. Apply N4-bias-correction (Avants, 2019; Tustison et al., 2010) to the registered T1-weighted image.
3. Conduct skull stripping using FSL BET (Brain Extraction Tool) (Smith, 2002) on the registered and N4 corrected T1-weighted image.
4. Erode the brain mask with a default $5 \times 5 \times 5$ kernel box.

Eroding a binary mask, A , with a kernel, B , centered at C consists of moving B by sliding its center C over all voxels in A . If all voxels in B are contained in A then the location of the center C is labeled as 1; otherwise, it is labeled 0 (erosion) (Haralick et al., 1987). The fs^lerode package in fs^lr (Muschelli et al., 2015) was used for the erosion procedure.

2.3. Intensity normalization

Using a method similar to the one used by [Shinohara et al. \(2012\)](#), images intensities for both T1-weighted and T2-FLAIR images were normalized as follows:

$$f_i^N(v) = \frac{f_i(v) - \mu_{i,M}}{\sigma_{i,M}},$$

where $\mu_{i,M}$ and $\sigma_{i,M}$ are the mean and standard deviation of the pre-processed image intensities for subject i from modality M . Note that [Shinohara et al. \(2012\)](#) used the normally appearing white matter (NAWM) as the reference set for normalization, which would require at least partial segmentation of NAWM; here we avoid this problem by using the entire brain as reference.

2.4. Smoothed volumes

Similarly to the original OASIS approach, we used smoothed volumes by applying Gaussian kernel smoothers both to the T1-weighted and T2-FLAIR images. Two 3D Gaussian filters with window sizes of 10 and 20 mm, respectively, were used to capture local inhomogeneity patterns that were not accounted by the N4 correction. We denote by $GM_i^N(v, k)$ the smoothed volume for subject i , image modality M , and kernel size k at voxel v . We fit models that include these smoothed volumes as well as models that do not (labeled reduced models), as the aggressive smoothing might actually remove subtle differences specific to the WM/GM boundary, which could further induce classification bias.

2.5. Logistic regression model

Two logistic regression models were used: **M1**, a full model based on OASIS and all the image modalities, and **M2**, a reduced model. The **M1** model for the probability that a voxel v for study participant i is in WMH area is

$$\begin{aligned} \mathbf{M1}: \text{logit}(P\{W_i(v) = 1\}) = & \beta_0 + \beta_1 \text{FLAIR}_i^N(v) + \beta_2 \text{GFLAIR}_i^N(v, 10) \\ & + \beta_3 \text{GFLAIR}_i^N(v, 20) + \beta_4 \text{T1}_i^N(v) \\ & + \beta_5 \text{GT1}_i^N(v, 10) + \beta_6 \text{GT1}_i^N(v, 20) \\ & + \beta_7 \text{FLAIR}_i^N(v) \text{GFLAIR}_i^N(v, 10) \\ & + \beta_8 \text{FLAIR}_i^N(v) \text{GFLAIR}_i^N(v, 20) \\ & + \beta_9 \text{T1}_i^N(v) \text{GT1}_i^N(v, 10) \\ & + \beta_{10} \text{T1}_i^N(v) \text{GT1}_i^N(v, 20). \end{aligned}$$

Model **M2** with the reduced predictors set is:

$$\mathbf{M2}: \text{logit}(P\{W_i(v) = 1\}) = \beta_0 + \beta_1 \text{FLAIR}_i^N(v) + \beta_2 \text{T1}_i^N(v),$$

where $\text{FLAIR}_i^N(v)$ is the normalized voxel's v FLAIR value, while $\text{GFLAIR}_i^N(v, 10)$ and $\text{GFLAIR}_i^N(v, 20)$ are smoothed normalized voxel's v FLAIR values with Gaussian kernels of size 10 mm and 20 mm, respectively. Notation for the other modalities follows the same convention.

2.6. Probability map refinement

The logistic regression models introduced in [Section 2.5](#) are used to produce an initial probability map for WMH at the voxel level. This probability map is then refined to reduce the false positive detection rate using two additional techniques: *Nearest Neighbor Refinement* and *Gaussian Filter Refinement* to remove false positives. We describe these in the next two sections.

2.6.1. Nearest neighbor refinement

The Nearest Neighbor Refinement (NNR) consists of first applying the FAST[™] ([Zhang et al., 2001](#)) algorithm, a popular brain tissue

segmentation based on T1-weighted images. The FAST[™] algorithm provides an estimated probability that the voxel v is in white matter, P_{wm}^v , gray matter, P_{gm}^v , CSF, P_{csf}^v , respectively. The sum of P_{wm}^v , P_{gm}^v and P_{csf}^v is equal to 1 for every voxel v . From these estimated tissue probability maps we estimate the tissue type of voxel v , denoted by T_v , as the tissue with highest probability at voxel v . Using the logistic models in [Section 2.5](#) we generate a probability that each voxel v is in WMH and denote it by P_{wmh}^v . We denote by N_v the 6 nearest neighbors (6NN) of voxel v . The idea is to use information from the neighboring voxels to reduce ‘‘speckling’’, the phenomenon where a few isolated voxels are identified as WMH when they should not be. Below we provide the detailed algorithm.

The algorithm starts with voxels whose estimated probability by FAST[™] of being in white matter is 1, $P_{wm}^v = 1$, and whose 6NN are all estimated to be in white matter by FAST[™], $T_v^{6NN} = wm$. Here, the last equality indicates that all entries of the six-dimensional vector T_v^{6NN} are estimated to be white matter by FAST[™]. For these voxels the estimated probability of the voxel being in WMH is exponentially reduced by simply raising the estimated probability of the voxel being in WMH using the logistic models to the power 10, $P_{wmh}^{rv} = (P_{wmh}^v)^{10}$. The net effect is to substantially reduce the estimated probability of this type of voxel to be in WMH. The second option is when the voxel is estimated by FAST[™] to be in white matter, $T_v = wm$, but not all its 6NN are estimated to be in white matter, $T_v^{6NN} \neq wm$. The last inequality indicates that at least one of the 6NN of the voxel v is not estimated to be in white matter by FAST[™]. In this case the estimated probability for the voxel to be in WMH is increased by raising it to the power average(P_{wm}^{6NN}), which is the average of the estimated probabilities for the voxel to be in white matter by FAST[™]. The average of these probabilities is a number less than one, indicating that the probability will be increased. The probability is increased more when there are more neighbors that are not estimated to be in white matter and when the estimated probabilities of these neighbors are further from 1, indicating increased probability that the voxels are not actually in the white matter. Both of these choices of powers were found empirically to work well and were validated using training/test data. If neither of these conditions are satisfied than the probability map obtained from the logistic models remains unchanged, $P_{wmh}^{rv} = P_{wmh}^v$.

2.6.2. Gaussian filter refinement

Once the NNR procedure is applied we apply a 3D Gaussian filter on the generated probability maps using the following sequence of operations: (1) create an eroded brain mask; (2) fill in the voxels in the eroded brain mask with the WMH probabilities estimated in [Section 2.6.1](#); and (3) apply a 3D Gaussian filter of size $5 \times 5 \times 5$ mm to the probability map on the eroded brain.

2.7. Binary segmentation and evaluation metrics

After creating the probability maps, a threshold value needs to be identified to classify voxels into classes. We use an approach proposed by [Valcarcel et al. \(2018\)](#), who proposed to use multiple threshold candidates and selected the optimal threshold based on the performance on the training set. We used the Dice Similarity Coefficient (DSC) ([Dice, 1945](#)) as the evaluation metric for selecting the optimal threshold.

Results were compared with manual segmentations performed by an experienced neuroradiologist, which provided the gold standard. The manual tracings of WMH were performed on 5 contiguous slices on the T2-FLAIR scans, the same for each subject. Models were compared in terms of the following metrics: (1) number of true positive voxels (TP); (2) number of false positive voxels (FP); (3) number of true negative voxels (TN); and (4) number of false negative voxels (FN). We computed four additional combined metrics commonly used for prediction performance evaluation ([Goutte and Gaussier, 2005](#)): (a) accuracy, defined as $\text{ACC} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$; (b) positive

predictive value, defined as $PPV = TP/(TP + FP)$; (c) true positive rate, defined as $TPR = TP/(TP + FN)$; (d) false positive rate, defined as $FPR = FP/(FP + TN)$; and (e) dice similarity coefficient, defined as $DSC = 2TP/(2TP + FP + FN)$ as well as 95% confidence interval (CI) computed using bootstrap. We also included the receiver operating characteristic curve (ROC curve), the precision-recall curve (PRC), and the area under these two curves (AUC) (Davis and Goadrich, 2006).

2.8. Comparison with other methods

We compared OASIS-AD with four other methods: OASIS (developed for MS lesion segmentation), MIMOSA (Valcarcel et al., 2018), the lesion segmentation tool (LST) (Schmidt, 2017), and the fuzzy connected algorithm of Wu et al. (2006) and labeled as fuzzy-c. Methods were compared on our study participants sample described in Section 2.1.

3. Results

Data were randomly split into training (15 study participants) and testing (5 study participants) and models were compared in terms of their performance on the testing data set.

3.1. Models performance comparison

3.1.1. OASIS-AD models comparison

We start by first evaluating the various types of the OASIS-AD model. Table 1 provides results for all model combinations considered, where the first column provides the label, while the second column provides the type of analysis conducted. For example, M2-NG is the OASIS-AD model using the logistic model M2 introduced in Section 2.5 combined with the NNR algorithm introduced in Section 1 followed by GFR algorithm introduced in Section 2.6.2. The acronym for this model could be OASIS-AD-M2-NG, but this is way too complex and we will use the M2-NG shortcut for presentation purposes, while understanding that all these models have the OASIS concept at the core with various refinements added to the resulting probability masks. The third column in Table 1 provides the optimal threshold obtained during training, while the fourth and fifth columns display the corresponding DSC and FPR on the test data.

Results indicate that the M1 model series (i.e., full models) outperforms the corresponding M2 series models (higher DSC and better FPR), but the differences are not very large. Taking into account that the M2 series models do not use smooth volumes, which can be time intensive on large datasets, we consider that the M2 series models provide an excellent first line approach for WMH segmentation. The M1-G model achieves the highest DSC (0.78), though it has a slightly higher FPR than the M1-NG model (0.009 compared to 0.007).

3.1.2. Comparisons with other models

Table 2 compares results for the best OASIS-AD model (M1-G) with the four other methods: OASIS, MIMOSA, LST and fuzzy-c, and Table 3 compares results for all the OASIS-AD models with other methods. For

Table 1
OASIS-AD models information.

OASIS-AD	Techniques	Optimal Threshold	DSC	FPR
M1	M1	0.17	0.72	0.017
M1-G	M1 + GFR	0.20	0.79	0.011
M1-NG	M1 + NNR + GFR	0.17	0.74	0.011
M1-GN	M1 + GFR + NNR	0.21	0.76	0.008
M2	M2	0.13	0.70	0.024
M2-G	M2 + GFR	0.14	0.77	0.017
M2-NG	M2 + NNR + GFR	0.13	0.72	0.016
M2-GN	M2 + GFR + NNR	0.16	0.74	0.013

the fuzzy-c method proposed by Wu's (Wu et al., 2006) we only have the binary brain masks and not the probability map. Therefore, it is not possible to compute the AUCs for fuzzy-c. The OASIS-AD (M1-G) model has the highest DSC at 0.78, with a 95% CI equal with (0.77, 0.79), with both MIMOSA and LST being close in second place (DSC = 0.71, 95% CI: (0.70, 0.77) and DSC = 0.76, 95% CI: (0.75, 0.80) respectively). The ROC-AUC (0.97) and ROC-PRC (0.86) for the M1-G model are substantially better than for MIMOSA (0.87 and 0.77, respectively) and LST (0.87 and 0.77, respectively.)

Fig. 2 displays the ROC and PRC for the four models OASIS-AD (M1-G), OASIS, MIMOSA, and LST, and Fig. 3 displays the ROC and PRC for all the models except fuzzy-c. The ROC curves are indistinguishable in the area of high specificity (specificity > 0.99), with the M1-G model performing slightly better. However, as specificity is allowed to be smaller (moving right on the 1-Specificity x-axis) the ROC of the OASIS-AD model is substantially better than for the other models. This indicates that small changes in specificity can lead to much larger improvements in sensitivity for the OASIS-AD model compared to the competing models. Both MIMOSA and LST seem to be tuned specifically for high specificity, whereas OASIS is has higher sensitivity for specificity areas that are not of practical interest. A similar result can be noted for the PRC in Fig. 2.

3.1.3. One slice comparison among models: case study

Fig. 4, showing true positives, false positives and false negatives color coded, compares the WMH segmentation results using two OASIS-AD methods (M1-G shown in panel C and M1-GN shown in panel D) with OASIS (panel E), MIMOSA (panel F), and LST (panel G), and fuzzy-c (panel H). Results are shown on one slice of a random subject from the training data. The corresponding FLAIR slice is shown in Panel A, while the manual segmentation of WMH is shown in Panel B. This slice contains both large and small contiguous WMHs regions and results indicate the good performance of both OASIS-AD approaches. The MIMOSA mask also looks very good, with slightly more speckling. The LST and OASIS estimators seem to contain many more spatially distributed false positive voxels, which may indicate a substantially different trade-off of false positives. Indeed, while the FPR was comparable between OASIS-AD and OASIS and LST, it seems that the false positives for OASIS-AD tend to cluster close to the true positives, whereas for the other two methods they are spread in areas that do not contain WMH. The fuzzy-c mask seems to be slightly conservative, misses important WMH clusters, and falsely identifies some WMH close to the cortical surface.

4. Discussion

We introduced OASIS-AD, a class of models designed to refine OASIS (Sweeney et al., 2013), an MS lesion segmentation approach for WMH in older adults with AD. OASIS-AD performed well in comparison with existing methods. OASIS-AD provides an interpretable solution based on logistic regression combined with two map refinement techniques designed to reduce the false positive rate. OASIS-AD is a significant improvement over OASIS both in terms of modeling techniques, which are adapted for the specific problems raised by WMH segmentation, and in terms of segmentation performance. OASIS-AD has three major advantages that are worth emphasizing. First, the logistic-based approach is highly flexible and it allows the use of any combination of multi-modal inputs, easy expansion of the predictor space, non-linearity, and potential interaction effects. Moreover, traditional methods for quantifying the relative importance of existing or new predictors can provide powerful insights into what and how new modalities and features are actually contributing to improved segmentation. Second, OASIS-AD can be trained with small, moderate, and large sample sizes, making it a very useful first line segmentation approach that can be easily deployed in new environments or sub-disease types. Third, and probably most importantly, OASIS-AD is easy to

Table 2
Performance evaluation metrics (reduced).

	ACC	PPV	TPR	FPR	DSC	ROC	PRC
M1-G	0.97(0.01) (0.96,0.98)	0.85(0.03) (0.83,0.88)	0.70(0.03) (0.69,0.72)	0.009(0.001) (0.008,0.01)	0.78(0.03) (0.77,0.79)	0.97	0.86
OASIS	0.95(0.01) (0.94,0.96)	0.75(0.04) (0.75,0.8)	0.58(0.04) (0.58,0.62)	0.014(0.002) (0.012,0.015)	0.65(0.04) (0.64,0.69)	0.92	0.74
MIMOSA	0.96(0.01) (0.96,0.97)	0.94(0.02) (0.93,0.97)	0.58(0.04) (0.56,0.64)	0.002(0.001) (0.001,0.003)	0.71(0.04) (0.70,0.77)	0.87	0.77
LST	0.97(0.01) (0.96,0.97)	0.83(0.05) (0.83,0.86)	0.72(0.04) (0.71,0.76)	0.012(0.005) (0.010,0.013)	0.76(0.03) (0.75,0.8)	0.87	0.77
fuzzy-c	0.95(0.002) (0.94,0.96)	0.88(0.13) (0.85,0.89)	0.51(0.13) (0.50,0.52)	0.018(0.015) (0.017,0.019)	0.62(0.11) (0.61,0.63)	NA	NA

Data is presented as mean (standard deviation) and 95% CI

generalize and interpret because it is based on a logistic regression model that accounts for the intensity of voxels in various disease tissues across image modalities.

Deep learning approaches can provide an alternative to OASIS-AD and we continue to investigate the added benefit of these techniques, including convolutional neural networks. So far, we have seen encouraging results, though much remains to be done in terms of increasing the sample size of the training data (not easy to achieve in low resource environments), performance (we have not yet matched OASIS-AD), interpretability (we would like to better understand what features of the data are actually contributing to improved prediction performance), and choices of the many tuning parameters (e.g., neighborhood size and filter types).

OASIS-AD uses NNR and GFR in combination to reduce the false positive rate, especially due to speckling. In particular, NNR uses neighborhood information combined with information from the FAST[™] segmentation algorithm to increase or decrease the estimated probability that a voxel should be identified as WMH. A potential disadvantage of NNR and GFR is that in certain situations they may lead to results that are too conservative when probabilities are shrunk too aggressively towards zero. There are several potential solutions that could be considered to help address these problems. For example, in the first step of the NNR(v) algorithm described in Section 1 we used the transformation $P_{wmh}^{rv} = (P_{wmh}^v)^{10}$ for voxels that were estimated by FAST[™]

Table 3
Performance evaluation metrics(full).

	ACC	PPV	TPR	FPR	DSC	ROC	PRC
M1	0.96(0.01) (0.95,0.97)	0.84(0.04) (0.81,0.88)	0.58(0.02) (0.58,0.61)	0.008(0.002) (0.006,0.008)	0.68(0.02) (0.67,0.72)	0.95	0.80
M1-G	0.97(0.01) (0.96,0.98)	0.85(0.03) (0.83,0.88)	0.70(0.03) (0.69,0.72)	0.009(0.001) (0.008,0.01)	0.78(0.03) (0.77,0.79)	0.97	0.86
M1-GN	0.96(0.01) (0.95,0.96)	0.85(0.04) (0.83,0.88)	0.63(0.05) (0.62,0.68)	0.005(0.001) (0.004,0.007)	0.73(0.03) (0.71,0.77)	0.97	0.86
M1-NG	0.96(0.01) (0.95,0.96)	0.84(0.04) (0.82,0.89)	0.58(0.05) (0.57,0.64)	0.007(0.001) (0.006,0.008)	0.68(0.05) (0.66,0.74)	0.82	0.73
M2	0.95(0.01) (0.95,0.95)	0.8(0.06) (0.75,0.86)	0.52(0.03) (0.51,0.56)	0.009(0.003) (0.007,0.011)	0.63(0.04) (0.61,0.68)	0.94	0.76
M2-G	0.96(0.01) (0.96,0.97)	0.84(0.04) (0.80,0.87)	0.65(0.04) (0.64,0.69)	0.009(0.003) (0.008,0.011)	0.72(0.03) (0.70,0.74)	0.96	0.86
M2-GN	0.96(0.003) (0.96,0.97)	0.86(0.05) (0.80,0.88)	0.58(0.062) (0.57,0.65)	0.006(0.003) (0.007,0.01)	0.68(0.05) (0.65,0.74)	0.94	0.81
M2-NG	0.95(0.003) (0.95,0.96)	0.84(0.05) (0.8,0.87)	0.54(0.06) (0.54,0.62)	0.007(0.002) (0.007,0.009)	0.65(0.06) (0.63,0.72)	0.82	0.72
OASIS	0.95(0.01) (0.94,0.96)	0.75(0.04) (0.75,0.8)	0.58(0.04) (0.58,0.62)	0.014(0.002) (0.012,0.015)	0.65(0.04) (0.64,0.69)	0.92	0.74
MIMOSA	0.96(0.01) (0.96,0.97)	0.94(0.02) (0.93,0.97)	0.58(0.04) (0.56,0.64)	0.002(0.001) (0.001,0.003)	0.71(0.04) (0.70,0.77)	0.87	0.77
LST	0.97(0.01) (0.96,0.97)	0.83(0.05) (0.83,0.86)	0.72(0.04) (0.71,0.76)	0.012(0.005) (0.010,0.013)	0.76(0.05) (0.75,0.8)	0.87	0.77
fuzzy-c	0.95(0.002) (0.94,0.96)	0.88(0.13) (0.85,0.89)	0.51(0.13) (0.50,0.52)	0.018(0.015) (0.017,0.019)	0.62(0.11) (0.61,0.63)	NA	NA

Data is presented as mean (standard deviation) and the 95% CI

to have probability 1 of being in white matter and all 6NN to be in white matter. One could use alternative transformations and one could better use FAST[™], or other segmentation algorithms, to inform the likelihood that the voxel is in WMH. One solution could be to use FAST[™] and OASIS-AD iteratively: first use FAST[™] to segment white matter, gray matter, and CSF and then use OASIS-AD to estimate the WMH. Once this is done the WMH region estimated via OASIS-AD can be filled in with normally appearing white matter and the process could be iterated until no differences are observed. In our study we only have two image modalities, T1 and FLAIR, though OASIS-AD can be easily extended to incorporate additional image modalities, while standard variable selection techniques as well as interaction terms could easily be embedded in the model structure.

In conclusion, our improved OASIS-AD method works better than the original OASIS, MIMOSA and LST for the aging population. Since this method was not tested on samples affected by diseases other than aging and, subsequently AD, we can only conclude that OASIS-AD performs better than existing methods in these types of population.

Conflict of Interest

Ciprian Crainiceanu is consulting with Bayer on unrelated methodological problems for wearable devices. Oscar Lopez is consulting with Grifols.

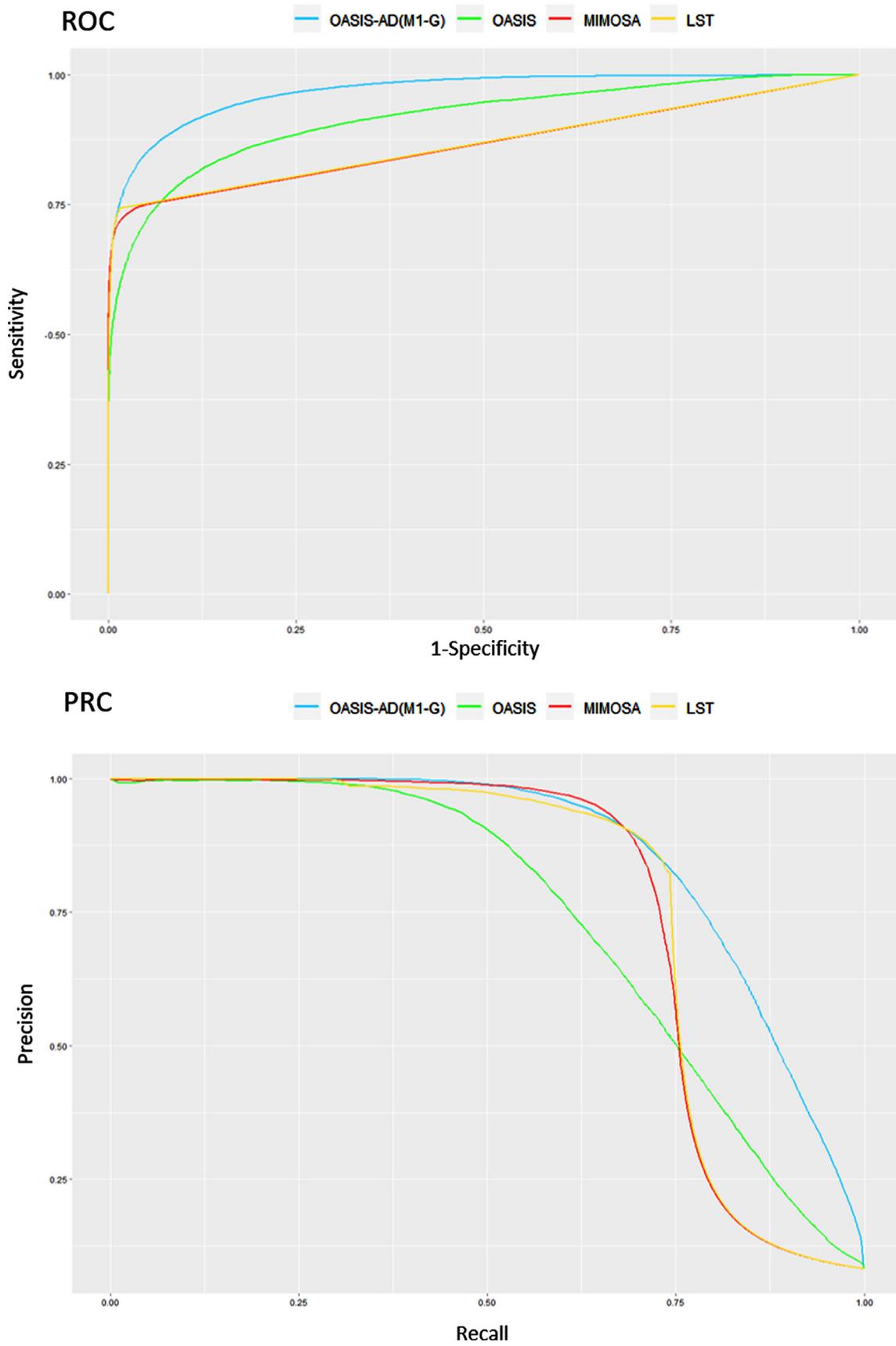


Fig. 2. ROC and PRC of models(reduced).

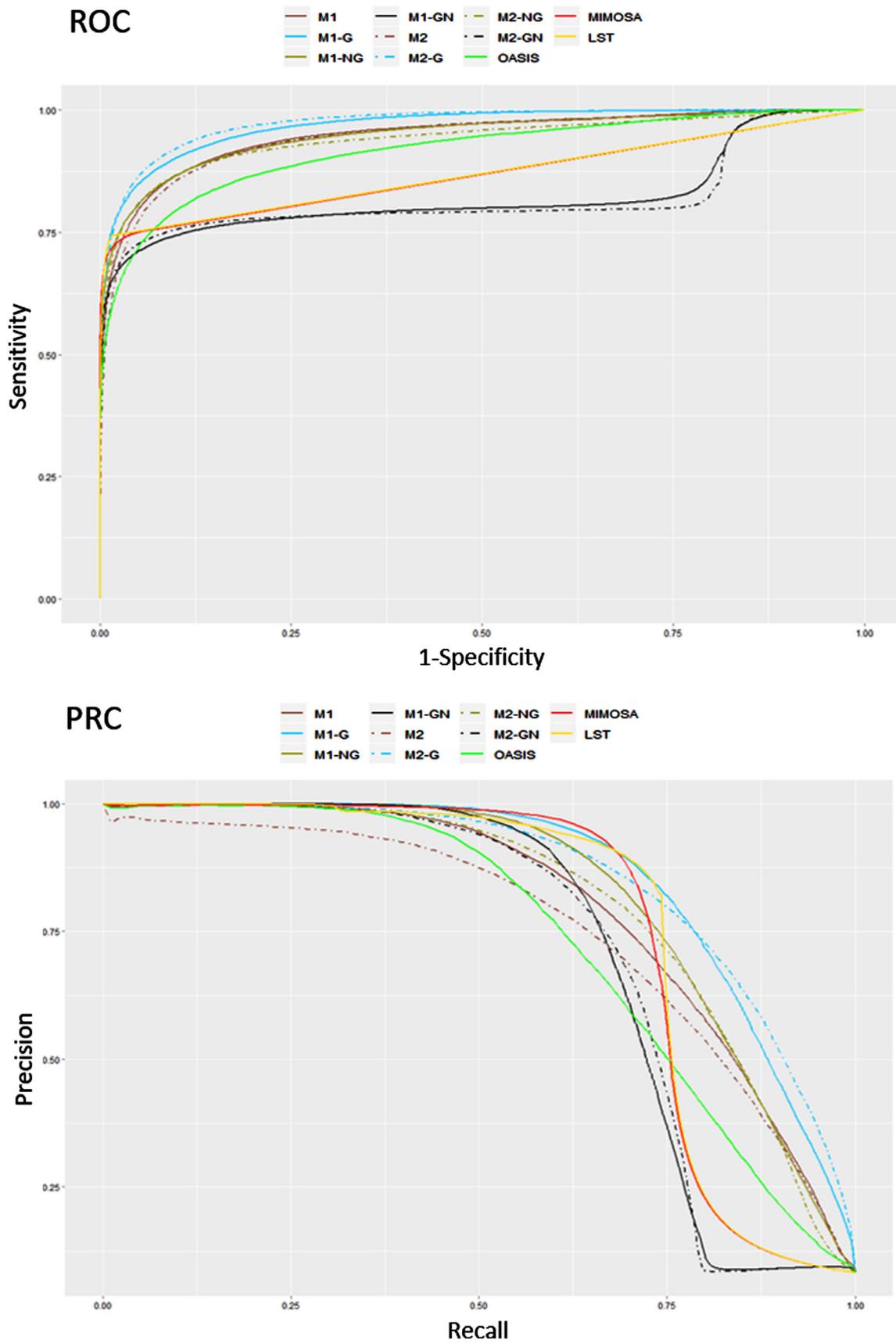


Fig. 3. ROC and PRC of models(full).

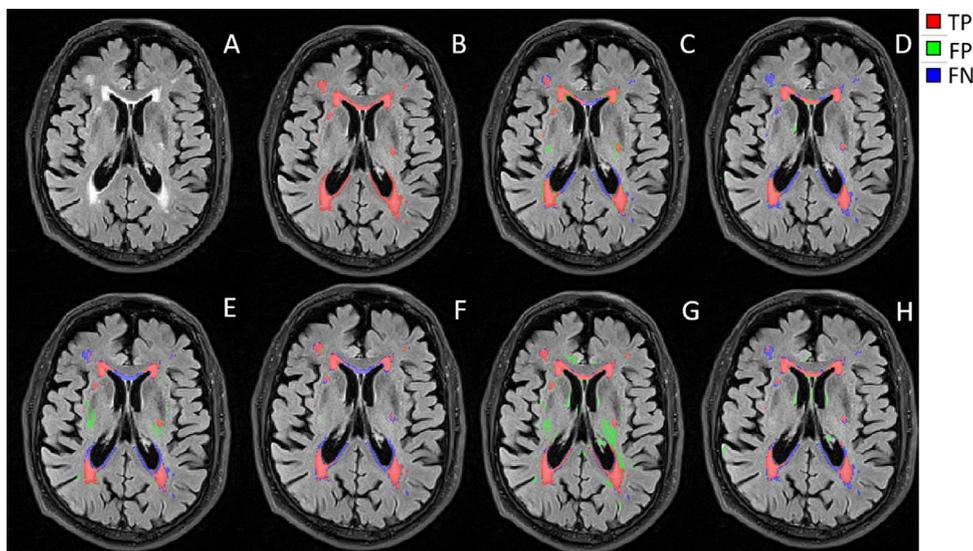


Fig. 4. A: FLAIR slice, B: manual, C: M1-G, D: M1-GN, E: OASIS, F: MIMOSA, G: LST, H: fuzzy-c.

For each input voxel v , we have:

Input: T_v , tissue type for voxel v estimated by "FAST"

T_v^{6NN} , tissue type set for the 6NN of voxel v estimated by "FAST"

p_{wm}^v , probability of voxel v being in white matter estimated by "FAST"

p_{wm}^{6NN} , probability set for the 6NN of voxel v estimated by "FAST"

P_{wmh}^v , probability of voxel v being WMH estimated by logistic models

Output: P_{wmh}^{rv} , probability of voxel being WMH estimated using NNR

- 1: **procedure** NNR(v)
- 2: **if** $p_{wm}^v = 1$ and $T_v^{6NN} = wm$, **then**
- 3: $P_{wmh}^{rv} = (P_{wmh}^v)^{10}$
- 4: **else if** $T_v = wm$ and $T_v^{6NN} \neq wm$, **then**
- 5: $P_{wmh}^{rv} = (P_{wmh}^v)^{\text{average}(p_{wm}^{6NN})}$
- 6: **else**
- 7: $P_{wmh}^{rv} = P_{wmh}^v$
- 8: **return** P_{wmh}^{rv}

Algorithm 1. Nearest Neighbor Refinement (NNR).

Acknowledgement

This work was supported by the following National Institutes of Health/NIA grants: R01 AG063752, P30 AG066468, RF1 AG025516, P01 AG025204, R01 AG034852, R01 GM113243.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.nicl.2019.102151](https://doi.org/10.1016/j.nicl.2019.102151)

References

Admiraal-Behloul, F., Van Den Heuvel, D., Olofsen, H., van Osch, M.J., van der Grond, J., Van Buchem, M., Reiber, J., 2005. Fully automatic segmentation of white matter hyperintensities in mr images of the elderly. *Neuroimage* 28 (3), 607–617.

Anbeek, P., Vincken, K.L., Van Osch, M.J., Bisschops, R.H., Van Der Grond, J., 2004. Probabilistic segmentation of white matter lesions in mr imaging. *NeuroImage* 21 (3), 1037–1044.

Anitha, M., Selvy, P.T., Palanisamy, V., 2012. Wml detection of brain images using fuzzy and possibilistic approach in feature space. *WSEAS Transactions on computers* 22242872. e-ISSN

Avants, B.B., 2019. ANTsR: ANTs in R: Quantification Tools for Biomedical Images. R package version 0.5.4.2

Caligiuri, M.E., Perrotta, P., Augimeri, A., Rocca, F., Quattrone, A., Cherubini, A., 2015. Automatic detection of white matter hyperintensities in healthy aging and pathology using magnetic resonance imaging: a review. *Neuroinformatics* 13 (3), 261–276.

Davis, J., Goadrich, M., 2006. The relationship between precision-recall and roc curves. *Proceedings of the 23rd International Conference on Machine Learning. ACM*, pp. 233–240.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology* 26 (3), 297–302.

Dyrby, T.B., Rostrup, E., Baaré, W.F., van Straaten, E.C., Barkhof, F., Vrenken, H., Ropele, S., Schmidt, R., Erkinjuntti, T., Wahlund, L.-O., et al., 2008. Segmentation of age-related white matter changes in a clinical multi-center study. *Neuroimage* 41 (2), 335–345.

Ghafoorian, M., Karssemeijer, N., Heskes, T., Uden, I.W., Sanchez, C.I., Litjens, G., Leeuw, F.-E., Ginneken, B., Marchiori, E., Platel, B., 2017. Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. *Sci. Rep.* 7 (1), 5110.

Goutte, C., Gaussier, E., 2005. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. *European Conference on Information Retrieval. Springer*, pp. 345–359.

Haralick, R.M., Sternberg, S.R., Zhuang, X., 1987. Image analysis using mathematical morphology. *IEEE Trans. Pattern Anal. Mach.Intell.* (4), 532–550.

Herskovits, E., Bryan, R., Yang, F., 2008. Automated bayesian segmentation of microvascular white-matter lesions in the accord-mind study. *Adv. Med. Sci.* 53 (2), 182–190.

Itti, L., Chang, L., Ernst, T., 2001. Segmentation of progressive multifocal leukoencephalopathy lesions in fluid-attenuated inversion recovery magnetic resonance imaging. *J. Neuroimaging* 11 (4), 412–417.

Karim, H., Tudorascu, D., Cohen, A., Price, J., Lopresti, B., Mathis, C., Klunk, W., Snitz, B.,

- Aizenstein, H., 2019. Relationships between executive control circuit activity, amyloid burden, and education in cognitively healthy older adults. *Am. J. Geriatr. Psychiatry* 12, 1360–1371.
- Kuijf, H., Tax, C., Zaanen, L., Bouvy, W., Bresser, J., Leemans, A., Viergever, M., Biessels, G., Vincken, K., 2016. Employing visual analytics to aid the design of white matter hyperintensity classifiers. *Med. Image Comput. Comput.-Assist. Intervention* 97–105.
- 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. *Acad. Radiol.* 15 (3), 300–313.
- Moeskops, P., Bresser, J., Kuijf, H., Mendrik, A., Biessels, G., Pluim, J., Igum, I., 2018. Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in mri. *Neuroimage* 17, 251–262.
- Muschelli, J., Gherman, A., Fortin, J.P., Avants, B., Whitcer, B., Clayden, J.D., Caffo, B.S., Crainiceanu, C.M., 2018. Neuroconductor: an r platform for medical imaging analysis. *Biostatistics*.
- Muschelli, J., Sweeney, E., Lindquist, M., Crainiceanu, C., 2015. Fslr: connecting the fsl software with r. *R J.* 7 (1), 163.
- Nadkarni, N., D, T., E, C., BE, S., AD, C., Halligan, E., aM C., HJ, A., WE, K., 2019. Association between amyloid-, small-vessel disease, and neurodegeneration biomarker positivity, and progression to mild cognitive impairment in cognitively normal individuals. *J. Gerontol. A Biol. Sci. Med. Sci.* 11, 1753–1760.
- Pillai, J.J., Friedman, L., Stuve, T.A., Trinidad, S., Jesberger, J.A., Lewin, J.S., Findling, R.L., Swales, T.P., Schulz, S.C., 2002. Increased presence of white matter hyperintensities in adolescent patients with bipolar disorder. *Psychiatry Res.* 114 (1), 51–56.
- Raidou, R., Kuijf, H., Sepasian, N., Pezzotti, N., WH, B., M, B., Vilanova, A., 2014. The added value of diffusion tensor imaging for automated white matter hyperintensity segmentation. *Computational Diffusion MRI* 45–53. *Mathematics and Visualization*
- Schmidt, P., 2017. Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging. *Imu Ph.D. thesis*.
- Sheline, Y.I., Price, J.L., Vaishnavi, S.N., Mintun, M.A., Barch, D.M., Epstein, A.A., Wilkins, C.H., Snyder, A.Z., Couture, L., Schechtman, K., et al., 2008. Regional white matter hyperintensity burden in automated segmentation distinguishes late-life depressed subjects from comparison subjects matched for vascular risk factors. *Am. J. Psychiatry* 165 (4), 524–532.
- Shinohara, R.T., Goldsmith, J., Mateen, F., Crainiceanu, C., Reich, D., 2012. Predicting breakdown of the blood-brain barrier in multiple sclerosis without contrast agents. *Am. J. Neuroradiol.* 33 (8), 1586–1590.
- Simões, R., Mönninghoff, C., Dlugaj, M., Weimar, C., Wanke, I., van Walsum, A.-M., Slump, C., 2013. Automatic segmentation of cerebral white matter hyperintensities using only 3d flair images. *Magnet. Resonance Imaging* 31 (7), 1182–1189.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Human Brain Mapping* 17 (3), 143–155.
- Sweeney, E.M., Shinohara, R.T., Shiee, N., Mateen, F.J., Chudgar, A.A., Cuzzocreo, J.L., Calabresi, P.A., Pham, D.L., Reich, D.S., Crainiceanu, C.M., 2013. Oasis is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in mri. *NeuroImage* 2, 402–413.
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4itk: Improved n3 bias correction. *IEEE Trans. Med. Imaging* 29 (6), 1310–1320.
- Valcarcel, A.M., Linn, K.A., Vandekar, S.N., Satterthwaite, T.D., Muschelli, J., Calabresi, P.A., Pham, D.L., Martin, M.L., Shinohara, R.T., 2018. Mimoso: an automated method for intermodal segmentation analysis of multiple sclerosis brain lesions. *J. Neuroimaging* 28 (4), 389–398.
- Van Den Heuvel, D., Admiraal-Behloul, F., Ten Dam, V., Olofsen, H., Bollen, E., Murray, H., Blauw, G., Westendorp, R., De Craen, A., Van Buchem, M., et al., 2004. Different progression rates for deep white matter hyperintensities in elderly men and women. *Neurology* 63 (9), 1699–1701.
- Wong, T.Y., Klein, R., Sharrett, A.R., Couper, D.J., Klein, B.E., Liao, D.-P., Hubbard, L.D., Mosley, T.H., investigators, A., et al., 2002. Cerebral white matter lesions, retinopathy, and incident clinical stroke. *Jama* 288 (1), 67–74.
- Wu, M., Rosano, C., Butters, M., Whyte, E., Nable, M., Crooks, R., Meltzer, C.C., Reynolds, C.F., Aizenstein, H.J., 2006. A fully automated method for quantifying and localizing white matter hyperintensities on mr images. *Psychiatry Res.* 148 (2), 133–142.
- Yoo, B.I., Lee, J.J., Han, J.W., Lee, E.Y., MacFall, J.R., Payne, M.E., Kim, T.H., Kim, J.H., Kim, K.W., et al., 2014. Application of variable threshold intensity to segmentation for white matter hyperintensities in fluid attenuated inversion recovery magnetic resonance images. *Neuroradiology* 56 (4), 265–281.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20 (1), 45–57.