Employing Visual Analytics to Aid the Design of White Matter Hyperintensity Classifiers

Concealed for Review

Institutes: Concealed for Review

Abstract. Accurate segmentation and quantification of brain white matter hyperintensities (WMHs) is important for prognosis and disease monitoring. Conventional WMH segmentation techniques employ classifiers, trained on T1 and FLAIR weighted MR images. Yet, there are indications that including features derived from diffusion weighted MRI, can improve classification. However, many features can be computed and feature selection is needed. Selecting adequate features is critical for the performance of WMH classification, but automated approaches might still be sub-optimal. In this work, we propose a novel pipeline for the interactive selection of optimal features for WMH classification, which involves the user in the process. A Visual Analytics (VA) system is employed to exploit the user's cognitive skills and interactively identify the most important features (T1, FLAIR, MD, and RD; and secondarily C_S and FA). Then, these features are used to train a classifier and its results are compared to the state-of-the-art. Finally, the VA is used to evaluate and provide insight into the classifier performance and results.

Keywords: White Matter Hyperintensities (WMHs), Visual Analytics (VA), Classification, Interactive Feature Selection

1 Introduction

White matter hyperintensities of presumed vascular origin (WMHs) are a common finding in conventional MR images of elderly subjects. They are a manifestation of cerebral small vessel disease (SVD) and can be associated with cognitive decline and dementia [1, 2]. Accurate segmentation and quantification of WMHs is needed for prognosis and disease monitoring. To this end, automated WMH classification techniques have been developed [3–5]. Conventional approaches include raw image intensities from T1 and FLAIR weighted MR images, but it is suggested that diffusion MRI can be beneficial for more accurate WMHs segmentation [6, 7]. However, diffusion MRI can provide many features, and feature selection is required. Recent approaches indicate that careful feature selection is more important than the actual classification algorithm [8].

In this work, we propose a novel user-driven, interactive pipeline for identifying an optimal feature list for WMH segmentation. Up to now, expert users have not been involved in the process of feature selection. However, taking advantage of their knowledge and skills can potentially outperform automated approaches.

For this reason, we employ a Visual Analytics (VA) system, where the users can interactively identify the most important features, which are different than the ones obtained automatically, in previous literature. Then, this feature list is used to train a classifier for WMH segmentation, achieving better or similar results than approaches with automatic feature extraction; however, requiring less scanning time and computations. Finally, the VA system is used again to evaluate and to provide understanding on the performance and results of the classifier.

2 Related Work

Visual Analytics (VA) refers to the field that combines in an interactive way, visualizations with pattern recognition, data mining and statistics, focusing on aiding exploration and analytical reasoning [9]. In the past, several VA systems have been proposed for the exploration of feature spaces, among which XmdvTool [10] and SimVis [11]. These approaches employ projections to visualize high-dimensional feature spaces, in combination with other visualization techniques, to explore the feature space of the data. Recently, more interactive approaches for the exploration of large multi-field medical data have emerged [12–14]. Some of these could partially aid the detection of WMHs, while others could partially allow us to explore the respective feature space of imaging modalities that can discriminate between WMHs and healthy tissue. Yet, none of these applications can support all required functionalities, together with an interactive validation with ground truth data. A recent system, proposed by Raidou et al. [15] fulfills these requirements. It was designed to tackle intra-tumor characterization, by employing as central view, a t-Distributed Stochastic Neighborhood Embedding [16] of the imaging-derived features, used in tumor diagnosis. It consists of multiple interactive views for the exploration and analysis of the underlying structure of the feature space, providing linking to anatomy and ground truth data. The application for this system is different, but the concept is similar and we adopt it into our pipeline. To the best of our knowledge, involving VA and users in selecting features for WMHs classifiers has not been addressed before.

3 Materials & Method

3.1 Subjects and MRI

We used the subjects of the MRBrainS13 challenge [17], with additional manual WMH delineations. Subjects included patients with diabetes and matched controls (men: 10, age: 71±4 y). All subjects underwent a standardized 3 T MR exam, including a 3D T1-weighted, a multi-slice FLAIR, a multi-slice IR, and a single-shot EPI DTI sequence with 45 directions. All sequences were aligned with the FLAIR sequence [18]. The diffusion images were corrected for subject motion, eddy current induced geometric distortions [19], and EPI distortions [20], including the required B-matrix adjustments [21], using ExploreDTI [22]. These

Fig. 1. The pipeline proposed for the user-driven, interactive feature selection.

subjects were previously reported in a study of Kuijf et al. [7] for the investigation of the added value of diffusion features in a WMH classifier. We obtained this exact dataset, including the T1, FLAIR, IR images, and the diffusion features FA, MD, axial (AD) and radial (RD) diffusivity, the Westin measures C_L, C_P, C_S [23], and the MNI152-normalized spatial coordinates [3, 7, 18, 24], and we will compare to their results.

3.2 Method

We propose a pipeline for user-driven, interactive selection of features that can be used in a classifier to differentiate between WMH and healthy brain tissue. Our pipeline consists of three steps, depicted in Figure 1. First, the data are explored and analyzed in the VA system proposed by Raidou et al. [15]. From this step, we obtain an optimal list of features for WMH detection, which is used to train a classifier. After classification, the VA system is used again to evaluate and provide a better understanding of the classification process and outcomes.

The adopted VA system [15] is employed to interactively explore the subject data (Figure 2). To easily visualize the otherwise complex high-dimensional feature space of imaging features of each subject, the VA system requires the use of dimensionality reduction. t-Distributed Stochastic Neighbor Embedding (t-SNE) [16] is used to map the high-dimensional feature space into a reduced 2D abstract embedding view, preserving the local structure of the initial space. This 2D embedding view is the central view of the system (Figure 2-ii), where close-by 2D data points reflect voxels with similar behavior in the high-dimensional feature space. Therefore, anatomical structures with similar high-dimensional features are expected to be grouped together in the embedding, in so-called visual clusters. These can be possibly divided in several clusters, also due to the nature of t-SNE. Additionally, the system incorporates anatomical views (Figure 2-i), where bi-directional linking, i.e., association, of the feature space to subject anatomy is possible. Ground truth data, i.e., manual delineations, can be reflected on the embedding view, to identify the visual cluster in the embedding map that contains most of the WMHs. Then, the visual cluster can be interactively selected and its intrinsic feature characteristics can be explored; for example, against (other structures of) the brain, or against WMHs voxels that are not in the selected visual cluster. Several linked views (Figure 2-iii) are interactively updated when the user selects one or more visual clusters in the embedding and depict complementary data information. This includes feature distributions and correlations, multidimensional data patterns, cluster validity

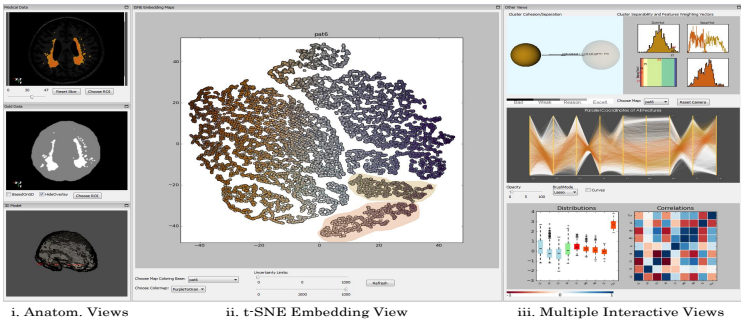

Fig. 2. The adopted VA system [15] during the exploration of the data of a subject from the MRBrainS13 challenge [17]. The three components of the system are denoted.

analysis and information on features that help separating visual clusters from each other, as given by the weights of the separation vector of Linear Discriminant Analysis (LDA). In this way, new aspects of the data are interactively discovered and features suitable for the detection of WMHs are identified.

The list of features resulting from the VA system is used to train a k-nearest-neighbor classifier for WMH segmentation, following the approach of Kuijf et al. [7]. For each set of features, six classifiers are trained with k = 50, 75, or 100, and the neighbor-weighted being either uniform or distance-based. After classification, the results need to be evaluated and understood. In many cases, classifiers are treated as black boxes and the process is limited to trial-and-error, without actual insight on the achieved result. To tackle this, we import the results of the classification into the VA system for additional interactive exploration.

4 Results

Initially, we performed for each one of the subjects a t-SNE with all the features described in Section 3.1, excluding the spatial coordinates. The ground truth, i.e., the manual delineations of WMHs, was initially used in the VA system, to determine whether the voxels of the WMHs form visual clusters, i.e., share similar imaging characteristics. In most of t-SNE embeddings of the subjects, the majority of voxels of the WMHs are grouped together, in one or two visual clusters. From selecting these visual clusters, we could identify that, for cases with two visual clusters, these either correspond to the core and the periphery, or to anterior and posterior WMHs. For large WMHs (top 50%), the visual clusters of the embedding identify 84-98% of the structures. For the rest, the visual clusters can at least detect the core, with a minimum detection percentage of 54%. The multiple interactive linked views of the VA system show that there are comparable behaviors within all cases of visual clusters, especially for larger WMH structures. The cluster analysis view provided information about the separation vector, resulting from LDA between the visual cluster containing most of WMHs and the visual cluster of the rest of the brain. This vector represents the optimal combination of features that is capable of discriminating the two

Subject Data			Features									
ID	Size	% Found	CL	CP	CS	FA	AD	MD	RD	IR	T1	FLAIR
1	1664	98	0,83	0,15	-1,65	-1,46	-0,06	-1,81	3,22	0,3	0,03	1,73
2	1370	93	0,95	0,08	-2,29	-1,99	-0,82	-4,26	5,93	0,16	0,56	2,43
3	1167	91	1,18	0,1	-3,39	-2,94	0	-6,51	7,86	0,1	1,72	2,97
4	819	89	0,75	0,06	-4,75	-4,27	-1,7	-23,91	19,35	0,12	9,48	5,14
5	613	87	1	-0,06	-4,75	-3,36	0,18	-16,87	17,29	-0,12	3,98	3,61
6	571	89	-2,22	-1,48	-8,66	-1,8	0,92	-15,7	21,23	0,46	2,52	6,7
7	494	87	-0,3	-0,17	-2,02	-0,84	0,48	-3,57	4,54	-0,34	1,3	1,83
8	461	84	2,39	1,22	6,61	0,35	-0,34	-8,63	13,37	-0,17	-0,47	-4,19
9	417	79	-3,96	-1,47	8,28	12,06	-0,22	-3,04	8,29	-0,11	-3,54	-7,51
10	353	60	2,6	1,68	-5,58	-9,09	-2,62	-25,04	5,9	-0,41	24,56	7,27
11	328	62	1,75	0,43	-2,07	-2,87	0,28	-8,88	8,56	0,43	2,34	1,6
12	273	65	0,1	-0,2	-2,17	-1,75	0	-10	8,06	-0,4	3,52	1,45
13	265	54	-0,54	-0,19	-4,28	-1,74	-0,84	-15,97	14,54	-0,19	4,87	2,73
14	239	68	0,36	-0,01	-1,79	-0,14	-0,62	-8,58	7,67	-0,01	2,47	1,51
15	228	63	2,82	1,64	7,67	-0,74	-2,06	-10,53	-2,19	1,28	7,31	-4,94
16	221	76	0,81	0,01	-5,09	-3,42	0,64	-8,19	10,32	0,02	2,01	3,35
17	220	84	-0,78	0,22	-5,53	-2,59	-1,74	-15,49	14,91	0,44	5,51	4,72

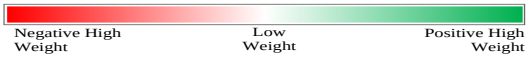

Fig. 3. Results for all subjects: the list of most important features, as resulting from the weights (showed per row, with the color encoding) of the separation vector of the LDA, performed for the detected visual clusters of WMH voxels against the rest of the brain. It is the set of MD, RD, T1 and FLAIR (and then, C_S and FA).

visual clusters from each other. In all, but three cases, the vectors of separation were comparable (Figure 3): T1, FLAIR, RD and MD are important, as they have a considerable weight in the vector of separation. For bigger WMHs, C_S and FA also become important. The contribution of other features such as AD, C_L, C_P and IR is not significant. Taking into account the relations between diffusion metrics, we decided on the optimal set of features for the classifier: MD, RD, T1 and FLAIR (and, secondarily, C_S and FA). Here, we added the spatial coordinates to better represent the brain volume and to suppress non-WMHs.

Based on the results of the VA system, the following six combinations of feature sets $s_i \in S$ were chosen for the classifier: s_1: FLAIR, RD, MD ; s_2: s_1+ T1 ; s_3: s_2+ x, y, z ; s_4: s_1+ FA, C_S ; s_5: s_4+ T1 ; s_6: s_5+ x, y, z. For each classifier trained on a feature set $s_i \in S$, the reported results in Table 1 are the mean $\pm$ sd. These results are compared to the traditional feature sets $f_i \in F$ used by Kuijf et al. [7]: f_1: T1, IR, FLAIR ; f_2: f_1+ x, y, z ; f_3: f_1+ FA, MD ; f_4: f_2+ FA, MD ; f_5: f_4+ C_L, C_P, C_S, AD, RD. A guided selection of features with the VA system can achieve similar or slightly better performance than the naïve selection of traditional features. The two best performing feature sets of Kuijf et al. [7] used 8 (f_4) and 13 (f_5) features respectively, while the current two best methods (s_3 and s_6) use 5 and 8 features, respectively, with comparable results. The VA system led us to discard C_L, C_P, AD and IR, which do not contribute in the classification; hereby, saving scanning and also computational time.

To evaluate the classification outcome, we introduce the results of the two best performing classifiers, s_3 and s_6, into the VA system. This can help us explore and analyze the parts of the WMHs that are missed, but also to understand better how classifiers work and how they can be improved. From an initial inspection, it results that classifier s_3 is restricted to the core of the WMHs,

Table 1. Sensitivity and Dice similarity coefficient (SI) (higher is better) for the classifiers trained on combinations of features $f_i \in F$ (light) and $s_i \in S$ (right).

F	Sensitivity (%)	Dice SI	S	Sensitivity (%)	Dice SI
f_1	59.7 ± 0.2	0.349 ± 0.001	s_1	58.2 ± 0.4	0.434 ± 0.004
f_2	73.4 ± 0.4	0.536 ± 0.005	s_2	64.8 ± 0.2	0.460 ± 0.003
f_3	67.8 ± 0.3	0.411 ± 0.003	s_3	76.2 ± 0.4	0.560 ± 0.005
f_4	77.2 ± 0.4	0.565 ± 0.004	s_4	61.2 ± 0.2	0.446 ± 0.003
f_5	75.2 ± 0.6	0.561 ± 0.003	s_5	66.3 ± 0.2	0.471 ± 0.004
			s_6	76.6 ± 0.5	0.576 ± 0.004

while s_6 detects an extension of it. The WMH core is always detected by both classifiers, as it has consistent imaging characteristics and is well-clustered in the t-SNE embeddings. In subjects with bigger WMHs, s_6 misses only small or thin structures and part of the periphery. In subjects with smaller WMHs, there is a tendency to miss periphery parts and posterior structures more often than the anterior. For bigger WMHs, the core differs in T1, MD, RD with the missed structures. Also, the latter are not as good clustered as the core, in the t-SNE embeddings, i.e., they are probably not coherent in their imaging characteristics. As WMHs become smaller, the influence of T1 becomes less strong, while MD and RD seem to become more important. The classifiers had a low performance for one subject (ID: 8), where all posterior WMHs were missed. The VA system showed that the anterior and posterior WMHs of this subject differ mainly due to MD and RD. Yet, the directional gradients might also be of influence.

5 Discussion & Conclusions

We proposed a user-driven pipeline for aiding the design of classifiers focusing on WMHs segmentation. Firstly, we identified using VA, the list of features (MD, RD, T1, FLAIR and secondarily, FA and C_S), suitable for the separation of WMHs, employing the cognitive skills of the user. Then, this list was used in the classification, with results similar or slightly better than state-of-art approaches. However, in our case less features are used, making the feature calculation less computationally intensive and time consuming. For example, we showed that C_L, C_P, AD and IR can be removed from the feature list, which also saves valuable scanning time (IR: 3:49.6 min). Classifier performance can be further improved by adding additional post-processing to remove false positive detection, which was not performed here, to be comparable to Kuijf et al. [7]. After classification, we evaluated the classifier outcome in the VA system, to understand and generate hypotheses of how it can be improved. The periphery is constantly missed, which could be an indication about the accuracy of the manual delineations of the WMHs. Thin and small structures could have been missed due to partial volume effect, while the directional gradients might be influencing the separation of posterior or anterior WMHs. For certain subjects, the missed structures have intrinsically different characteristics. In this case, more features, such as texture or tensor information, could show if these are indeed different kinds of WMHs or if the currently employed feature space does not entirely describe these WMHs.

Our entire pipeline is user-driven, since the user interacts and guides the analysis. This has the advantage that the cognitive capabilities of the user, which are not easily automatized in most of the cases, can be included in feature selection. However, the results are user-dependent and it remains important to analyze the bias introduced by the user, in such a pipeline. Additionally, our approach can be applied to other classifiers and applications, where imaging feature selection is required. Despite t-SNE is widely used [25] for understanding high dimensional data, errors can also be introduced due to its use. Adding more features for exploration in the VA system, such as textural features or information from tensors, could give interesting results. However, certain visualizations of the VA system do not scale well to a high number of features; thus, new visualizations would be needed to tackle hundreds of features. Finally, evaluating the use of the pipeline with a user study to define its general usefulness and also to assess the performance for other application, is another point for future work. Nevertheless, employing VA in the design of classifiers has potential for better understanding the data under exploration and for obtaining more insight into classifiers.

References

1. Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., Black, S. E. (2013). Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. The Lancet Neurology, 12(8), 822-838.
2. Pantoni, L. (2010). Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. The Lancet Neurology, 9(7), 689-701.
3. Steenwijk, M. D., Pouwels, P. J., Daams, M., van Dalen, J. W., Caan, M. W., Richard, E., Vrenken, H. (2013). Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). NeuroImage: Clinical, 3, 462-469.
4. 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.
5. de Boer, R., Vrooman, H. A., van der Lijn, F., Vernooij, M. W., Ikram, M. A., van der Lugt, A., Niessen, W. J. (2009). White matter lesion extension to automatic brain tissue segmentation on MRI. Neuroimage, 45(4), 1151-1161.
6. Maillard, P., Carmichael, O., Harvey, D., Fletcher, E., Reed, B., Mungas, D., DeCarli, C. (2013). FLAIR and diffusion MRI signals are independent predictors of white matter hyperintensities. American Journal of Neuroradiology, 34(1), 54-61.
7. Kuijf, H. J., Tax, C. M., Zaanen, L. K., Bouvy, W. H., de Bresser, J., Leemans, A., Vincken, K. L. (2014). The Added Value of Diffusion Tensor Imaging for Automated White Matter Hyperintensity Segmentation. In Computational Diffusion MRI (pp. 45-53). Springer International Publishing.
8. Sweeney, E. M., Vogelstein, J. T., Cuzzocreo, J. L., Calabresi, P. A., Reich, D. S., Crainiceanu, C. M., Shinohara, R. T. (2014). A comparison of supervised machine learning algorithms and feature vectors for MS lesion segmentation using multimodal structural MRI. PloS one, 9(4), e95753.
9. Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H. (2008). Visual analytics: Scope and challenges (pp. 76-90). Springer Berlin Heidelberg.

10. Ward, M. O. (1994, October). XmdvTool: Integrating multiple methods for visualizing multivariate data. In Proceedings of the Conference on Visualization'94 (pp. 326-333). IEEE Computer Society Press.
11. Oeltze, S., Doleisch, H., Hauser, H., Muigg, P., Preim, B. (2007). Interactive visual analysis of perfusion data. Visualization and Computer Graphics, IEEE Transactions on, 13(6), 1392-1399.
12. Blaas, J., Botha, C. P., Post, F. H. (2007, May). Interactive Visualization of Multi-Field Medical Data Using Linked Physical and Feature-Space Views. In EuroVis (pp. 123-130).
13. Jeong, D. H., Ziemkiewicz, C., Fisher, B., Ribarsky, W., Chang, R. (2009, June). iPCA: An Interactive System for PCA based Visual Analytics. In Computer Graphics Forum (Vol. 28, No. 3, pp. 767-774). Blackwell Publishing Ltd.
14. Choo, J., Lee, H., Kihm, J., Park, H. (2010, October). iVisClassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on (pp. 27-34). IEEE.
15. Raidou, R. G., Van Der Heide, U. A., Dinh, C. V., Ghobadi, G., Kallehauge, J. F., Breeuwer, M., Vilanova, A. (2015, June). Visual Analytics for the Exploration of Tumor Tissue Characterization. In Computer Graphics Forum (Vol. 34, No. 3, pp. 11-20).
16. Van der Maaten, L., Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9(2579-2605), 85.
17. Mendrik, A. M., Vincken, K. L., Kuijf, H. J., Breeuwer, M., Bouvy, W. H., de Bresser, J., Jog, A. (2015). MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans. Computational Intelligence and Neuroscience, special issue on Simulation and Validation in Brain Image Analysis, 2015, vol. Article ID 813696, p.16
18. Klein, S., Staring, M., Murphy, K., Viergever, M. A., Pluim, J. P. (2010). Elastix: a toolbox for intensity-based medical image registration. Medical Imaging, IEEE Transactions on, 29(1), 196-205.
19. Rohde, G. K., Barnett, A. S., Basser, P. J., Marenco, S., Pierpaoli, C. (2004). Comprehensive approach for correction of motion and distortion in diffusionweighted MRI. Magnetic resonance in medicine, 51(1), 103-114.
20. Irfanoglu, M. O., Walker, L., Sarlls, J., Marenco, S., Pierpaoli, C. (2012). Effects of image distortions originating from susceptibility variations and concomitant fields on diffusion MRI tractography results. Neuroimage, 61(1), 275-288.
21. Leemans, A., Jones, D. K. (2009). The B matrix must be rotated when correcting for subject motion in DTI data. Magnetic Resonance in Medicine, 61(6), 1336-1349.
22. Leemans, A., Jeurissen, B., Sijbers, J., Jones, D. K. (2009, April). ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In 17th Annual Meeting of Intl Soc Mag Reson Med (Vol. 209, p. 3537).
23. Westin, C. F., Maier, S. E., Mamata, H., Nabavi, A., Jolesz, F. A., Kikinis, R. (2002). Processing and visualization for diffusion tensor MRI. Medical image analysis, 6(2), 93-108.
24. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage, 47, S102.
25. Amir, E. A. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nature biotechnology, 31(6), 545-552.