NeuroImage 157 (2017) 233–249

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Performance comparison of 10 different classification techniques in segmenting white matter hyperintensities in aging

Mahsa Dadar[a], Josefina Maranzano[b], Karen Misquitta[d], Cassandra J. Anor[d], Vladimir S. Fonov[a], M. Carmela Tartaglia[d], Owen T. Carmichael[c], Charles Decarli[e], D. Louis Collins[a,*], Alzheimer's Disease Neuroimaging Initiative[1]

[a] NeuroImaging and Surgical Tools Laboratory, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada
[b] Magnetic Resonance Studies Laboratory, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada
[c] Pennington Biomedical Research Center, Baton Rouge, LA, USA
[d] Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada
[e] University of California, Davis, CA, USA

ARTICLE INFO

Keywords:
White matter hyperintensities
Segmentation
Classification
Alzheimer's disease

ABSTRACT

Introduction: White matter hyperintensities (WMHs) are areas of abnormal signal on magnetic resonance images (MRIs) that characterize various types of histopathological lesions. The load and location of WMHs are important clinical measures that may indicate the presence of small vessel disease in aging and Alzheimer's disease (AD) patients. Manually segmenting WMHs is time consuming and prone to inter-rater and intra-rater variabilities. Automated tools that can accurately and robustly detect these lesions can be used to measure the vascular burden in individuals with AD or the elderly population in general. Many WMH segmentation techniques use a classifier in combination with a set of intensity and location features to segment WMHs, however, the optimal choice of classifier is unknown.

Methods: We compare 10 different linear and nonlinear classification techniques to identify WMHs from MRI data. Each classifier is trained and optimized based on a set of features obtained from co-registered MR images containing spatial location and intensity information. We further assess the performance of the classifiers using different combinations of MRI contrast information. The performances of the different classifiers were compared on three heterogeneous multi-site datasets, including images acquired with different scanners and different scan-parameters. These included data from the ADC study from University of California Davis, the NACC database and the ADNI study. The classifiers (naïve Bayes, logistic regression, decision trees, random forests, support vector machines, k-nearest neighbors, bagging, and boosting) were evaluated using a variety of voxel-wise and volumetric similarity measures such as Dice Kappa similarity index (SI), Intra-Class Correlation (ICC), and sensitivity as well as computational burden and processing times. These investigations enable meaningful comparisons between the performances of different classifiers to determine the most suitable classifiers for segmentation of WMHs. In the spirit of open-source science, we also make available a fully automated tool for segmentation of WMHs with pre-trained classifiers for all these techniques.

Results: Random Forests yielded the best performance among all classifiers with mean Dice Kappa (SI) of 0.66 ± 0.17 and ICC=0.99 for the ADC dataset (using T1w, T2w, PD, and FLAIR scans), SI=0.72 ± 0.10, ICC=0.93 for the NACC dataset (using T1w and FLAIR scans), SI=0.66 ± 0.23, ICC=0.94 for ADNI1 dataset (using T1w, T2w, and PD scans) and SI=0.72 ± 0.19, ICC=0.96 for ADNI2/GO dataset (using T1w and FLAIR scans). Not using the T2w/PD information did not change the performance of the Random Forest classifier (SI=0.66 ± 0.17, ICC=0.99). However, not using FLAIR information in the ADC dataset significantly decreased the Dice Kappa, but the volumetric correlation did not drastically change (SI=0.47 ± 0.21, ICC=0.95).

* Correspondence to: Magnetic Resonance Imaging (MRI), Montreal Neurological Institute, 3801 University Street, Room WB315, Montréal, QC, Canada H3A 2B4.
E-mail addresses: mahsa.dadar@mail.mcgill.ca (M. Dadar), jmaranzano@mail.mcgill.ca (J. Maranzano), karen.misquitta@mail.utoronto.ca (K. Misquitta), c.anor@mail.utoronto.ca (C.J. Anor), vladimir.fonov@mcgill.ca (V.S. Fonov), carmela.tartaglia@utoronto.ca (M.C. Tartaglia), owen.carmichael@pbrc.edu (O.T. Carmichael), charles.decarli@ucdmc.ucdavis.edu (C. Decarli), louis.collins@mcgill.ca (D.L. Collins).

[1] Part of the data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:
http://adni.loni.usc.edu/wp-ontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

http://dx.doi.org/10.1016/j.neuroimage.2017.06.009
Received 27 April 2017; Received in revised form 30 May 2017; Accepted 2 June 2017
Available online 03 July 2017

Conclusion: Our investigations showed that with appropriate features, most off-the-shelf classifiers are able to accurately detect WMHs in presence of FLAIR scan information, while Random Forests had the best performance across all datasets. However, we observed that the performances of most linear classifiers and some nonlinear classifiers drastically decline in absence of FLAIR information, with Random Forest still retaining the best performance.

Introduction

White matter hyperintensities (WMHs), commonly identified as areas of increased signal in relation with the surrounding white matter regions on T2w, PD and FLAIR MRIs, are one of the non-specific yet typical and constant MRI expressions of cerebral small vessel disease (CSVD), along with lacunar infarcts and microhemorrhages (Conklin et al., 2014; Gouw et al., 2010). They have been shown to be more extensive in patients with Alzheimer's disease compared to age-matched healthy normal populations (Yoshita et al., 2005). WMHs reflect ischemic injury in the elderly and AD populations and the existence and severity of WMHs can lead to or accelerate decline in cognitive as well as executive functions (Dubois et al., 2014). As a result, the location and load of WMHs are important clinical measures, raising substantial need for their accurate quantifications. WMHs are generally detected using fluid attenuated inversion recovery (FLAIR) or T2w/PD scans. Manually labeling WMHs is challenging due to time constraints as well as inter-rater and intra-rater variabilities (Grimaud et al., 1996). As a result, automated tools that can segment WMHs robustly and with high accuracy are extremely useful, particularly in large scale studies such the Alzheimer's Disease Neuroimaging Initiative (http://www.loni.ucla.edu/ADNI/), the National Alzheimer's Coordinating Center (NACC) database (https://www.alz.washington.edu/) and others where it is desired to estimate the contribution of neurovascular disease to cognitive decline.

The heterogeneity in the distribution and patterns of WMHs makes the segmentation task intrinsically complex (Caligiuri et al., 2015). Automated segmentation tools usually integrate information from multiple complementary MRI contrasts including T1w, T2w, PD and FLAIR to reduce uncertainty and improve segmentation accuracy. Most successful fully automated WMH segmentation techniques extract a combination of location and intensity features from these images and use them as inputs to a linear or nonlinear classifier. Here we review the most commonly used linear and nonlinear classifiers in general as well as their application to the task of segmenting lesions in general or WMHs of vascular etiology specifically.

While there have been many studies attempting to segment WMHs using these classification techniques, drawing meaningful comparisons between their performances is not possible since they have been applied to different datasets and results are highly variable across different populations and imaging protocols (García-Lorenzo et al., 2013; Caligiuri et al., 2015). To our knowledge, no studies have compared the performance of these classification techniques for detecting WMHs against one another on the same datasets, especially for cases where classification is attempted without using the optimal FLAIR information. In this paper, we have extensively compared the performance of these different classification techniques in detecting WMHs with and without FLAIR information using 3 different large publicly available datasets with different scanners and acquisition protocols. This enables us to draw more generalizable conclusions regarding the performance of the classifiers. Our contributions include an extensive comparison of 10 widely used classification techniques in detecting WMHs across 4 different datasets, three of which are from multi-site and multi-scanner studies and across different combinations of imaging modalities. In addition, we make publicly available an implementation of the segmentation tool along with all the pre-trained classifiers (http://nist.mni.mcgill.ca/?p=221). The proposed tool is generalizable to data from different scanners since it has been trained on data from multiple scanners.

Materials and methods

Subjects

The performances of the different classifiers were assessed based on four datasets of subjects with different ranges of WMH loads. Table 1 shows the demographic information for each dataset.

ADC

This dataset consists of 70 individuals (70–90 years old) with normal cognition, mild cognitive impairment (MCI), and AD dementia from University of California, Davis Alzheimer's Disease Center (ADC) who were scanned using T1w, double-echo T2w/PD, and FLAIR MRI modalities.

NACC

This dataset consists of a patient sample of 32 MCI and AD subjects obtained from the National Alzheimer's Coordinating Center (NACC) database which is a database of subjects with a range of cognitive status, i.e. normal cognition, MCI, and demented who received T1w, and FLAIR MRI scans (https://www.alz.washington.edu/). Data consisted of variables from a Uniform Data Set collected from more than 30 Alzheimer's disease centers (ADC) throughout the United States and cataloged at the NACC. ADCs are National Institute on Aging–funded centers that enroll patients using different participation recruiting practices. A full description of the NACC data set has been previously provided (Beekly et al., 2004; Morris et al., 2006). NACC data used here has been acquired at six different ADCs using eight different scanner models of three different manufacturers. Subjects were selected to have low, medium, and large WMH loads.

ADNI

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

ADNI1

This dataset consists of T1w, T2w, and PD scans of 53 subjects from ADNI1 study. Despite the fact that all subjects had to have Hachinski Ischemic Score of less than or equal to 4 as part of the inclusion criteria (Petersen et al., 2010), we found many subjects that had high WMH loads. Subjects were selected from different sites and scanners and a preliminary assessment was performed to evaluate their WMH load

Table 1
Demographic information for ADC, NACC, ADNI1 and ADNI2/GO datasets.

Dataset	ADC	NACC	ADNI1	ADNI2/GO
N	70	32	53	46
Sex	35 M	15 M	27 M	25 M
Age	78.0 ± 7.3	74.9 ± 8.0	75.7 ± 6.6	74.1 ± 6.5