Multiple Sclerosis: Identification of Temporal Changes in Brain Lesions with Computer-Assisted Detection Software

M. BILELLO¹, M. ARKUSZEWSKI^{1,2}, P. NUCIFORA¹, I. NASRALLAH¹, E.R. MELHEM³, L. CIRILLO⁴, J. KREJZA^{3,5}

¹Department of Radiology, Division of Neuroradiology, University of Pennsylvania; Philadelphia, Pennsylvania, USA

²Department of Neurology, Medical University of Silesia; Katowice, Poland

³Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine; Baltimore, Maryland, USA

⁴Department of Radiology, University of Bologna; Bologna, Italy

⁵Al-Imam Muhammad Ibn Saud Islamic University; Riyadh, Saudi Arabia

Key words: multiple sclerosis, brain lesions, magnetic resonance, imaging, computer assessment

SUMMARY – Multiple sclerosis (MS) is a chronic disease with a progressing and evolving course. Serial imaging with MRI is the mainstay in monitoring and managing MS patients. In this work we demonstrate the performance of a locally developed computer-assisted detection (CAD) software used to track temporal changes in brain MS lesions. CAD tracks changes in T2-bright MS lesions between two time points on a 3D high-resolution isotropic FLAIR MR sequence of the brain acquired at 3 Tesla. The program consists of an image-processing pipeline, and displays scrollable difference maps used as an aid to the neuroradiologist for assessing lesional change. To assess the value of the software we have compared diagnostic accuracy and duration of interpretation of the CAD-assisted and routine clinical interpretations in 98 randomly chosen, paired MR examinations from 88 patients (68 women, 20 men, mean age 43.5, age range 21-75) with a diagnosis of definite MS. The ground truth was determined by a three-expert panel. In case-wise analysis, CAD interpretation showed higher sensitivity than a clinical report (87% vs 77%, respectively). Lesion-wise analysis demonstrated improved sensitivity of CAD over a routine clinical interpretation of 40%-48%. Mean software-assisted interpretation time was 2.7 min. Our study demonstrates the potential of including CAD software in the workflow of neuroradiology practice for the detection of MS lesional change. Automated quantification of temporal change in MS lesion load may also be used in clinical research, e.g., in drug trials.

Introduction

Multiple Sclerosis (MS) is a chronic inflammatory disease of the central nervous system that affects more than 400,000 people in the US¹. MRI is being used to detect MS lesions and to identify their temporal changes at regular follow-ups. Reliable identification of new or resolving lesions is important in making treatment decisions². However, such an analysis poses a time-consuming challenge to the radiologist because of the high number of MR sequences and images required to fully assess temporal changes and the difficulty to make a

side-by-side comparison of the same sequence at two time points. Small lesion changes can be difficult to identify, especially in heavy pre-existing lesion burdens^{3,4}. Software that can help identify temporal changes reliably in a time efficient and automated manner can have a significant clinical impact especially at centers with large MS clinics.

We have developed a computer-assisted detection (CAD) software tool to automatically co-register and subtract the images of the brain across two time points. CAD works with 3D isotropic high-resolution FLAIR images acquired with a Siemens 3T Verio scanner via a new

protocol optimized for the visualization of MS lesions⁵. To help the neuroradiologist quickly identify and assess any change in MS lesions that may have occurred in the patient's brain, the software highlights the progressed or regressed lesions on the acquired images with a two-color code. We hypothesized that implementation of the CAD software in routine clinical use may aid neuroradiologists in evaluating the temporal change in brain MS lesions and shorten the assessment time. Thus, we compared the accuracy and duration of a software-assisted detection with a routine clinical report against the "reference standard" established by a panel of experts.

Materials and Methods

Patients and Database Acquisition

Our Institution Review Board approved the protocol of the study. Ninety-eight MR paired examinations from 88 patients (68 women, 20 men, mean age 43.5, age range 21-75) with a diagnosis of definite MS were randomly selected from our institution's PACS. These patients had been scanned between January 2009 and April 2010 using our multi-sequence MS protocol on a Siemens Verio 3T scanner (Siemens, Erlangen, Germany). The software system uses only one sequence, the 3D fluid-attenuated inversion recovery (FLAIR) (TR/TE = 5000/395 ms, TI = 1800 ms), which uses a field of view of $250 \times 250 \times 160$ mm, and a matrix of $256 \times 256 \times 160$, resulting in a near isotropic $1 \times 1 \times 1$ mm voxel size.

CAD System

The software runs on a Linux workstation, and the only operation needed prior to running the software is to export the two time point DICOM studies through the network from any PACS workstation to the Linux station. The software has a GUI console that allows the user to load the DICOM studies into the program and scroll through them, displays the patient's name and the study date, and has a graphical cropping tool that is used to remove unnecessary slices from the volumes. After cropping is performed, the user clicks on the "START" button to begin processing. The processing pipeline (Figure 1) starts with rigid registration of the "time 2" (follow-up study) data to the "time 1"

(baseline study), followed by skull removal and inhomogeneity correction. The second phase is image subtraction and generation of candidate new or resolved lesions. The range of lesion intensities for a particular scan is estimated dynamically using linear regression with parameters that had been determined off-line during training. Based on intensity estimation, the software computes thresholds dynamically and applies them to the output of the subtraction operation to eliminate false positives. It then displays the results in the form of a scrollable axial volume with colored new and resolved lesion voxels superimposed on the "time 1" reference scan (Figure 2). Currently, the software takes about ten minutes to process a case on a workstation with an Intel core i7 processor, and 8 GB of RAM. The software has been previously described⁵.

Study Design

The aim of the study was to compare the accuracy of the routine clinical radiological evaluation and software-aided only evaluation in the relation to the "ground truth". At the time of acquisition, an attending neuroradiologist helped by a neuroradiology fellow (a radiologist who has completed at least four years of general radiology training and nine months of subspecialty neuroradiology training) evaluated the current study in comparison to the prior one, using all sequences available in PACS, but without assistance from the CAD software, and generated a clinical report. Independently, the same 98 studies with prior scans were assessed by one neuroradiology fellow with one year of subspecialty training, using only the software output. For the software assessment, the fellow used a form on which he reported new (or enlarging) and resolved (or improving) lesions detected by the software, and specified the location and the side of the brain involved. The image-processing component of the software had been completed prior to this evaluation, so the fellow needed only scroll through the display panel for the assessment. The fellow considered only the flagged (colored) regions on the results display panel that were located within the brain tissue. The reader ignored the perceived false positives located outside the brain, e.g. within skull, and entered only the perceived true positives on the form. The fellow had reasonable exposure to the software prior to the experiment. The duration of software-

Figure 1 Pre-processing pipeline. The software first performs co-registration of the time 2 volume, taking time 1 as the reference. Then skull removal is performed followed by bias field correction. The next step is image subtraction, which generates candidates for new lesions (red) and resolved lesions (green). Finally, a false positive removal step, based on intensity estimates both for normal lesions and normal structures is applied to increase specificity. In this example, the software found a resolved lesion in the right genu of the corpus callosum, and new lesions in the posterior periventricular white matter. Note that the skull is put back before visualization to maintain the visual “feel”.

assisted interpretation was recorded for each case, but this time did not include the image processing time.

To establish the “ground truth”, all 98 pairs of MR studies were assessed by a panel of experts (one neurologist with five years of residency training in neurology and seven years of practice, and two neuroradiologists with two years of fellowship training and three years of practice) by reviewing all PACS images as well as the software-analysis results. First, the panel decided in every individual case, whether there was any change in brain MS lesions in the follow-up study in the relation to the prior scan. Second, the number of new (or growing) and/or resolved (or resolving) lesions was determined. Discrepancies were resolved by consensus.

Data Analysis

Statistical analyses were performed using SYSTAT 12 statistical software (SPSS Science, Chicago, IL), for Windows (Microsoft Corpora-

tion, Redmond, WA, USA). We had three sets of data reflecting detected changes in brain MS lesions: the result of the software-assisted analysis, the routine clinical reports and the expert panel’s assessment; we therefore compared each dataset with the two others.

The differences in the number of cases diagnosed as “changed” between both detection methods and the panel of experts were analyzed with McNemar’s test for symmetry. Differences in the number of individual lesions detected with both methods and by the expert panel were analyzed with Wilcoxon’s signed rank test. For the correction related to multiple comparisons, a probability less than 0.017 was considered significant.

As a measure of diagnostic accuracy of the clinical report and software-assisted analysis, for case-wise analysis we calculated sensitivity, specificity, positive and negative predictive values (PPV and NPV respectively), efficiency and area under receiver-operator characteristic curve (AUC), and for lesion-wise analysis we calculated sensitivity, the ratio of false positives per case and PPV.

Figure 2 Results panel. This panel opens at the end of processing. The small images on the left represent the original time 1 and the co-registered time 2. The larger image on the right contains the 3D “lesions” flagged by the software (red for new and green for resolved) over the time 1 volume as background brain. The scrolling bar acts synchronically on the 3 volumes. In this example, the software found two new lesions, one adjacent to the right trigone, and another in the left peritrigonal region, and a resolved lesion in the right genu of the corpus callosum.

Results

Identification of temporal changes in MS lesions with CAD software was consistent with the expert panel’s assessment (Table 1). The

routine clinical reports, however, identified significantly fewer patients with changed status in the follow-up MR in relation to the “ground truth”. The difference between CAD and the clinical report in identification of temporal le-

Table 1 Results of software-assisted analysis of 88 patients with a definite diagnosis of MS, who underwent 98 follow-up studies with 3D FLAIR MR imaging. Software-assisted detection of cases with worsening or resolving lesions was compared with routine clinical and experts' detection.

Follow-up MR examinations of MS patients	Panel of experts	Clinical report	Software-assisted analysis	Differences (p value)		
	I	II	III	I vs. II	II vs. III	III vs. I
Number of cases with unchanged status	58	66	58	0.011	0.021	1.000
Number of cases with changed status	40	32	40			
• Cases with new or worsening lesions	35	25	32	0.022	0.035	0.317
• Cases with resolved or resolving lesions	18	11	16	0.035	0.166	0.414

Statistics: McNemar's test for symmetry.

Table 2 Measures of accuracy of the routine clinical report versus software-assisted analysis in the detection of changes in brain MS lesions in 98 follow-up 3D FLAIR MR imaging examinations from 88 patients with a definite diagnosis of MS (case-wise).

Follow-up MR examinations of MS patients	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Efficiency (%)	AUC
Any lesion changes						
• Clinical report	77	98	97	86	90	0.88
• Software-assisted analysis	87	91	87	91	90	0.89
New and worsening lesions						
• Clinical report	71	100	100	86	90	0.86
• Software-assisted analysis	83	95	91	91	91	0.89
Resolved and resolving lesions						
• Clinical report	50	97	82	90	89	0.74
• Software-assisted analysis	78	97	87	95	94	0.88

Note: PPV; positive predictive value; NPV; negative predictive value; AUC: area under ROC curve.

sion changes showed a trend toward statistical significance (Table 1). CAD was found to be 10% to 28% more sensitive in detection of new or resolved lesions than the clinical report with a comparable efficiency and only slightly lower specificity (Table 2).

Lesion-wise, the CAD detected a significantly higher number of individual changes in brain MS lesions than the clinical report. However, neither CAD nor clinical reports detected all changes compared to the "ground truth" (Table 3). The CAD was 48% more sensitive than the routine clinical interpretation in the detection of any change, and 40% to 47% higher for new and resolved lesions respectively (Table 4). We observed a small change in the false positives rate with software assistance of -0.01 for new lesions and 0.05 for resolved ones (Table 4).

The mean interpretation time with help of the software was 2.7 min (1-18 min; median = 2 min).

Discussion

The study showed that the CAD software was better able to identify temporal changes in MS lesions using a 3D FLAIR MR sequence in patients with definite MS compared to routine clinical interpretation. The sensitivity and false-positive ratio in identifying patients with new (or growing) lesions was improved with the software assistance. The detection of individual lesions that resolved (or shrank) showed a twofold higher sensitivity, but with a 2.5 times higher false positives ratio (Table 3 and 4). We

Table 3 Results of software-assisted analysis of 98 follow-up 3D FLAIR MR imaging examinations from 88 patients with a definite diagnosis of MS. Software-assisted detection of individual worsening or resolving lesions was compared with routine clinical and experts' detection.

Follow-up MR examinations of MS patients	Panel of experts	Clinical report	Software-assisted analysis	Differences (p value)		
	I	II	III	I vs. II	II vs. III	III vs. I
Number of all lesion changes	268	88	217	<0.001	0.006	0.001
• Number of new or worsening lesions	241	74	187	<0.001	0.009	<0.001
• Number of resolved or resolving lesions	27	14	30	0.033	0.087	0.582

Statistics: Wilcoxon's signed rank test

Table 4 Sensitivity of the routine clinical report and software-assisted imaging analysis in the detection of distinct lesion changes in 98 follow-up 3D FLAIR MR imaging examinations from 88 patients with a definite diagnosis of MS.

Follow-up MR examinations of MS patients	TP	Missed	FP	Sensitivity	FP/case	PPV
Any lesion changes						
• Clinical report	79	189	9	29%	0.09	90%
• Software-assisted analysis	206	62	11	77%	0.11	95%
New and worsening lesions						
• Clinical report	68	173	6	28%	0.06	92%
• Software-assisted analysis	182	59	5	75%	0.05	97%
Resolved and resolving lesions						
• Clinical report	11	16	3	41%	0.03	79%
• Software-assisted analysis	22	5	8	81%	0.08	73%

Note: Unless indicated, data are numbers of changed brain MS lesions; TP: true positive; FP: false positive; PPV: positive predictive value.

believe this result could be improved by optimizing the program parameters further, and with additional training for interpreting neuroradiologists. This study shows the potential of using detection software in both clinical radiology and research. The program is primarily designed to work as a clinical aid for the neuroradiologist. The processing time is short enough for cases to be run in real time within the clinical workflow. The program parameters are tilted towards high sensitivity, but nevertheless have a low rate of false positives, allowing for efficient assessment of temporal changes. The program could also be used in clinical research, e.g., in drug trials, where consistent and reproducible detection of temporal changes with reliable sensitivity and specificity is of great value. The program presented here does not quantify the exact volume of each lesion, or the total change in lesion load, but this capability could be readily introduced.

CAD software exists for a variety of imaging modalities, i.e., mammography, chest radiology, and CT colonography, to improve the efficiency and inter-rater agreement of radiological interpretations. In most cases, CAD software operates in concert with a radiologist in order to generate an interpretation. CAD software can be tested by first establishing a database of studies with "ground truth" interpretations by a panel of experts, followed by testing with readers who interpret the studies with and without the help of the software⁶⁻⁸. These studies have the advantage of producing quantifiable, lesion-specific results.

Automated methods of detection, identification and quantification of MS lesions have been presented by many researchers^{6,9-12}, most of them focusing on enhancing lesions¹³⁻¹⁷. The number of enhancing lesions on a follow-up examination has been considered an important parameter in assessing treatment efficacy or disease activity.

Figure 3 False positives. False positives commonly occur in the anterior temporal lobes because of phase artifacts, and in the brainstem.

From a clinical perspective, non-enhancing new lesions are also of great interest. In fact, as the arsenal of available treatment options is growing, neurologists caring for MS patients are even more interested than ever before in knowing whether new lesions have arisen from the prior scan, regardless of their enhancing status. The program is not designed to detect enhancing lesions, however, we do not believe that those represent a clinical issue. It is relatively easy to detect enhancing lesions “manually”. More difficult is spotting non-enhancing new lesions, especially in a background of heavy, pre-existing lesion load. The capability to detect enhancing lesions could nevertheless be added to the package in a relatively straightforward manner but it would increase the computing time. Most of the similar work done previously has also focused on automatic detection. We depart from this approach, and propose CAD software aimed at helping the neuroradiologist in clinical interpretation. This approach allows for faster computing time, and, we believe, is a more practical and realistic approach that can make a clinical impact. In addition, when compared to similar published work, our study is the largest to have compared software-assisted interpretation with interpretation performed in a real clinical setting. In addition, our CAD software identifies any change in lesion load: both enlarging and shrinking lesions.

One limitation of our study is the lack of confidence index associated with the perceived lesional changes recorded by the fellow on the study form. This missing information would have allowed a free-response receiver operating characteristic (FROC) curve to be generated to show the dependence of sensitivity with the rate of false positives. Our presumption

is that a significant number of false positives would have been associated with a fairly low confidence, contrary to the true positives, and therefore it would have been possible to retain high sensitivity while eliminating low-confidence perceived lesions. The reason for this conjecture is that most of the false positives, as explained above, occurred in specific areas. We believe that a significant number of false positives could have been avoided had the fellow been more experienced with the software’s behavior, especially as it generates a relatively high number of flagged regions – and a relatively high number of false positives – in two specific locations because of phase artifacts: the anterior temporal lobes, and the brainstem (Figure 3). The second area, the brainstem, is a notoriously difficult area for FLAIR^{18,19}. The need to rule out false positives by human readers has the potential to increase reading time. However, a previous study of CT colonography CAD software demonstrated that false positives had limited effect on the interpretations of experienced radiologists²⁰.

A second limitation is the undercount stemming from the occasional failure of clinical reports to state a precise number of new lesions, particularly when the number is greater than a handful. This situation occurred eight times out of 98 cases. As stated previously, if the report mentions “at least n ” new lesions, we assigned $(n + 1)$ to the number of new lesions detected in that case. This explains the drop in sensitivity for clinical interpretation from 0.72 for case-wise performance to 0.26 for lesion-wise performance. Of course, this penalizes the clinical interpretation unfairly, since no explicit demand of accurate count was made to clinical neuroradiologists.

A third limitation is the lack of interobserver agreement analysis, since each case was reviewed only once with software assistance, and naturally, only once in the clinical setting. Fourthly and finally, the software was trained to work on imaging data acquired with a specific MR scanner, and under a specific scanning protocol. However, it would be relatively straightforward to train the program parameters on a different FLAIR sequence, as long as it is 3D, which is probably a requirement for good performance.

Conclusion

We showed the superior accuracy of the software-assisted interpretation over the clinical interpretation in the detection of MS lesion changes over time. Our study demonstrates the potential of including CAD-like software in the workflow of neuroradiology practice, especially in the assessment of change in serial studies. Such software may improve the accuracy of the detection, and shorten the time of interpretation.

References

- 1 Inglese M. Multiple sclerosis: new insights and trends. *Am J Neuroradiol.* 2006; 27: 954-957.
- 2 Simon JH, Li D, Traboulsee A, et al. Standardized MR imaging protocol for multiple sclerosis: Consortium of MS Centers consensus guidelines. *Am J Neuroradiol.* 2006; 27: 455-461.
- 3 Filippi M, Horsfield MA, Bressi S, et al. Intra- and inter-observer agreement of brain MRI lesion volume measurements in multiple sclerosis. A comparison of techniques. *Brain.* 1995; 118: 1593-1600.
- 4 Molyneux PD, Miller DH, Filippi M, et al. Visual analysis of serial T2-weighted MRI in multiple sclerosis: intra- and interobserver reproducibility. *Neuroradiology.* 1999; 41: 882-888.
- 5 Bilello M, Arkuszewski M, Nasrallah I, et al. Multiple Sclerosis Lesions in the Brain: Computer-Assisted assessment of Lesion Load Dynamics on 3D FLAIR MR Images. *Neuroradiol J.* 2012; 2: 412-417.
- 6 Moraal B, Meier DS, Poppe PA, et al. Subtraction MR images in a multiple sclerosis multicenter clinical trial setting. *Radiology.* 2009; 250: 506-514.
- 7 Brown MS, Goldin JG, Rogers S, et al. Computer-aided lung nodule detection in CT: results of large-scale observer test. *Acad Radiol.* 2005; 12: 681-686.
- 8 Taylor SA, Charman SC, Lefere P, et al. CT colonography: investigation of the optimum reader paradigm by using computer-aided detection software. *Radiology.* 2008; 246: 463-471.
- 9 Goldberg-Zimring D, Achiron A, Miron S, et al. Automated detection and characterization of multiple sclerosis lesions in brain MR images. *Magn Reson Imaging.* 1998; 16: 311-318.
- 10 Khayati R, Vafadust M, Towhidkhan F, et al. A novel method for automatic determination of different stages of multiple sclerosis lesions in brain MR FLAIR images. *Comput Med Imaging Graph.* 2008; 32: 124-133.
- 11 Elliott C, Francis SJ, Arnold DL, et al. Bayesian classification of multiple sclerosis lesions in longitudinal MRI using subtraction images. *Med Image Comput Assist Interv.* 2010; 13: 290-297.
- 12 Shah M, Xiao Y, Subbanna N, et al. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. *Med Image Anal.* 2011; 15: 267-282.
- 13 Bedell BJ, Narayana PA. Automatic segmentation of gadolinium-enhanced multiple sclerosis lesions. *Magn Reson Med.* 1998; 39: 935-940.
- 14 Karimaghloo Z, Shah M, Francis S, et al. Automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain MRI using conditional random fields. *IEEE Trans Med Imaging.* 2012; DOI: 10.1109/TMI.2012.2186639.
- 15 Karimaghloo Z, Shah M, Francis SJ, et al. Detection of Gad-enhancing lesions in multiple sclerosis using conditional random fields. *Med Image Comput Assist Interv.* 2010; 13: 41-48.
- 16 Moraal B, Wattjes MP, Geurts JJ, et al. Improved detection of active multiple sclerosis lesions: 3D subtraction imaging. *Radiology.* 2010; 255: 154-163.
- 17 Tan IL, van Schijndel RA, Fazekas F, et al. Image registration and subtraction to detect active T(2) lesions in MS: an interobserver study. *J Neurol.* 2002; 249: 767-773.
- 18 Gawne-Cain ML, O'Riordan JI, Thompson AJ, et al. Multiple sclerosis lesion detection in the brain: a comparison of fast fluid-attenuated inversion recovery and conventional T2-weighted dual spin echo. *Neurology.* 1997; 49: 364-370.
- 19 Wattjes MP, Lutterbey GG, Harzheim M, et al. Imaging of inflammatory lesions at 3.0 Tesla in patients with clinically isolated syndromes suggestive of multiple sclerosis: a comparison of fluid-attenuated inversion recovery with T2 turbo spin-echo. *Eur Radiol.* 2006; 16: 1494-1500.
- 20 Taylor SA, Greenhalgh R, Ilangoan R, et al. CT colonography and computer-aided detection: effect of false-positive results on reader specificity and reading efficiency in a low-prevalence screening population. *Radiology.* 2008; 247: 133-140.

Michał Arkuszewski, MD, PhD
Department of Neurology
Medical University of Silesia
Central University Hospital
Medykow 14, 40-752 Katowice, Poland
Tel.: +48 32 789 46 01
Fax: +48 32 789 45 55
E-mail: michal.arkuszewski@gmail.com