The added value of diffusion tensor imaging for automated white matter hyperintensity segmentation

Hugo J. Kuijf, Chantal M. W. Tax, L. Karlijn Zaanen, Willem H. Bouvy, Jeroen de Bresser, Alexander Leemans, Max A. Viergever, Geert Jan Biessels, and Koen L. Vincken

Abstract Automated white matter hyperintensity (WMH) segmentation techniques for brain MRI often employ voxel-wise classifiers, trained on traditional features such as: multi-spectral MR image intensities, spatial location, texture, or shape. Recent studies show that diffusion tensor imaging (DTI) provides a measure for WMH, independent from the commonly used FLAIR images. Hence, we hypothesized that adding features derived from DTI to a voxel-wise classifier for WMH segmentation may have added value and improve segmentation results.
A k nearest neighbour (kNN) classifier was implemented and trained on various combinations of features. Manual delineations of WMH were available for 20 subjects. Classifiers trained with diffusion features, such as fractional anisotropy and mean diffusivity, are compared to an equivalent classifier without diffusion features. Evaluation measures are sensitivity and Dice similarity coefficient (SI).
Adding diffusion features to a kNN classifier significantly (Student's t-test, $p < 0.0001$) improved the quality of the segmentation. Depending on the chosen kNN parameters and features, improvements in sensitivity ranged from 2.4% to 13.5% and in SI from 4.7% to 18.0%.
In conclusion, adding diffusion features derived from DTI to a voxel-wise classifier for WMH segmentation significantly improves the quality of the segmentation.

1 Introduction

Vascular lesions in the brain, like white matter hyperintensities (WMH), are commonly encountered on MR images of elderly people and are associated with cog-

University Medical Center Utrecht, the Netherlands, e-mail: hugok@isi.uu.nl

The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-11182-7_5

nitive impairment, vascular risk factors, and an increased risk for future stroke, dementia, and death [1, 2, 3].

WMH are usually visualized on MRI with a FLAIR sequence, on which they appear as hyperintense regions in the white matter. The assessment of WMH can be performed manually, with visual rating scales such as the Age-Related White Matter Changes rating scale [3]. Manual delineations of WMH can be used to quantitatively assess volume, shape, or location. However, this task is tedious, time-consuming, and observer dependent, making it infeasible for use in large research studies or daily clinical practice.

Automated segmentation techniques for WMH have been proposed in the literature for many years [4]. Many of these techniques employ a voxel-wise classification (e.g. k nearest neighbours (kNN), neural networks, support vector machines, random forest), using features such as intensity, spatial location, texture, or shape [5, 6, 7, 8]. Such features are derived from the traditional MR sequences that are available, including T1-weighted, T2-weighted, PD-weighted, FLAIR, or IR sequences. Often, post processing routines are applied to improve the resulting segmentation.

Diffusion tensor imaging (DTI) can be used to study the microstructural organization of white matter by sensitizing the MR sequence to diffusion [9]. Measures derived from DTI, such as fractional anisotropy (FA) and mean diffusivity (MD), are known to detect changes in the white matter microstructure [10]. A recent study by Maillard *et al.* [1] showed that FLAIR and DTI signals are independent predictors of white matter hyperintensities [1, 2].

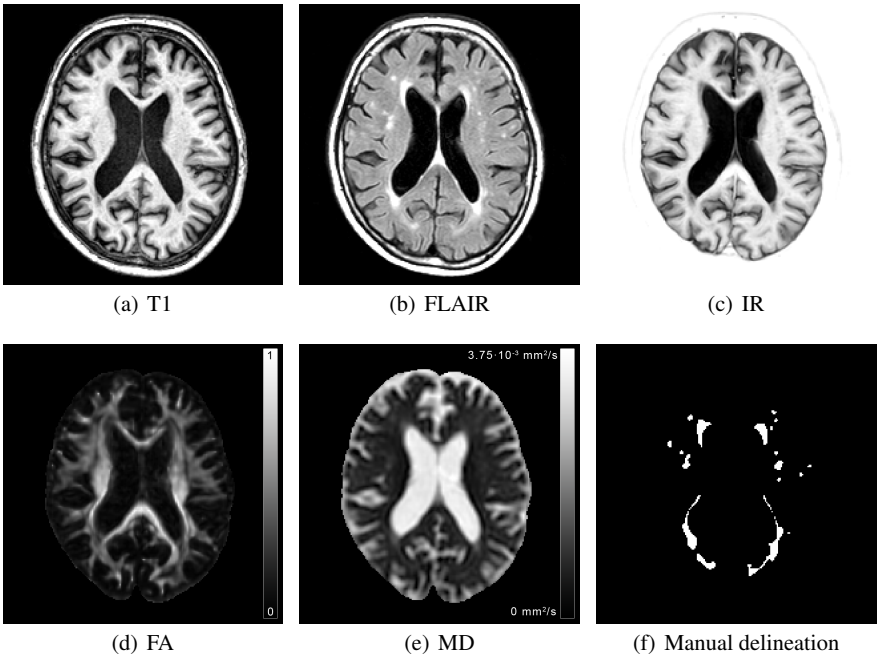
Based on these findings, we hypothesize that it can be beneficial to add DTI measures to a voxel-wise classification technique for WMH segmentation. In this study, we will use a popular classifier for WMH segmentation, the kNN classifier [4, 5, 7], and investigate the added value of DTI measures.

2 Methods and Materials

2.1 Participants & MRI

For the present study, 20 subjects (mean age: 71 years, sd: 4 years, 10 men, 10 women) were recruited at the University Medical Center Utrecht, the Netherlands. Subjects included patients with diabetes and matched controls, all with varying degrees of atrophy and WMH. The study was approved by the medical ethics committee of the University Medical Center Utrecht. All subjects gave written informed consent.

The subjects underwent a standardized MR exam on a 3.0 T Philips Achieva MR scanner using an eight-channel head coil, including, amongst others, a 3D T1-weighted turbo field echo sequence (TR: 7.9 ms, TE: 4.5 ms), a multi-slice FLAIR sequence (TR: 11 000 ms, TE: 125 ms, TI: 2800 ms), a multi-slice IR sequence (TR:



(a) T1 (b) FLAIR (c) IR

(d) FA (e) MD (f) Manual delineation

Fig. 1 Example image data of a subject with a high white matter hyperintensity (WMH) load (WMH volume: 34 ml). The top row shows the intensity features, where WMH appear dark on T1 and IR, and bright on FLAIR. Bottom row shows diffusion features, where WMH appear dark on FA and bright on MD, with the manual delineation of the WMH.

4416 ms, TE: 15 ms, TI: 400 ms), and a single-shot EPI DTI sequence with 45 directions at b = 1200 s mm^{-2} and a b = 0 s mm^{-2} image (TR: 6638 ms, TE: 73 ms).

Using the `elastix` toolbox for medical image registration [11], the 3D T1, IR, and b = 0 s mm^{-2} DTI images were aligned with the FLAIR image and resampled to a voxel size of $0.96 \times 0.95 \times 3.00$ mm^3. This matched the voxel size of the FLAIR and established a voxel-wise correspondence between all the images. The diffusion weighted scans were corrected for subject motion, eddy current induced geometric distortions [12], and EPI distortions [13], including the required B-matrix adjustments [14]. These corrections were performed in one interpolation step to minimize blurring effects, using ExploreDTI [15]. A binary mask including all intracranial structures (the brain and cerebrospinal fluid) was available for all subjects.

WMH were delineated manually by an experienced human observer on the FLAIR image using a freehand spline drawing tool. All delineations were verified by a second experienced human observer and adapted if needed.

2.2 Classification

A *k*NN classifier, as implemented in scikit-learn [16], was used to perform a voxel-wise classification. Here, only two classes were considered: a voxel is either WMH or non-WMH (background, brain tissue, et cetera). The *k*NN classifier is trained with labelled samples, where each sample is an individual voxel for which the true class is known from the manual delineations. Each sample has certain features, which are explained below.

To perform WMH segmentation for a new subject, each voxel in the image data of that subject is evaluated by the *k*NN classifier. It will find k training samples having features that are the most similar to the new voxel (the nearest neighbours). The most frequent class amongst the nearest neighbours is assigned to the new voxel.

A leave-one-out cross-validation strategy was used to classify the data of each subject. The *k*NN classifier was trained on the data of 19 subjects and then applied to segment the data of 1 subject. This procedure was repeated 20 times, so that the data of each individual subject was classified using the data of the other 19 subjects. The quality of the resulting segmentation could be assessed by comparing the results of the *k*NN classifier against the manual delineations.

Included in the training data were all voxels labelled as WMH and 10% of the non-WMH voxels within the binary mask. A k-d tree with a leaf size of 30 was used to perform the nearest neighbour search, with an Euclidean distance metric. To determine the robustness of the technique, various settings for the number of neighbours k and the neighbour weighting function w were used, as explained below in the Experiments section.

2.3 Features

A number of features was computed for each voxel in the image data of a subject, including: intensity, spatial, and diffusion features. All features were scaled to have values between 0 and 1, giving them equal weights within the *k*NN classification. Intensity and diffusion features for an example subject are shown in Figure 1, together with the manual delineation of the WMH.

2.3.1 Intensity features

The T1, FLAIR, and IR intensities were included as features. For each image, the intensities within the binary mask were normalized using the method proposed by Cocosco *et al.* [17], clipping the first and last percentile of the histogram and rescaling all intensities in between.

2.3.2 Spatial features

As spatial features, the MNI-normalized spatial coordinates x, y, and z of each voxel were included. The T1 sequence of each subject was transformed to the MNI152 standard-space atlas by computing a linear registration using `elastix` [11, 18, 19]. By applying the inverse transformation, the MNI152 coordinates were "warped" to the coordinate space of the subject data and no interpolation or resampling of the original data was required. This approach is nicely demonstrated by Steenwijk *et al.* [7]. All values for x, y, and z within the binary mask were normalized.

By warping the MNI152 atlas, the spatial location of each voxel was comparable between all subjects and differences in, e.g., position in the scanner or head size were normalized.

2.3.3 Diffusion features

Diffusion tensors were estimated using a weighted linear least squares approach [20]. The diffusion measures FA, MD, axial diffusivity (AD), radial diffusivity (RD) [21], and the Westin measures C_L, C_P, and C_S [22] were computed and subsequently normalized.

2.4 Experiments

Multiple settings for the kNN parameters k and w were applied. k is the number of neighbours used by the classifier and was set to 50, 75, or 100. w is the weighting function that was applied to the nearest neighbours to select a final class and was set to either 'uniform' or 'distance'. With uniform weighting, the class that occurs most amongst the nearest neighbours is assigned to the voxel being tested. With distance weighting, the inverse of the distance to each training sample is used as a weight: closer neighbours will have a greater influence in determining the final class than neighbours further away.

Various combinations of features were used, where combinations consisting of 'traditional' features will be compared to combinations with diffusion features. Selected combinations $f_i \in F$ included:

f_1: T1, IR, FLAIR
f_2: T1, IR, FLAIR, x, y, and z
f_3: T1, IR, FLAIR, with FA and MD
f_4: T1, IR, FLAIR, x, y, z, with FA and MD
f_5: T1, IR, FLAIR, x, y, z, with FA, MD, C_L, C_P, C_S, AD, and RD

In total, 6 different kNN classifiers will be evaluated for each $f_i \in F$. Each classifier employed a leave-one-out cross-validation approach. For each resulting segmentation, the sensitivity (percentage of true WMH voxels detected) and the commonly

Table 1 Left: sensitivity and Dice similarity coefficient (SI) (higher is better) for the classifiers trained on combinations of features $f_i \in F$. **Right:** relative improvement between classifiers with diffusion features versus the equivalent classifier without diffusion features.

F	Sensitivity (%)	SI
f_1	59.7 ± 0.2	0.349 ± 0.001
f_2	73.4 ± 0.4	0.536 ± 0.005
f_3	67.8 ± 0.3	0.411 ± 0.003
f_4	77.2 ± 0.4	0.565 ± 0.004
f_5	75.2 ± 0.6	0.561 ± 0.003

	Relative improvement	
	Sensitivity (%)	Dice (%)
f_3 vs f_1	$13.5 \pm 0.2^*$	$18.0 \pm 0.6^*$
f_4 vs f_2	$5.1 \pm 0.1^*$	$5.4 \pm 0.2^*$
f_5 vs f_2	$2.4 \pm 0.3^*$	$4.7 \pm 0.5^*$

* The difference is statistically significant, using a paired samples two-tailed Student's t-test with $p < 0.0001$.

reported Dice similarity coefficient (SI, overlap between manual delineation and resulting segmentation) [23] will be assessed. Differences between the classifiers will be tested for significance using a paired samples two-tailed Student's t-test.

3 Results

For each classifier trained on a combination of features $f_i \in F$, the reported results are the mean $\pm$ sd obtained by averaging the results for each setting of k and w. These results are reported in Table 1, together with the difference between classifiers using diffusion features versus the equivalent classifier without diffusion features.

Adding DTI measures as features to a kNN classifier results in statistically significant ($p < 0.0001$) improvements in both sensitivity and SI.

Including more diffusion features besides FA and MD, as is done with f_5, diminished the measured improvement, although the results are still significantly better than without these features.

The inclusion of spatial features (f_2 vs f_1 and f_4 vs f_3) resulted in a statistically significant ($p < 0.0001$) improvement in sensitivity and SI.

In cases where the sensitivity is high or approaches the inter observer variability, it is informative to inspect the false negative rate / type II error (defined as $1-$ sensitivity; the percentage of undetected WMH voxels). The relative reductions in false negative rate are: f_3 vs f_1: $-20.0 \pm 0.4\,\%$, f_4 vs f_2: $-14.2 \pm 0.2\,\%$, and f_5 vs f_2: $-6.5 \pm 1.0\,\%$. All reductions are statistically significant ($p < 0.0001$).

4 Discussion

It is clear that adding diffusion features to a voxel-wise classifier for WMH segmentation significantly improves the segmentation results. This is in line with the

reports from literature that FLAIR and DTI are independent predictors and confirms our hypothesis that DTI has added value in automated WMH segmentation.

The *k*NN classifiers as presented herein do not have a high performance in terms of SI. However, the current classifiers are only used to demonstrate the added value of DTI measures. More advanced segmentation methods that are proposed in literature are also likely to benefit from diffusion features. Post processing steps that reduce the number of false positive detections might increase the SI to a level comparable with literature, since SI is a measure that combines both sensitivity and specificity. However, adding these post processing routines is not relevant to demonstrate the added value of DTI.

Adding FA and MD features to the classifiers proved to be valuable. The addition of more DTI measures, such as C_L, C_P, C_S, AD, and RD, still improved the results with respect to an equivalent classifier without any DTI measures, but showed a diminished improvement with respect to the classifiers that only included FA and MD. A possible explanation for this can be found in the high correlation that exists between the DTI measures, since they are all derived from the eigenvalues of the diffusion tensor. Adding a feature to the classifier increases the dimensionality of the feature space and slightly reduces the influence of all other features. When a feature is added that (strongly) correlates with an existing feature, the influence of all other features is reduced. Therefore, only features that provide new information should be added and in the present situation, this is provided by FA and MD.

Future work will consist of improving the overall performance of the classifiers, by including more features and post processing routines that are known from literature to improve the segmentation quality. Possible improvements are the inclusion of tissue type priors [7], morphological operations to close segmentation gaps and remove too small hyperintensities, or optimize the probability threshold of the *k*NN classifier [5]. The latter option has the potential to further increase the sensitivity (by lowering the threshold), but will introduce more false positives that need to be removed by post processing. Increased involuntary head motion or altered cardiac pulsatile motion in these elderly patients may cause acquisition artefacts. This can result in large biases in the diffusion measures, which can be prevented by using robust tensor estimation procedures in future work [24].

5 Conclusion

Including DTI measures as features in a classification technique for WMH segmentation significantly improves the results of the final segmentation.

Acknowledgements We gratefully acknowledge the Utrecht Vascular Cognitive Impairment Study Group for recruiting the patients who were included as subjects in our study. This study was financially supported by the project Brainbox (Quantitative analysis of MR brain images for cerebrovascular disease management), funded by the Netherlands Organisation for Health Research and Development (ZonMw) in the framework of the research programme IMDI (Innovative Medical Devices Initiative); project 104002002.

References

1. Maillard, P., Carmichael, O., Harvey, D., Fletcher, E., Reed, B., Mungas, D., DeCarli, C.: Flair and diffusion mri signals are independent predictors of white matter hyperintensities. American Journal of Neuroradiology **34**(1) (2013) 54–61
2. de Groot, M., Verhaaren, B.F., de Boer, R., Klein, S., Hofman, A., van der Lugt, A., Ikram, M.A., Niessen, W.J., Vernooij, M.W.: Changes in normal-appearing white matter precede development of white matter lesions. Stroke **44**(4) (2013) 1037–1042
3. Wahlund, L.O., Barkhof, F., Fazekas, F., Bronge, L., Augustin, M., Sjgren, M., Wallin, A., Ader, H., Leys, D., Pantoni, L., Pasquier, F., Erkinjuntti, T., Scheltens, P., on behalf of the European Task Force on Age-Related White Matter Changes: A new rating scale for age-related white matter changes applicable to mri and ct. Stroke **32**(6) (2001) 1318–1322
4. Mortazavi, D., Kouzani, A., Soltanian-Zadeh, H.: Segmentation of multiple sclerosis lesions in mr images: a review. Neuroradiology **54**(4) (2012) 299–320
5. Anbeek, P., Vincken, K.L., van Osch, M.J., Bisschops, R.H., van der Grond, J.: Probabilistic segmentation of white matter lesions in MR imaging. NeuroImage **21**(3) (2004) 1037 – 1044
6. de Boer, R., Vrooman, H.A., van der Lijn, F., Vernooij, M.W., Ikram, M.A., van der Lugt, A., Breteler, M.M., Niessen, W.J.: White matter lesion extension to automatic brain tissue segmentation on mri. NeuroImage **45**(4) (2009) 1151 – 1161
7. Steenwijk, M.D., Pouwels, P.J., Daams, M., van Dalen, J.W., Caan, M.W., Richard, E., Barkhof, F., Vrenken, H.: Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (knn-ttps). NeuroImage: Clinical **3**(0) (2013) 462 – 469
8. Admiraal-Behloul, F., van den Heuvel, D., Olofsen, H., van Osch, M., van der Grond, J., van Buchem, M., Reiber, J.: Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage **28**(3) (2005) 607 – 617
9. Basser, P.J., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the nmr spin echo. Journal of magnetic resonance. Series B **103**(3) (1994) 247–254
10. Beaulieu, C.: The basis of anisotropic water diffusion in the nervous system a technical review. NMR in Biomedicine **15**(7-8) (2002) 435–455
11. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: elastix: A toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1) (jan. 2010) 196 –205
12. Rohde, G., Barnett, A., Basser, P., Marenco, S., Pierpaoli, C.: Comprehensive approach for correction of motion and distortion in diffusion-weighted mri. Magnetic Resonance in Medicine **51**(1) (2004) 103–114
13. Irfanoglu, M.O., Walker, L., Sarlls, J., Marenco, S., Pierpaoli, C.: Effects of image distortions originating from susceptibility variations and concomitant fields on diffusion mri tractography results. NeuroImage **61**(1) (2012) 275 – 288
14. Leemans, A., Jones, D.K.: The b-matrix must be rotated when correcting for subject motion in dti data. Magnetic Resonance in Medicine **61**(6) (2009) 1336–1349
15. Leemans, A., Jeurissen, B., Sijbers, J., Jones, D.K.: Exploredti: a graphical toolbox for processing, analyzing, and visualizing diffusion mr data. In: Proc. Intl. Soc. Mag. Reson. Med. Volume 17. (2009) 3537
16. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
17. Cocosco, C.A., Zijdenbos, A.P., Evans, A.: A fully automatic and robust brain mri tissue classification method. Medical Image Analysis **7**(4) (2003) 513 – 527 Medical Image Computing and Computer Assisted Intervention.
18. Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., Collins, D.L.: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage **47, Supplement 1** (2009) S102 Organization for Human Brain Mapping 2009 Annual Meeting.

19. Fonov, V.S., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L.: Unbiased average age-appropriate atlases for pediatric studies. NeuroImage **54**(1) (2011) 313 – 327
20. Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., Jeurissen, B.: Weighted linear least squares estimation of diffusion mri parameters: Strengths, limitations, and pitfalls. NeuroImage **81**(0) (2013) 335 – 346
21. Basser, P.J.: Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. NMR Biomed **8** (1995) 333–344
22. Westin, C., Maier, S., Mamata, H., Nabavi, A., Jolesz, F., Kikinis, R.: Processing and visualization for diffusion tensor mri. Med Image Anal **6**(2) (2002) 93–108
23. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3) (1945) 297–302
24. Tax, C.M., Otte, W.M., Viergever, M.A., Dijkhuizen, R.M., Leemans, A.: Rekindle: Robust extraction of kurtosis indices with linear estimation. Magnetic Resonance in Medicine **(in press)** (2014)