Computer-Assisted Segmentation of White Matter Lesions in 3D MR Images Using Support Vector Machine[1]

Zhiqiang Lao, Dinggang Shen, Dengfeng Liu, Abbas F. Jawad, Elias R. Melhem, Lenore J. Launer
R. Nick Bryan, Christos Davatzikos

Rationale and Objectives. Brain lesions, especially white matter lesions (WMLs), are associated with cardiac and vascular disease, but also with normal aging. Quantitative analysis of WML in large clinical trials is becoming more and more important.

Materials and Methods. In this article, we present a computer-assisted WML segmentation method, based on local features extracted from multiparametric magnetic resonance imaging (MRI) sequences (ie, T1-weighted, T2-weighted, proton density-weighted, and fluid attenuation inversion recovery MRI scans). A support vector machine classifier is first trained on expert-defined WMLs, and is then used to classify new scans.

Results. Postprocessing analysis further reduces false positives by using anatomic knowledge and measures of distance from the training set.

Conclusions. Cross-validation on a population of 35 patients from three different imaging sites with WMLs of varying sizes, shapes, and locations tests the robustness and accuracy of the proposed segmentation method, compared with the manual segmentation results from two experienced neuroradiologists.

Key Words. White matter lesion segmentation; support vector machine; machine learning

Cerebrovascular disease (CVD) in elderly individuals is very important. In particular, CVD increases the likelihood of clinical dementia (1–4) even in the absence of clinical stroke (5), albeit the literature is somewhat inconclusive as to whether CVD has simply an additive role to Alzheimer's disease (AD) or there are interactions between the two. Approximately one third of patients that meet clinical and pathologic diagnostic criteria for AD have some degree of vascular pathology (6,7). The impact of CVD on mild cognitive impairment—in which the etiology of the cognitive deficit is generally less clear—is likely to be even greater. Therefore, to identify biologic markers specific to the AD process, it is critical to also identify the extent of concurrent CVD related brain injury that is often clinically silent (8–11), because, at the very least, CVD increases the likelihood of clinical presentation of dementia, for the same level of AD-related pathology.

Population studies, such as the Cardiovascular Health Study or the Rotterdam Scan Study, have shown that brain lesions, especially white matter lesions (WMLs), are

Acad Radiol 2008; 15:300–313

[1] Department of Radiology, 3600 Market Street, Suite 380, University of Pennsylvania, Philadelphia, PA 19104 (Z.L., D.S., A.F.J., E.R.M., R.N.B., C.D.); Department of Biostatistics, Children's Hospital of Philadelphia, Philadelphia, PA (A.F.J.); Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, MD (L.J.L.); Lister Hill National Center for Biomedical Communications, National Library of Medicine/National Institute of Health, Bethesda, MD (D.L.). Received August 23, 2007; accepted October 1, 2007. Supported (in part) by the Intramural Research Program of the NIH, National Institute of Aging contract N01-HC-95178. Image analysis was supported in part by R01-AG-1497. **Address correspondence to:** Z.L. e-mail: Zhiqiang.Lao@uphs.upenn.edu

doi:10.1016/j.acra.2007.10.012

associated with age, clinically silent stroke, higher systolic blood pressure, lower forced expiratory volume in 1 second, hypertension, atrial fibrillation, carotid and peripheral arterioscleroses, impaired cognition, and depression (12–14). Furthermore, it has been shown that stroke patients with a large WML load have an increased risk of hemorrhagic transformation, higher preoperative risk of a disabling or fatal stroke during endarterectomy, or intercerebral hemorrhage during anticoagulation therapy (15). The increased interest in brain lesion research may improve diagnosis and prognosis possibilities for patients with cardiovascular symptoms.

The relationship between diabetes mellitus and cognitive impairment, as well as with increased risk for dementia, has been documented by several clinical studies (16–18). This relationship is mediated by brain pathology, including cerebral infarcts, leukoaraiosis, and tissue atrophy (19–24). Precise measurement of such pathology from magnetic resonance imaging (MRI), and more importantly measurement of evolution of pathology over time, is very important for disease monitoring and evaluation of treatments for diabetes mellitus, such as controlling blood pressure and glycemia. All of these previous studies have employed subjective evaluation of brain lesions, such as the scale of de Groot et al (12), which examined the relationship between periventricular and subcortical WMLs and cognitive functioning in 1,077 elderly subjects randomly sampled from the general population, and are hampered by variations in the anatomical definition of brain abnormalities. Therefore such methods of evaluation of brain abnormality in diabetes mellitus are not easily reproducible, qualitative, and nearly impossible to use without paired reading and high level of quality control in large multisite studies and in longitudinal evaluations that might span several years (25). There is an increasing need for development of highly automated, validated, and reproducible computer-based image analysis tools, especially in large-scale longitudinal studies evaluating brain pathology in diabetes mellitus.

Because brain lesion patterns are very heterogeneous, ranging from punctuate lesions in the deep white matter to large confluent periventricular lesions, the scoring of such lesions is complicated. Moreover, it has been shown that different visual rating scales lead to inconsistencies among studies (26). Commonly used ordinal brain lesion scoring methods, such as used in the Cardiovascular Health Study (27) or the Rotterdam Scan Stud (12,28), offer semiquantitative information on the prevalence of such lesions. Exact spatial information is useful because it has been suggested that specific lesion patterns are associated with specific symptoms (29,30). Moreover, for longitudinal studies aiming to capture relatively small changes in brain lesion patterns, accurate information of lesion volume and location is essential. Expert-based delineation of brain lesions is known to be difficult to reproduce across raters, or even within the same rater, which makes it problematic and that combination of readings from independent reader may be necessary in a longitudinal study.

The use of an automated segmentation method that detects brain lesions with a high sensitivity and specificity could be advantageous. Most of the successful methods in the literature have been developed for the detection of multiple sclerosis (MS) lesions (31–46). In early approaches when multi-modality images are not easily available, features describing normal tissue statistics (either intensity property alone or both intensity and spatial properties) are usually extracted from available modality and then combined with various classifiers, such as: minimum distance classifier, Bayesian classifier, decision tree, for MS lesion segmentation purpose. In (31), Kamber et al built a voxel-wise probability normal tissue (GM [gray matter], WM [white matter], VN [ventricle]) distribution model in Talairach space and then use a decision tree to discriminate MS lesion tissue from normal tissue based on entropy minimization. A similar approach was pursued elsewhere (46), in which spatial statistics of normal brain tissues were first determined from a training set, and deviations from normal variation were flagged as lesions. In Udupa et al (33), major brain tissues (WM, GM, and CSF [cerebrospinal fluid]) were modeled as fuzzy connected regions; potential MS lesions are regarded as isolated islands and were further refined by human judgment.

Most current imaging studies offer the potential to combine multiparametric MRIs (ie, images obtained via different MRI protocols). The advantage of integrating information from multiple sequences is that it can reduce the uncertainty and increase the accuracy of the segmentation. One can categorize most state-of-the-art lesion segmentation algorithms in two main categories: supervised voxel-wise classification and unsupervised clustering. Leemput et al (37) proposed an unsupervised WML segmentation model via setting up a multivariate Gaussian model to describe normal tissue signal distribution, and using it to detect MS lesions as outliers. In supervised methods, a set of images in which the desired segmentation is known (expert manual delineation) is used as a training set to build and fine-tune the segmentation algorithm (35,44). Based on the well-known medical image processing system INSECT, Zijdenbos et al (41) proposed a