NeuroImage 148 (2017) 77–102

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Longitudinal multiple sclerosis lesion segmentation: Resource and challenge

Aaron Carass[a,b,*,1], Snehashis Roy[c,1], Amod Jog[b,1], Jennifer L. Cuzzocreo[d,1], Elizabeth Magrath[c,1], Adrian Gherman[e,1], Julia Button[d,1], James Nguyen[d,1], Ferran Prados[f,g], Carole H. Sudre[f], Manuel Jorge Cardoso[f,h], Niamh Cawley[g], Olga Ciccarelli[g], Claudia A.M. Wheeler-Kingshott[g], Sébastien Ourselin[f,h], Laurence Catanese[i], Hrishikesh Deshpande[i], Pierre Maurel[i], Olivier Commowick[i], Christian Barillot[i], Xavier Tomas-Fernandez[j,k], Simon K. Warfield[j,k], Suthirth Vaidya[l], Abhijith Chunduru[l], Ramanathan Muthuganapathy[l], Ganapathy Krishnamurthi[l], Andrew Jesson[m], Tal Arbel[m], Oskar Maier[n], Heinz Handels[n], Leonardo O. Iheme[o], Devrim Unay[o], Saurabh Jain[p], Diana M. Sima[p], Dirk Smeets[p], Mohsen Ghafoorian[q], Bram Platel[r], Ariel Birenbaum[s], Hayit Greenspan[t], Pierre-Louis Bazin[u,1], Peter A. Calabresi[d,1], Ciprian M. Crainiceanu[e,1], Lotta M. Ellingsen[a,v,1], Daniel S. Reich[d,w,1], Jerry L. Prince[a,b,1], Dzung L. Pham[c,1]

[a] *Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA*
[b] *Department of Computer Science, The Johns Hopkins University, Baltimore, MD 21218, USA*
[c] *CNRM, The Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD 20892, USA*
[d] *Department of Radiology, The Johns Hopkins School of Medicine, Baltimore, MD 21287, USA*
[e] *Department of Biostatistics, The Johns Hopkins University, Baltimore, MD 21205, USA*
[f] *Translational Imaging Group, CMIC, UCL, NW1 2HE London, UK*
[g] *NMR Research Unit, UCL Institute of Neurology, WC1N 3BG London, UK*
[h] *Dementia Research Centre, UCL Institute of Neurology, WC1N 3BG London, UK*
[i] *VisAGeS: INSERM U746, CNRS UMR6074, INRIA, University of Rennes I, France*
[j] *Computational Radiology Laboratory, Boston Childrens Hospital, Boston, MA 02115, USA*
[k] *Harvard Medical School, Boston, MA 02115, USA*
[l] *Biomedical Imaging Lab, Department of Engineering Design, Indian Institute of Technology, Chennai 600036, India*
[m] *Centre For Intelligent Machines, McGill University, Montréal, QC H3A 0E9, Canada*
[n] *Institute of Medical Informatics, University of Lübeck, 23538 Lübeck, Germany*
[o] *Bahçeşehir University, Faculty of Engineering and Natural Sciences, 34349 Beşiktaş, Turkey*
[p] *icometrix, 3012 Leuven, Belgium*
[q] *Institute for Computing and Information Sciences, Radboud University, 6525 HP Nijmegen, Netherlands*
[r] *Diagnostic Image Analysis Group, Radboud University Medical Center, 6525 GA Nijmegen, Netherlands*
[s] *Department of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel*
[t] *Department of Biomedical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel*
[u] *Department of Neurophysics, Max Planck Institute, 04103 Leipzig, Germany*
[v] *Department of Electrical and Computer Engineering, University of Iceland, 107 Reykjavík, Iceland*
[w] *Translational Neuroradiology Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD 20892, USA*

ARTICLE INFO

Keywords:
Magnetic resonance imaging
Multiple sclerosis

ABSTRACT

In conjunction with the ISBI 2015 conference, we organized a longitudinal lesion segmentation challenge providing training and test data to registered participants. The training data consisted of five subjects with a mean of 4.4 time-points, and test data of fourteen subjects with a mean of 4.4 time-points. All 82 data sets had the white matter lesions associated with multiple sclerosis delineated by two human expert raters. Eleven teams submitted results using state-of-the-art lesion segmentation algorithms to the challenge, with ten teams presenting their results at the conference. We present a quantitative evaluation comparing the consistency of the two raters as well as exploring the performance of the eleven submitted results in addition to three other lesion

* Correspondence to: Department of Electrical and Computer Engineering, The Johns Hopkins University, 105 Barton Hall, 3400 N. Charles St., Baltimore, MD 21218, USA.
E-mail address: aaron_carass@jhu.edu (A. Carass).
[1] These authors co-organized the challenge, all others contributed results.

http://dx.doi.org/10.1016/j.neuroimage.2016.12.064
Received 19 August 2016; Accepted 19 December 2016
Available online 11 January 2017

segmentation algorithms. The challenge presented three unique opportunities: (1) the sharing of a rich data set; (2) collaboration and comparison of the various avenues of research being pursued in the community; and (3) a review and refinement of the evaluation metrics currently in use. We report on the performance of the challenge participants, as well as the construction and evaluation of a consensus delineation. The image data and manual delineations will continue to be available for download, through an evaluation website[2] as a resource for future researchers in the area. This data resource provides a platform to compare existing methods in a fair and consistent manner to each other and multiple manual raters.

1. Introduction

Multiple sclerosis (MS) is a disease of the central nervous system (CNS) that is characterized by inflammation and neuroaxonal degeneration in both gray matter (GM) and white matter (WM) (Compston and Coles, 2008). MS is the most prevalent autoimmune disorder affecting the CNS, with an estimated 2.5 million cases worldwide (World Health Organization, 2008; Confavreux and Vukusic, 2008) and was responsible for approximately 20,000 deaths in 2013 (Global Burden of Disease Study 2013 Mortality and Causes of Death Collaborators, 2015). MS has a relatively young age of onset with an average age of 29.2 years and interquartile onset range of 25.3 and 31.8 years (World Health Organization, 2008). Symptoms of MS include cognitive impairment, vision loss, weakness in limbs, dizziness, and fatigue. The term multiple sclerosis originates from the scars (known as lesions) in the WM of the CNS that are formed by the demyelination process, which can be quantified through magnetic resonance imaging (MRI) of the brain and spinal cord. T_2-weighted (T_2-w) lesions within the WM (or WMLs), so called because of their hyperintense appearance on T_2-w MRI, have become a standard part of the diagnostic criteria (Polman et al., 2011). However, it is a labor intensive and somewhat subjective task to identify and manually delineate or segment WM hyperintensities from normal tissue in MR images. This objective is made more difficult when considering a longitudinal series of data, particularly when each data set at a given time-point for an individual consists of several scan modalities of varying quality (Vrenken et al., 2013). MS frequently involves lesions that may be readily apparent on a scan at one time-point, but not in subsequent time-points (He et al., 2001; Gaitán et al., 2011; Qian et al., 2011). Delineating the scans individually without reference to previous images, may lead to errors in detection of damaged tissue; such as previously lesioned areas that have contracted, undergone remyelination, are no longer inflamed, or a combination thereof. These damaged areas may correlate with disability, although it is as yet unclear precisely how they are related and through what exact mechanism they affect changes in symptoms (Meier et al., 2007; Filippi et al., 2012). Thus there is an apparent need for the automatic detection and segmentation of WMLs in longitudinal CNS scans of MS patients.

Three major subtypes or stages of WMLs can be visualized using MR imaging (Filippi and Grossman, 2002; Wu et al., 2006): (1) gadolinium-enhancing lesions, which demonstrate blood-brain barrier leakage, (2) hypointense T_1-w lesions, also called *black holes* that possess prolonged T_1-w relaxation times, and (3) hyperintense T_2-w lesions, which likely reflect increased water content stemming from inflammation and/or demyelination. These latter lesions are the most prevalent type (Bakshi, 2005) and are hyperintense on proton density weighted (*PD*-w), T_2-w, and fluid attenuated inversion recovery (FLAIR) images. Both enhancing and black hole lesions typically form a subset of T_2-w lesions. Quantification of T_2-w lesion volume and identification of new T_2-w and enhancing lesions in longitudinal data are commonly used to gauge disease severity and monitor therapies, although these metrics have largely been shown to only weakly correlate with clinical disability (Filippi et al., 2014). Pathologically, we can differentiate the different stages of an MS WML as pre-active, active, chronic active, or chronic inactive depending on the demyelination status, adaptive immune response, and microglia behavior. Lesions with normal myelin density and activated microglia are termed pre-active, while sharp bordered demyelination reflects active lesions. Chronic active lesions have a fully demyelinated center and are hypocellular, and chronic inactive lesions have complete demyelination and an absence of any microglia. Current MRI technologies are very sensitive to T_2-w WMLs, however they do not provide any insight about pathological heterogeneity (Jonkman et al., 2015).

Despite this, MRI has gained prominence as an important tool for the clinical diagnosis of MS (Polman et al., 2011), as well as understanding the progression of the disease (Buonanno et al., 1983; Paty, 1988; Filippi et al., 1995; Evans et al., 1997; Collins et al., 2001). A variety of techniques are being used for automated MS lesion segmentation (Anbeek et al., 2004; Brosch et al., 2015, 2016; Deshpande et al., 2015; Dugas-Phocion et al., 2004; Elliott et al., 2013, 2014; Ferrari et al., 2003; Geremia et al., 2010; Havaei et al., 2016; Jain et al., 2015; Jog et al., 2015; Johnston et al., 1996; Kamber et al., 1996; Khayati et al., 2008; Rey et al., 1999, 2002; Roy et al., 2010, 2014b; Schmidt et al., 2012; Shiee et al., 2010; Subbanna et al., 2015; Sudre et al., 2015; Tomas-Fernandez and Warfield, 2011, 2012; Valverde et al., 2017; Weiss et al., 2013; Welti et al., 2001; Xie and Tao, 2011) with several review articles available that describe and evaluate the utility of these methods (García-Lorenzo et al., 2013; Lladó et al., 2012), though semi-automated approaches have also been reported (Udupa et al., 1997; Wu et al., 2006; Zijdenbos et al., 1994). The early work on WML segmentation used the principle of modeling the distributions of intensities of healthy brain tissues and segmenting outliers to those distributions as lesions. An early example of this is Van Leemput et al. (2001), which augmented the outlier detection with contextual information using a Markov random field (MRF). This idea was extended by Aït-Ali et al. (2005) by using an entire time series for a subject, estimating the tissue distributions using an iterative Trimmed Likelihood Estimator (TLE), followed by a segmentation refinement step based on the Mahalanobis distance and prior information from clinical knowledge. Later improvements to the TLE based model include mean shift (García-Lorenzo et al., 2008, 2011) and Hidden Markov chains (Bricq et al., 2008). Other approaches to treating the WM lesions as an outlier class include methods based on support vector machines (SVM) (Ferrari et al., 2003), coupling of local and global intensity models in a Gaussian Mixture Model (GMM) (Tomas-Fernandez and Warfield, 2011, 2012) and using adaptive outlier detection (Ong et al., 2012).

As an alternative to the outlier detection approach other methods create models with lesions as an additional class. Examples of this include: *k*-nearest neighbors (*k*-NN) (Anbeek et al., 2004), a hierarchical Hidden Markov random field (Sajja et al., 2004, 2006); an unsupervised Bayesian lesion classifier with various regions of the brain having different intensity distributions (Harmouche et al., 2006); a Bayesian classifier based on the adaptive mixtures method and an MRF (Khayati et al., 2008); a constrained GMM based on posterior probabilities followed by a level set method for lesion boundary refinement (Freifeld et al., 2009); a fuzzy C-means model with a

[2] The Challenge Evaluation Website is: http://smart-stats-tools.org/lesion-challenge-2015