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

Renata Georgia Raidou[1,2], Hugo J. Kuijf[3], Neda Sepasian[1], Nicola Pezzotti[2], Willem H. Bouvy[4], Marcel Breeuwer[5,1], and Anna Vilanova[2,1]

[1] Eindhoven University of Technology, The Netherlands
[2] Delft University of Technology, The Netherlands
[3] Image Sciences Institute, University Medical Center Utrecht, The Netherlands
[4] Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, The Netherlands
[5] Philips Healthcare, Best, The Netherlands

Abstract. Accurate segmentation of brain white matter hyperintensities (WMHs) is important for prognosis and disease monitoring. To this end, classifiers are often trained – usually, using T1 and FLAIR weighted MR images. Incorporating additional features, derived from diffusion weighted MRI, could improve classification. However, the multitude of diffusion-derived features requires selecting the most adequate. For this, automated feature selection is commonly employed, which can often be sub-optimal. In this work, we propose a different approach, introducing a semi-automated pipeline to select interactively features for WMH classification. The advantage of this solution is the integration of the knowledge and skills of experts in the process. In our pipeline, a Visual Analytics (VA) system is employed, to enable user-driven feature selection. The resulting features are T1, FLAIR, Mean Diffusivity (MD), and Radial Diffusivity (RD) – and secondarily, C_S and Fractional Anisotropy (FA). The next step in the pipeline is to train a classifier with these features, and compare its results to a similar classifier, used in previous work with automated feature selection. Finally, VA is employed again, to analyze and understand 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 MR images of elderly subjects. They are a manifestation of cerebral small vessel disease (SVD) and are associated with cognitive decline and dementia [1]. Accurate segmentation of WMHs is important for prognosis and disease monitoring. To this end, automated WMH classification techniques have been developed [2]. Conventional approaches include raw image intensities from T1 and FLAIR weighted MR images, but recently, it has been suggested that diffusion MRI can improve the segmentation [3, 4]. Multiple features can be derived from this imaging modality; thus, careful feature selection is required.

In this work, we propose a semi-automated approach, to aid the design of WMH classifiers. Our novelty is the introduction of a user-driven, interactive pipeline that provides new insight into the entire classification procedure – especially, in the identification of an adequate feature list, and the analysis of the outcome. Up to now, the knowledge and cognitive skills of experts have not been intensively involved in the process. In the *first step* of our pipeline, we employ a Visual Analytics (VA) system [5], where expert users select interactively the most important features. In the *second step*, the resulting feature list is used to train a classifier for WMH segmentation. The performance and results of this classifier can be analyzed and interpreted in the *final step* of the pipeline, using again VA.

2   Related Work

Visual Analytics (VA) refers to the field that combines, through interaction, visualizations with pattern recognition, data mining and statistics, and focuses on aiding exploration and analytical reasoning [6]. Recently, Raidou et al. proposed a highly interactive VA system for the exploration of intra-tumor tissue characteristics [5]. The system employs a t-Distributed Stochastic Neighbor Embedding [7] of several imaging-derived features, used in tumor diagnosis. It also 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. Yet, to the best of our knowledge, involving users through VA and interaction in an entire pipeline for feature selection, classification and outcome evaluation for WMH structures has not been addressed before.

3   Materials & Method

3.1   Subjects and MRI Data

We used the subjects of the MRBrainS13 challenge [8], 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 [9]. The diffusion images were corrected for subject motion, eddy current induced geometric distortions, and EPI distortions, including the required B-matrix adjustments, using ExploreDTI [10].

The dataset includes T1, FLAIR and IR weighted images, as well as the following diffusion features: Fractional Anisotropy (FA), Mean Diffusivity (MD), Axial Diffusivity (AD), Radial Diffusivity (RD), the Westin measures C_L, C_P, C_S [11], and MNI152-normalized spatial coordinates [9, 12]. This exact dataset has been previously reported in a study of Kuijf et al. [3], for the investigation of the added value of diffusion features in a WMH classifier. Since we could have access to the exact same data and we share the same goal, we will use the previous work of Kuijf et al., as a baseline for the evaluation of our results.
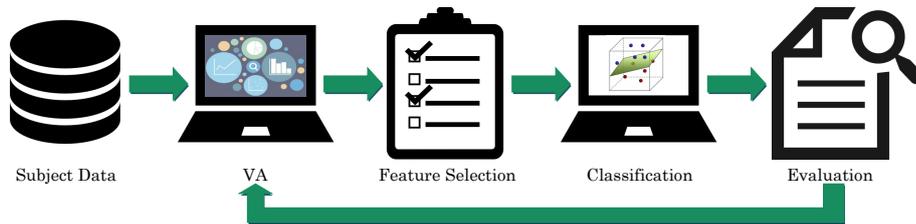
Fig. 1. The pipeline proposed for the user-driven feature selection, classification and outcome evaluation for the segmentation of White Matter Hyperintensities (WMHs).

3.2 Method

In this section, we describe our new pipeline⋆ for the user-driven, interactive selection of features that can differentiate WMHs from healthy brain tissue. Our pipeline consists of three steps, depicted in Figure 1. First, the data are interactively explored and analyzed by expert users, in the VA system proposed by Raidou et al. [5]. From this step, we obtain through interaction and visual analysis, a list of features, adequate for WMH detection. These features are subsequently used to train a classifier. After classification, the VA system is used again to evaluate and better understand the classification process and outcome.

Feature Selection using VA. The VA system of Raidou et al. [5] is employed to interactively explore the data of each one of the available subjects (Figure 2). Initially, t-Distributed Stochastic Neighbor Embedding (t-SNE) [7] is used to map the high-dimensional feature space of each subject (described in Section 3.1) into a reduced 2D abstract embedding view, preserving the local structure of the feature space. Spatial coordinates are excluded, as we are interested in preserving similarities in the feature space, and the voxel positions could introduce bias. In the resulting embedding view (Figure 2-ii), close-by 2D data points reflect voxels with similar behavior in the high-dimensional feature space. Therefore, voxels from structures with similar imaging characteristics are expected to be grouped together in the embedding, in so-called visual clusters. Having available ground truth data, i.e., manual delineations of the WMHs, allows to associate visual clusters from the feature space to anatomy, and vice versa (Figure 2-i).

When a WMH-containing visual cluster is interactively selected, its intrinsic feature characteristics are explored; for example, against other structures of the brain, or against WMHs voxels that are not within the selected visual cluster. Then, several linked views (Figure 2-iii) are interactively updated with complementary data information. This includes feature distributions and correlations, multidimensional data patterns, cluster validity 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, features suitable for the detection of WMHs are interactively identified.

⋆An interactive demo can be found here: `https://vimeo.com/170609498`

i. Anatom. Views   ii. t-SNE Embedding View   iii. Multiple Interactive Views
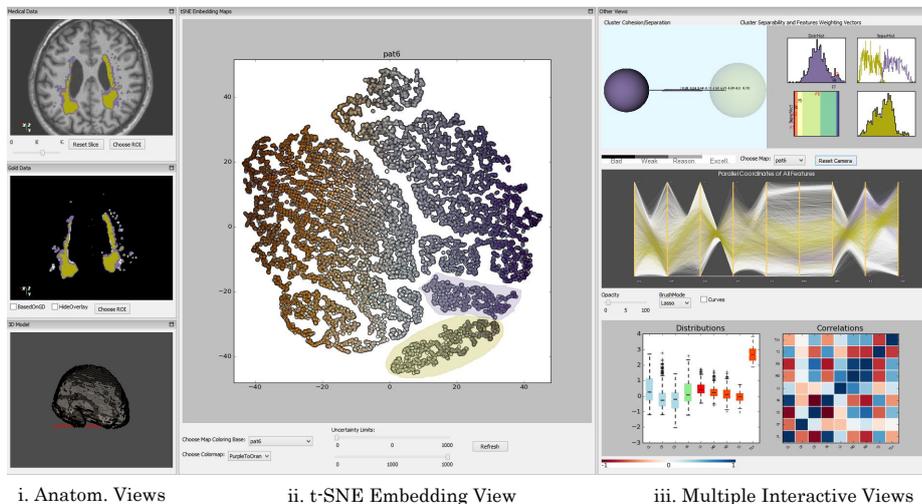
Fig. 2. The adopted VA system [5] during the exploration of the data of a subject from the MRBrainS13 challenge [8]. The three components of the system are denoted.

For example, for the subject of Figure 2, two visual clusters have been selected in the t-SNE of the middle view. As depicted in the anatomical views, one corresponds to the WMH core (green) and the other to the periphery (purple). Together, they represent the biggest part of the structure. Still, several small parts are missed. The separation vector, resulting from LDA between the two visual clusters containing the WMHs against the rest of the brain, is extracted. From the weights of this vector, features adequate for differentiating the detected WMHs from the rest of the brain, are identified. This analysis is subject-specific and has to be performed on a single-subject basis. When all subjects have been explored, the user decides on the most suitable feature list, overall.

Classification. In this step, many different classification approaches could be followed, but comparing to all would be out of scope, for this work. Recently, Kuijf et al. [3] presented an approach for WMH classification, using the same set of diffusion features. To evaluate whether our user-driven feature selection outperforms automated feature selection, we adopt a similar classification approach, as in the previous work of Kuijf et al. The list of features resulting from the VA system is used to train a k-nearest-neighbor classifier for WMH segmentation. For different feature combinations, several classifiers are trained with $k = 50$, 75, or 100, and the neighbor-weighted is either uniform or distance-based [3].

Evaluation of Classification. In many cases, classifiers are treated as black boxes, and users do not have actual insight into the achieved result. With this step, we want to provide a way for evaluating and understanding both the results of the classifier and the classifier itself. To this end, we import the binary

masks resulting from the classification (detected vs. missed WMHs) into the VA system [5]. The user can interactively explore the high-dimensional feature space of the two regions of the mask, and generate hypotheses about why the classifier failed to detect parts of the WMHs, with respect to the imaging features.
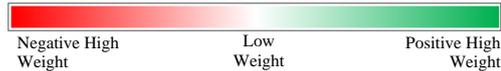
4   Results

Feature Selection using VA. 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, similar to the case depicted in Figure 2. From selecting these visual clusters, we could identify that, for subjects 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

Table 1. The most important features for each subject, as resulting from the weights of the LDA separation vector, performed for the detected visual clusters of WMH voxels against the rest of the brain. The second column denotes the size of WMHs in voxels. The third column shows the percentage of WMHs detected by visual clusters in the VA tool. The other columns represent features, and their weights are color encoded per row. The resulting feature list is the set MD, RD, T1 and FLAIR (then, C_S and FA).

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

Negative High Weight	Low Weight	Positive High Weight

WMH structures. As mentioned before, the cluster analysis view of the VA system provides the separation vector, resulting from LDA between the visual cluster containing most of WMHs and the visual cluster of the rest of the brain. Table 1 depicts, for all investigated subjects, the weights of separation for these two visual clusters. In all – but three – cases, T1, FLAIR, RD and MD are more important, as they have a considerable weight. For bigger WMHs, C_S and FA also become important. The contribution of other features such as AD, C_L, C_P and IR seems not significant.

Considering also the (cor-)relations between diffusion features, we decide on the overall set of features for the classifier: MD, RD, T1 and FLAIR (secondarily, C_S and FA). Here, we add the MNI152-normalized spatial coordinates (x, y, z) to better represent the brain volume and to suppress non-WMH structures.

Classification. Based on the results of the VA system, the following four combinations of feature sets $s_i \in S$ are chosen for our k-NN classifier: s_1: MD, RD, T1, FLAIR; s_2: $s_1+ (x, y, z)$; s_3: $s_1+ C_S$, FA; s_4: $s_3+ (x, y, z)$. For each classifier trained on a feature set $s_i \in S$, we measure the sensitivity and Dice similarity coefficient (mean ± standard deviation), as shown in Table 2. These measurements are performed, with respect to the available manual delineations of the WMH structures. Furthermore, our results are compared to the feature sets $f_i \in F$, previously used by Kuijf et al. [3]: 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.

The results of Table 2 demonstrate that our proposed VA-guided feature selection can achieve similar or slightly better performance than the automated feature selection, presented by Kuijf et al. [3]. The two best performing feature sets of Kuijf et al. used 8 (f_4) and 13 (f_5) features, while our current two best methods use 7 (s_2) and 8 (s_4) features only, with comparable results. Our approach allows to discard C_L, C_P, AD and IR, which do not contribute in the classification; hereby, saving scanning and also computational time.

Evaluation of Classification. To evaluate the classification outcome, we introduce the results of the two best performing classifiers, s_2 and s_4, into the VA

Table 2. Sensitivity, Dice similarity coefficient (SI, higher is better) and number of features for the classifiers, trained on combinations of features $s_i \in S$ (left, from our VA-driven approach) and $f_i \in F$ (right, from [3]), with respect to the available manual delineations.

Our VA-driven approach				Automated approach from [2]			
S	Sensitivity (%)	Dice SI	Features	F	Sensitivity (%)	Dice SI	Features
s_1	64.8 ± 0.2	0.460 ± 0.003	4	f_1	59.7 ± 0.2	0.349 ± 0.001	3
s_2	76.2 ± 0.4	0.560 ± 0.005	7	f_2	73.4 ± 0.4	0.536 ± 0.005	6
s_3	66.3 ± 0.2	0.471 ± 0.004	5	f_3	67.8 ± 0.3	0.411 ± 0.003	5
s_4	76.6 ± 0.5	0.576 ± 0.004	8	f_4	77.2 ± 0.4	0.565 ± 0.004	8
				f_5	75.2 ± 0.6	0.561 ± 0.003	13

system. One of the goals is to explore and analyze the parts of the WMHs that are missed, but also to understand better how these classifiers work and how they can be improved. From an initial inspection, it results that classifier s_2 is restricted to the core of the WMHs, while s_4 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_4 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 in the t-SNE embeddings as the core, i.e., they are 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.

5   Discussion & Conclusions

We proposed a user-driven pipeline for aiding the design of classifiers, focusing on WMH segmentation. Using VA and the cognitive skills of an expert user, we initially identified the list of features (MD, RD, T1, FLAIR, and secondarily, FA and C_S) that are suitable for the separation of WMHs. Then, this list was used for WMH classification. In respect of previous work [3], our results are comparable. Yet, our results are not achieved through trial-and-error, but after a justifiable and understandable, interactive feature selection. Additionally, our approach requires less features, which allows to skip several imaging sequences, making the feature calculation less computationally intensive and time consuming. For example, we concluded that C_L, C_P, AD and IR can be omitted, which saves valuable scanning time (IR: 3:49.6 min).

After classification, we evaluated the classifier outcome in the VA system. The periphery is constantly missed. Thin and small structures can be missed due to partial volume effect, while the MNI152-normalized spatial coordinates can influence the separation of posterior or anterior WMHs. For certain subjects, the missed structures have intrinsically different imaging characteristics. In this case, more features, such as texture or tensor information, should be further investigated. The performance of the classifier could 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. [3]. Also, it would be interesting to investigate what happens, when our VA-selected features are used with more sophisticated classification algorithms.

In the entire pipeline, the user interacts and guides the analysis. This has the advantage that the cognitive capabilities of the user, which are not easily automatized, can be included in feature selection. However, the results are user-dependent and it remains important to analyze the bias introduced by the user. Although t-SNE is widely used [13] 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, 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 and the frequently exploding set of imaging features.

References

1. Pantoni, L.: Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. The Lancet Neurology, 9(7), 689-701 (2010).
2. 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), 1037-1044 (2004).
3. Kuijf, H. J., Tax, C. M., Zaanen, L. K., Bouvy, W. H., de Bresser, J., Leemans, A., Vincken, K. L.: The Added Value of Diffusion Tensor Imaging for Automated White Matter Hyperintensity Segmentation. In Computational Diffusion MRI (pp. 45-53). Springer International Publishing (2014).
4. 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), 54-61 (2013).
5. Raidou, R. G., Van Der Heide, U. A., Dinh, C. V., Ghobadi, G., Kallehauge, J. F., Breeuwer, M., Vilanova, A.: Visual Analytics for the Exploration of Tumor Tissue Characterization. In Computer Graphics Forum (Vol. 34, No. 3, pp. 11-20) (2015).
6. Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H.: Visual analytics: Scope and challenges (pp. 76-90). Springer Berlin Heidelberg (2008).
7. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research, 9(2579-2605), 85 (2008).
8. Mendrik, A. M., Vincken, K. L., Kuijf, H. J., Breeuwer, M., Bouvy, W. H., de Bresser, J., Jog, A.: 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, vol. Article ID 813696, p.16 (2015).
9. Klein, S., Staring, M., Murphy, K., Viergever, M. A., Pluim, J. P.: Elastix: a toolbox for intensity-based medical image registration. Medical Imaging, IEEE Transactions on, 29(1), 196-205 (2010).
10. Leemans, A., Jeurissen, B., Sijbers, J., Jones, D. K.: 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) (2009).
11. Westin, C. F., Maier, S. E., Mamata, H., Nabavi, A., Jolesz, F. A., Kikinis, R.: Processing and visualization for diffusion tensor MRI. Medical image analysis, 6(2), 93-108 (2002).
12. 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, S102 (2009).
13. Amir, E. A. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., Pe'er, D.: viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nature biotechnology, 31(6), 545-552 (2013).