NeuroImage: Clinical 17 (2018) 918–934

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks

R. Guerrero[a,*,1], C. Qin[a,1], O. Oktay[a], C. Bowles[a], L. Chen[a], R. Joules[b], R. Wolz[b,a], M.C. Valdés-Hernández[c], D.A. Dickie[c], J. Wardlaw[c], D. Rueckert[a]

[a] Department of Computing, Imperial College London, UK
[b] IXICO plc., UK
[c] UK Dementia Research Institute at The University of Edinburgh, Edinburgh Medical School, 47 Little France Crescent, Edinburgh EH16 4TJ, UK

ARTICLE INFO

Keywords:
White matter hyperintensity
Stroke
CNN
Segmentation

ABSTRACT

White matter hyperintensities (WMH) are a feature of sporadic small vessel disease also frequently observed in magnetic resonance images (MRI) of healthy elderly subjects. The accurate assessment of WMH burden is of crucial importance for epidemiological studies to determine association between WMHs, cognitive and clinical data; their causes, and the effects of new treatments in randomized trials. The manual delineation of WMHs is a very tedious, costly and time consuming process, that needs to be carried out by an expert annotator (e.g. a trained image analyst or radiologist). The problem of WMH delineation is further complicated by the fact that other pathological features (i.e. stroke lesions) often also appear as hyperintense regions. Recently, several automated methods aiming to tackle the challenges of WMH segmentation have been proposed. Most of these methods have been specifically developed to segment WMH in MRI but cannot differentiate between WMHs and strokes. Other methods, capable of distinguishing between different pathologies in brain MRI, are not designed with simultaneous WMH and stroke segmentation in mind. Therefore, a task specific, reliable, fully automated method that can segment and differentiate between these two pathological manifestations on MRI has not yet been fully identified. In this work we propose to use a convolutional neural network (CNN) that is able to segment hyperintensities and differentiate between WMHs and stroke lesions. Specifically, we aim to distinguish between WMH pathologies from those caused by stroke lesions due to either cortical, large or small subcortical infarcts. The proposed fully convolutional CNN architecture, called uResNet, that comprised an analysis path, that gradually learns low and high level features, followed by a synthesis path, that gradually combines and upsamples the low and high level features into a class likelihood semantic segmentation. Quantitatively, the proposed CNN architecture is shown to outperform other well established and state-of-the-art algorithms in terms of overlap with manual expert annotations. Clinically, the extracted WMH volumes were found to correlate better with the Fazekas visual rating score than competing methods or the expert-annotated volumes. Additionally, a comparison of the associations found between clinical risk-factors and the WMH volumes generated by the proposed method, was found to be in line with the associations found with the expert-annotated volumes.

1. Introduction

1.1. Clinical motivation

White matter hyperintensities (WMH), referred to in the clinical literature as leukoaraiosis, white matter lesions or white matter disease (Wardlaw et al., 2013), are a characteristic of small vessel disease (Wardlaw and Pantoni, 2014) commonly observed in elderly subjects on fluid-attenuated inversion recovery (FLAIR) magnetic resonance (MR) images, which, as the name suggests, they appear as hyperintense regions. Moreover, stroke lesions of cortical, large subcortical (striatocapsular) or small subcortical infarct origin can also often appear as hyperintense regions in FLAIR MR images and can coexist and coalesce with WMHs. The accurate assessment of WMH burden is of crucial importance for epidemiological studies to determine associations between WMHs, cognitive and clinical data. Similarly, it would help

* Corresponding author.
E-mail address: reg09@imperial.ac.uk (R. Guerrero).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.nicl.2017.12.022
Received 9 June 2017; Received in revised form 11 December 2017; Accepted 16 December 2017
Available online 20 December 2017

discover their causes, and the effects of new treatments in randomized trials. In the assessment of WMH burden it is important to exclude stroke lesions as they have different underlying pathologies, and failure to account for this may have important implications for the design and sample size calculations of observational studies and randomized trials using WMH quantitative measures, WMH progression or brain atrophy as outcome measures (Wang et al., 2012). One of the most widely used metrics to assess WMH burden and severity is the Fazekas visual rating scale (i.e. score) (Fazekas et al., 1987). In this scale, a radiologist visually rates deep white matter and peri-ventricular areas of a MR scan into four possible categories each depending on the size, location and confluence of lesions. The combination of both deep white matter and peri-ventricular ratings yields a combined zero to six scale. In the vast majority of clinical trials and in general clinical practice visual rating scores are used (such as the Fazekas score). WMHs are very variable in size, appearance and location, and therefore the categorical nature of the Fazekas scale has limitations for studying their progression in relation with other clinical parameters. WMH volume has been demonstrated to correlate with severity of symptoms, progression of disability and clinical outcome (Bendfeldt et al., 2010; Chard et al., 2002; Löuvbld et al., 1997). Accordingly, determining WMH volume has been of interest in clinical research as well as in clinical trials on disease-modifying drugs (Löuvbld et al., 1997; van Gijn, 1998; Brott et al., 1989; SPIRIT, 1997). For some studies, lesions have been traced manually (sometimes with the help of semi-automated tools for contour detection) slice by slice. This process can easily become prohibitively expensive for even moderately large datasets. It is therefore obvious that the accurate automatic quantification of WMH volume would be highly desirable, as this will undoubtedly lead to savings in both time and cost. Recently, several automated and semi-automated methods have been put forward to address the coarseness of the visual assessments (e.g. Fazekas score), as well as the dependence on highly qualified experts to perform such assessments. These methods can be broadly classified into *supervised*, when a training "gold-standard" is available (Van Nguyen et al., 2015; Ghafoorian et al., 2016), i.e. when one or more human experts have annotated data, *unsupervised*, when no such gold-standard exists (Ye et al., 2013; Cardoso et al., 2015; Bowles et al., 2016), and *semi-supervised*, when only a small portion of available data has been expertly annotated (Kawata et al., 2010; Qin et al., 2016). However, despite the number of proposed methods, no automated solution is currently widely used in clinical practice and only a few of them are publicly available (Shiee et al., 2010a; Damangir et al., 2012; Schmidt et al., 2012). This is partly because lesion load, as defined in most previously proposed automatic WMH segmentation algorithms, does not take into account the contribution of strokes lesion, as these methods are generally unable to differentiate between these two types of lesions.

1.2. Related work

In the following we review existing methods and challenges that are related to our work, especially on Multiple sclerosis (MS), WMH and stroke lesion segmentation in MR imaging. Additionally, some more general CNN segmentation approaches that share architectural similarities with the method we propose here are also reviewed in this section. Over the last few years, there has been an increased amount of research going on in these areas (García-Lorenzo et al., 2013; Caligiuri et al., 2015; Maier et al., 2017; Rekik et al., 2012). Although some of the methods mentioned here were proposed for segmenting different pathologies rather than the ones we explore in this work, they can in fact be applied to different tasks. As mentioned before, these methods can be broadly classified into *unsupervised*, *semi-automatic*, *semi-supervised* and *supervised*, depending on the amount of expertly annotated data available.

1.2.1. Unsupervised segmentation

Unsupervised segmentation methods do not require labeled data to perform the segmentation. Most of these approaches employ clustering methods based on intensity information or some anatomical knowledge to group similar voxels into clusters, such as fuzzy C-means methods (Gibson et al., 2010), EM-based algorithms (Dugas-Phocion et al., 2004; Forbes et al., 2010; Kikinis et al., 1999) and Gaussian mixture models (Freifeld et al., 2009; Khayati et al., 2008). Some of the probabilistic generative models of the lesion formation for stroke lesion segmentation were also designed, such as Forbes et al. (2010); Derntl et al. (2015). Forbes et al. (2010) proposed a Bayesian multi-sequence Markov model for fusing multiple MR sequences to robustly and accurately segment brain lesions. Derntl et al. (2015) proposed to combine standard atlas-based segmentation with a stroke lesion occurrence atlas, in a patient-specific iterative procedure. Some authors have also proposed to model lesions as outliers to normal tissues. Van Leemput et al. (2001) employed a weighted EM framework in which voxels far from the model were weighted less in the estimation and considered potential lesions. Weiss et al. (2013) proposed to use dictionary learning to learn a sparse representation from pathology free brain T1-weighted MR scans and then applied this dictionary to sparsely reconstruct brain MR images that contain pathologies, where the lesions were identified using the reconstruction error. Additionally, several works have also focused on exploiting the fact that WMHs are best observed in FLAIR MR images, while being difficult to identify in T1-weighted MR images. Some of these methods rely on generating a synthetic FLAIR image based on observed T1-weighted MR image using random forests (Ye et al., 2013), generative mixture-models (Cardoso et al., 2015), support vector regression (SVR) (Bowles et al., 2016) or convolutional neural networks (CNN) (Van Nguyen et al., 2015). Both synthetic (healthy looking) and real FLAIR (with pathologies) images are then compared to detect any abnormalities. Other method like lesion-TOADS (Shiee et al., 2010b) combines atlas segmentation with statistical intensity modeling to simultaneously segment major brain structures as well as lesions. The lesion growth algorithm (LGA), proposed by Schmidt et al. (2012) and part of SPM's LST toolbox (www.statistical-modelling.de/lst.html), constructs a conservative lesion belief map with a pre-chosen threshold (κ), followed by the initial map being grown along voxels that appear hyperintense in the FLAIR image. In essence, LGA is a self-seeded algorithm and it tends to have difficulties detecting subtle WMHs. An important drawback of all these methods is that they are in fact abnormality detection algorithms and not specifically WMH segmentation methods, hence in principle they detect any pathology, whether or not is a WMH-related pathology.

1.2.2. Semi-automatic and semi-supervised segmentation

Several semi-automatic algorithms proposed in the literature for WMH segmentation rely on region growing techniques that require initial seed points to be placed by an operator. Kawata et al. (2010) introduced a region growing method for adaptive selection of segmentation by using a SVM with image features extracted from initially identified WMH candidates. Itti et al. (2001) proposed another region growing algorithm that extracts WMHs by propagating seed points into neighboring voxels whose intensity is above an optimized threshold. The process iterates until convergence, i.e. all voxels above the threshold that are connected to the initial seed point had been annotated. Aside from the drawback of requiring per image expert inputs, semi-automatic methods have the additional potential drawback that seeds points could easily be selected in obvious regions, while the biggest challenge of WMH segmentation can arguably be found in the more confusing border regions. Qin et al. (2016) proposed a semi-supervised algorithm that optimizes a kernel based max-margin objective function which aims to maximize the margin averaged over inliers and outliers while exploiting a limited amount of available labeled data. Although theoretically interesting and well motivated, the problem of transferring useful knowledge from unlabeled data to a task defined by