Identification of Cerebral Small Vessel Disease Using Multiple Instance Learning

Liang Chen[1], Tong Tong[1], Chin Pang Ho[1], Rajiv Patel[2], David Cohen[2], Angela C. Dawson[3], Omid Halse[3], Olivia Geraghty[1], Paul E.M. Rinne[1], Christopher J. White[1], Tagore Nakornchai[1], Paul Bentley[1], and Daniel Rueckert[1]

[1] Imperial College London, London, United Kingdom
[2] Northwick Park Hospital, London, United Kingdom
[3] Imperial College Healthcare NHS Trust, London, United Kingdom

Abstract. Cerebral small vessel disease (SVD) is a common cause of ageing-associated physical and cognitive impairment. Identifying SVD is important for both clinical and research purposes but is usually dependent on radiologists' evaluation of brain scans. Computer tomography (CT) is the most widely used brain imaging technique but for SVD it shows a low signal-to-noise ratio, and consequently poor inter-rater reliability. We therefore propose a novel framework based on multiple instance learning (MIL) to distinguish between absent/mild SVD and moderate/severe SVD. Intensity patches are extracted from regions with high probability of containing lesions. These are then used as instances in MIL for the identification of SVD. A large baseline CT dataset, consisting of 590 CT scans, was used for evaluation. We achieved approximately 75% accuracy in classifying two different types of SVD, which is high for this challenging problem. Our results outperform those obtained by either standard machine learning methods or current clinical practice.

1 Introduction

The World Health Organization (WHO) states that stroke is the second major cause of death in the world during 2000 and 2012. Stroke, which is a cerebrovascular accident, is the loss of brain function caused by the lack of blood supply [16]. It may lead to long-term disability. Ischemic stroke and hemorrhagic stroke are two different categories of strokes that require different treatments [6]. Ischemic stroke accounts for approximately 80% of all strokes [7]. Intravenous thrombolysis with recombinant tissue plasminogen activator (rt-PA) is the recommended therapy for acute ischemic stroke that reduces severe disability but causes deterioration due to symptomatic intracranial hemorrhage (SICH) in approximately 6% [18]. [4] demonstrates that cerebral SVD is associated with increased risk of ischemic stroke. Hypertensive SVD is the most common mechanism of hemorrhagic stroke [6]. In order to reduce the rate of SICH, which is associated with the worst outcome of stroke, management of SVD is pivotal. Cerebral SVD refers to a group of pathological aetiologies that affect the brain [12]. However, in this

N. Navab et al. (Eds.): MICCAI 2015, Part I, LNCS 9349, pp. 523–530, 2015.
DOI: 10.1007/978-3-319-24553-9_64

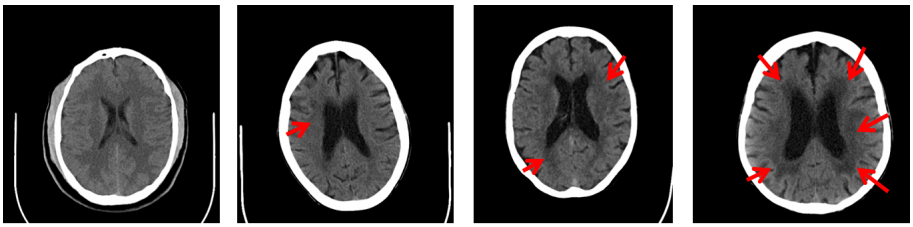

Fig. 1. Examples of CT images of the brain: (a) normal brain appearance, (b) brain with mild cerebral SVD, (c) brain with moderate SVD, and (d) cerebrum with severe SVD. The red arrows point out where the lesions are.

paper we will use the term to describe ischemic consequences of white matter (WM) lesions. Figure 1 presents examples of cerebrums with different kinds of SVD.

Advanced neuroimaging techniques have been widely used in the diagnosis of stroke. It is normally recommended that patients should undergo either magnetic resonance (MR) or CT imaging [19]. Diffusion-weighted imaging (DWI) and T2-fluid attenuated inversion recovery (FLAIR) should be included in the MR sequences, which are able to show any acute or chronic lesions. Although MR is the gold standard, CT is more frequently used in the acute phase of stroke treatment. This is due to the fact that there is typically no MR scanner access in emergency rooms in hospitals. For patients suffering from acute stroke it is therefore desirable to have reliable and automatic image analysis techniques for CT images.

There have been a large number of studies focusing on the automatic analysis of brain MR images. For instance, in Alzheimer's Disease (AD), machine learning techniques have been extensively used to classify controls and patients. However, there are very few works that focus on the classification of subjects suffering from stroke and even fewer which use CT images [5]. The cutting-edge studies on CT images, including [3], [17], and [9], are typically based on statistical values and threshold. These methods are fairly simple that it is difficult to apply to large datasets. To the best of our knowledge no machine learning approach has been proposed for the identification of SVD in a large dataset of CT images.

Fazekas et al. [8] proposed a standard approach for SVD grading. In this approach, SVD is divided into four categories according to the degree of the lesions: absent, mild, moderate, and severe. Generally, mild SVD is associated with normal brain ageing while moderate or severe SVD suggests potential risks for diseases such as stroke. One of the challenges for the grading of SVD is that the difference between absent/mild and moderate/severe SVD are often very subtle. In the context of similar challenging classification problems in medical images, semi-supervised machine learning approaches have been very successful. MIL is one example of such a semi-supervised learning method [21]. It solves the

problem that standard approaches are difficult to distinguish lesions and normal tissues at a voxel or a patch.

In MIL, instances are contained in bags. A bag is positive if there is one positive instance in it; otherwise the bag is negative. Different bags may contain various numbers of instances. Compared to other standard supervised and unsupervised learning algorithms, there are many MIL methods that have been developed and applied, e.g. MIS-Boost [1], MIForest [11], and EM-Diverse Density [20]. In [1], the authors proposed a boosting based MIL, which outperforms a number of other similar algorithms on several benchmark datasets. This approach aims to learn a specific instance for each weak classifier, which is able to discriminate two categories of instances.

In this paper, we tackle the problem of automatic SVD identification. An MIL framework is formulated to classify SVD into normal (absent and mild) and abnormal (moderate and severe) groups, which is based on a large dataset of CT images. In Section 2, details of the MIS-Boost algorithm and our further optimization will be presented. Section 3 demonstrates how patches were extracted. We will show our imaging dataset, the pre-processing techniques, and our model based on MIS-Boost in Section 4. Comparisons between our model and other state-of-the-art algorithms will also be shown.

2 Methods

Given bags and their labels, MIL is recognized as a supervised learning method, which learns the mapping $\mathbf{X} \rightarrow \mathbf{Y}$, where $\mathbf{X}$ is a set of training data and $\mathbf{Y} = \{-1, +1\}$ is the set of corresponding labels. In this case, $\mathbf{X} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_N\}$ and for each bag $\mathbf{B}_i = \{\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_{n_i}\}$, where $\mathbf{I}_k \in \Re^r$ is the k-th instance in bag $\mathbf{B}_i$. N is the number of bags. n_i is the number of instances in the i-th bag. r is the size of a patch. The boosting-based MIL proposed in [1] aims to learn a 'bag-level' classifier

$$F(\mathbf{B}) = \text{sign}\left(\sum_{m=1}^{M} f_m(\mathbf{B})\right), \tag{1}$$

where $f_m(\cdot), m = 1, 2, \ldots, M$, are weak classifiers defined as

$$f_m(\mathbf{B}) = \frac{2}{1 + e^{-(\beta_1 D(\mathbf{p}_m, \mathbf{B}) + \beta_0)}} - 1. \tag{2}$$

The task of each weak classifier is to find a patch $\mathbf{p}_m$, which serves as an instance, to discriminate different bags. In [1], the distance from an instance to a bag is defined as below

$$D(\mathbf{p}_m, \mathbf{B}) = \sum_{k=1}^{n} \pi_k d(\mathbf{p}_m, \mathbf{I}_k), \tag{3}$$

where $d(\mathbf{p}_m, \mathbf{I}_k) = \|\mathbf{p}_m - \mathbf{I}_k\|_2$ and $\pi_k = \frac{e^{-\alpha d(\mathbf{p}_m, \mathbf{I}_k)}}{\sum_{l=1}^{n} e^{-\alpha d(\mathbf{p}_m, \mathbf{I}_l)}}$. $d(\mathbf{p}_m, \mathbf{I}_k)$ is the distance between the specific instance and the k-th instance in the bag, which is

the standard Euclidean distance and π_k is its weight. α is a constant. $D(\mathbf{p}_m, \mathbf{B})$ is the weighted average distance of $\mathbf{p}_m$ to each instance in the bag and $f_m(\cdot)$ maps the distance into the range $[-1, 1]$.

In order to learn $\mathbf{p}_m$, [1] defined an error function based on Gentle Adaboost. We obtained the parameters β_0, β_1, and $\mathbf{p}_m$ by minimizing the weighted error between the ground truth labels and the decision made by weak classifiers.

$$\min_{\mathbf{p}_m, \beta_0, \beta_1} \varepsilon_m = \sum_{i=1}^{N} w_i (y_i - f_m(\mathbf{B}_i))^2 \tag{4}$$

In [1] the optimization problem is solved via a coordinate descent algorithm. This uses a line-search method and therefore does not require the calculation of derivatives. However, each iteration is very time-consuming. In this work, we propose an optimization using a region-trust-reflective method [2] to allow a more efficient optimization that exploits the fact that the function above is differentiable. We formulated optimization of the objective function as a non-linear least square fitting problem.

For initialization, we performed k-means clustering for all instances in all the bags. The resulting K clustering centres were used as input for the initial $\mathbf{p}_m$ and we selected one leading to the minimum error ε_m among them. In order to decide on the number of weak classifiers M, we split the training dataset into sub-training and validation sets and pick up final M with minimum validation error.

3 Patch Extraction

In MIL, each bag contains a number of instances, which are patches in our case. Patches were extracted from original CT images since the slice thickness varies between different scans and resampling them to a constant voxel size will reduce the image quality. The extraction was guided by an atlas, which shows the regions with high probability of lesions. In order to construct such an atlas, we collected 277 MR images with SVD. For all MR images, clinical experts manually outlined regions of interests (ROIs) corresponding to the SVD lesions. They were then registered and normalized onto a standard space so that we are able to obtain the lesion atlas. The atlas constructed shows the probability for each voxel in the brain to be part of an SVD lesion. We excluded the regions with very low abnormal probability ($< 4\%$) in the atlas since they are likely to be outliers. Finally, the lesion atlas was mapped back to each individual CT image so that for each CT image a lesion atlas is available which shows regions with high probability of lesions. Figure 2 visualizes this processing pipeline.

4 Experiments and Results

4.1 Imaging Data and Pre-processing

In this study, all the data was collected from a local hospital. We collected 627 baseline CT brain images with stroke. For all patients, the imaging was carried

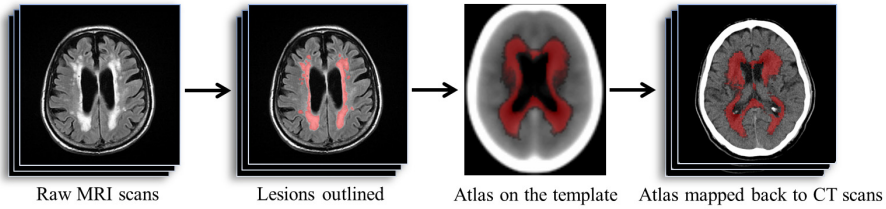

Fig. 2. The process of atlas construction and mapping back. The red regions are the ROIs for patch extraction.

out within a short time window after stroke (4.5 hours). The average age of these subjects is 70.75 ± 10.83. There are 326 male and 301 female participants. The labels of these images were assessed by an expert according to [8] and there are inter-rater consistencies of about 75% between experts.

We developed a pipeline to normalize the images before analysis. All images were registered to a CT-based template and then resampled to a uniform voxel size. The template was developed by [14]. In order to reduce the radiation burden for patients, in some subjects the brains were scanned in two separate volumes including the cerebrum and the base using different voxel sizes. For the images scanned separately, the voxel sizes of cerebrum and base are approximately $0.45 \times 0.45 \times 7.2$ mm and $0.45 \times 0.45 \times 2.4$ mm; the voxel size of the whole-brain scans is approximately $0.38 \times 0.38 \times 3$ mm, while the template's voxel size is $2 \times 2 \times 2$ mm. We combined these sub-volumes into single volumes with constant voxel size. Subsequently we corrected the gantry tilt and rigidly co-registered all images to the template. Following this step, a non-rigid registration [15] was performed between all images and the template. Finally, all images were resampled onto the voxel grid of the template. The processing pipeline failed for 37 CT scans because of poor image quality and/or patient movement. These subjects were excluded and we used the remaining 590 scans in the following experiments, which consists of 350 with absent/mild SVD and 240 with moderate/severe SVD.

4.2 Patch-Based Identification of SVD

In order to have two SVD groups which are balanced in terms of number of subjects, we randomly sampled 240 subjects from the absent or mild group and performed leave-10%-out cross-validation. The random sampling was repeated for $T = 10$ times and the final results are average values of the T repeats. In this paper, abnormal bags and instances are regarded as positive.

In MIS-Boost, each subject is modelled as a bag, which can contain a number of patches as the instances. The patches were extracted from an ROI according to the atlas. The ROI is defined by those voxels in which the prior probability for lesions is not low. As different original CT scans have different numbers of slices and the size of the brain varies, different bags contain different numbers of instances. We obtain on average 2313 patches (SD: 762) in a bag. Given the

Table 1. Classification performance of different classifiers and features. Results of MIS-Boost and random forest are based on T times cross-validation.

Classifier	Feature	Accuracy(%)	Sensitivity(%)	Specificity(%)
MIS-Boost	Patch in ROI	**75.04**±1.37	**80.17**±1.65	69.92 ± 1.37
Random forest	Voxel in ROI	70.65 ± 0.03	69.63 ± 0.04	**71.67**±0.04
	Voxel in whole brain	65.25 ± 0.02	65.64 ± 0.04	64.96 ± 0.04
Threshold	t-Score	54.07	5.42	48.64

different slice thickness of the different scans, 2D patches were extracted with a patch size of 15×15. The performance is shown in Table 1.

In order to demonstrate the performance of our model, we compared the results to those obtained using alternative approaches. We compared our approach to random forests [10]. It is one of the most popular standard machine learning methods and has achieved a notable success in classification of AD patients and controls using imaging data [13]. As the CT images have been registered and normalized to the template, the voxels of processed images were selected as features for the random forests. Voxels were extracted from the whole brain and the ROI, respectively. We also compared the approach by [9] which has shown the ability for automated stroke lesion delineation using brain CT images. In this approach a t-score map is calculated, which when combined with a carefully selected threshold, can be used to delineate stroke lesions. Since acute stroke lesions are similar to SVD in terms of intensity and texture, this approach can be tested in terms of its performance for the evaluation of SVD. We collected 307 CT images without SVD to calculate the standard t-score map in template's space and mapped it back to each native image space. For each individual subject, we delineated the potential SVD lesions by applying the selected threshold to its t-score map and therefore obtained the volume of the lesions. We then sorted the volumes of all subjects and chose the median as the threshold to distinguish normal and abnormal subjects in terms of SVD.

According to Table 1, our implementation of patch-based MIS-Boost outperforms the other two methods. It is clear that a simple method based on thresholds is unreliable since its sensitivity is low. This means it cannot detect abnormal subjects. Compared with the threshold-based method, the random-forest-based method improves the accuracy by 10%. In addition, the random forest classifier is sufficiently robust as the gap between sensitivity and specificity is small. The use of voxels from the ROI defined by the atlas enhances the accuracy by 5% compared to using voxels from the whole brain. Furthermore, our proposed model boosts the classification accuracy by an additional 5%. Apart from the high accuracy of classification, the sensitivity of MIS-Boost is high.

5 Discussion and Conclusion

We have presented a framework in which boosting based MIL is used to learn patches for discrimination of normal or abnormal brain degeneration. A key feature of the proposed method is that it has been applied a large clinical CT dataset for automatic clinical identification of SVD. In addition, patches from original CT scans were employed, which avoids additional errors. To the best of our knowledge of this is the first such application of automated detection of SVD in such a large dataset. We have also shown that the classification results obtained using a state-of-the-art classification technique such as random forests is not as good as the proposed approach.

The proposed approach uses an atlas of SVD lesions derived from MR images. Compared with the low resolution of CT images, MR images are able to show brain lesions in detail. MR imaging is therefore regarded as the gold standard in the assessment of SVD. This provides prior knowledge where lesions occur frequently in the brain. The MR images used for the atlas construction are separate from the images that are used for training and/or testing.

The proposed method also showed its strength compared to standard clinical approaches, where basic statistical features are used. Since CT images show a low signal-to-noise ratio, small lesions like SVD are difficult to be identified at a voxel. In contrast, patch-based features decrease the effect of noise.

In the future, the proposed method will be applied to a larger dataset including data from different clinical centres so that the framework can be tested more widely in terms of robustness and accuracy. More importantly, our final goal is to predict the outcome of stroke - whether the stroke patients will hemorrhage or not. This will help to reduce the rate of SICH significantly, which will improve quality of patients' lives and reduce the pressure for the public health services.

References

[1] Akbas, E., Ghanem, B., Ahuja, N.: MIS-Boost: Multiple instance selection boosting. arXiv preprint:1109.2388 (2011)

[2] Byrd, R.H., Schnabel, R.B., Shultz, G.A.: Approximate solution of the trust region problem by minimization over two-dimensional subspaces. Mathematical Programming 40(1–3), 247–263 (1988)

[3] Chawla, M., Sharma, S., Sivaswamy, J., Kishore, L.: A method for automatic detection and classification of stroke from brain CT images. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3581–3584. IEEE (2009)

[4] Conijn, M.M.A., Kloppenborg, R.P., Algra, A., Mali, W.P.T.M., Kappelle, L.J., Vincken, K.L., Van Der Graaf, Y., Geerlings, M.I.: Cerebral small vessel disease and risk of death, ischemic stroke, and cardiac complications in patients with atherosclerotic disease: The second manifestations of arterial disease-magnetic resonance (SMART-MR) study. Stroke 42, 3105–3109 (2011)

[5] Dalca, A.V., et al.: Segmentation of cerebrovascular pathologies in stroke patients with spatial and shape priors. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 773–780. Springer, Heidelberg (2014)

[6] Donnan, G., Fisher, M., Macleod, M., Davis, S.M.: Stroke. The Lancet 371, 1612–1623 (2008)
[7] Durukan, A., Tatlisumak, T.: Acute ischemic stroke: Overview of major experimental rodent models, pathophysiology, and therapy of focal cerebral ischemia. Pharmacology Biochemistry and Behavior 87, 179–197 (2007)
[8] Fazekas, F., Chawluk, J.B., Alavi, A., Hurtig, H.I., Zimmerman, R.A.: MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. American Journal of Neuroradiology 8(3), 421–426 (1987)
[9] Gillebert, C.R., Humphreys, G.W., Mantini, D.: Automated delineation of stroke lesions using brain CT images. NeuroImage: Clinical 4, 540–548 (2014)
[10] Ho, T.K.: The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8), 832–844 (1998)
[11] Leistner, C., Saffari, A., Bischof, H.: MIForests: Multiple-instance learning with randomized trees. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 29–42. Springer, Heidelberg (2010)
[12] Pantoni, L.: Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. The Lancet Neurology 9(7), 689–701 (2010)
[13] Ramírez, J., Górriz, J., Segovia, F., Chaves, R., Salas-Gonzalez, D., López, M., Álvarez, I., Padilla, P.: Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification. Neuroscience Letters 472(2), 99–103 (2010)
[14] Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., Karnath, H.O.: Age-specific CT and MRI templates for spatial normalization. NeuroImage 61, 957–965 (2012)
[15] Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging 18(8), 712–721 (1999)
[16] Sims, N.R., Muyderman, H.: Mitochondria, oxidative metabolism and cell death in stroke. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1802(1), 80–91 (2010)
[17] Takahashi, N., Tsai, D.Y., Lee, Y., Kinoshita, T., Ishii, K.: Z-score mapping method for extracting hypoattenuation areas of hyperacute stroke in unenhanced CT. Academic Radiology 17(1), 84–92 (2010)
[18] Wardlaw, J.M., Murray, V., Berge, E., Del Zoppo, G., Sandercock, P., Lindley, R.L., Cohen, G.: Recombinant tissue plasminogen activator for acute ischaemic stroke: An updated systematic review and meta-analysis. The Lancet 379, 2364–2372 (2012)
[19] Wintermark, M., Albers, G.W., Alexandrov, A.V., Alger, J.R., Bammer, R., Baron, J.C., Davis, S., Demaerschalk, B.M., Derdeyn, C.P., Donnan, G.A., et al.: Acute stroke imaging research roadmap. American Journal of Neuroradiology 29(5), e23–e30 (2008)
[20] Zhang, Q., Goldman, S.: EM-DD: An improved multiple-instance learning technique. In: Advances in Neural Information Processing Systems, pp. 1073–1080 (2001)
[21] Zhou, Z.H.: Multi-instance learning: a survey. Tech. rep., National Laboratory for Novel Software Technology, Nanjing (2004)