Segmentation of Age-Related White-Matter Changes in a large-scale, multi-centre study.

T. Dyrby[1], E. Rostrup[1], E. C. van Straaten[2], F. Barkhof[3], S. Ropele[4], L. K. Hansen[5], G. Waldemar[6]

[1]MR-department, Copenhagen University Hospital, Hvidovre, Copenhagen, Denmark, [2]Department of Neurology and Alzheimer Center, VU Medical Center, Amsterdam, Netherlands, [3]Department of Radiology and Image Analysis Center, VU Medical Center, Amsterdam, Netherlands, [4]Karl Franzens University, Graz, Austria, [5]Danish Technical University, Copenhagen, Denmark, [6]Copenhagen Univ. Hospital, Rigshospitalet, Copenhagen, Denmark

Introduction

Age-related white-matter changes (ARWMC) are often seen on brain MRI of elderly people and are known to be associated with a number of vascular risk factors. The possible direct association with cognitive and motor deficits is also increasingly recognised. ARWMC is often quantified through visual rating scales, but quantitative volumetric measurements have recently been shown to provide some increase in sensitivity towards clinical correlation, and also opens the possibility of voxel based studies. However, manual volumetric procedures are highly time consuming, and an automated method would therefore be useful both in research and for clinical evaluation of elderly patients with cognitive deficits. In the present study the robustness of an artificial neural network (ANN) model is tested, and a direct comparison between automated and manual volumetry is performed in a large multi-centre study of elderly subjects. The paper describes the success and methodology in dealing with the trade-off between the large number of subjects, giving statistical power, and the large variability within centres and subjects.

Method

MR scans were acquired on a population consisting of 639 non-demented elderly subjects (65 – 84 years) at 11 different centres as part of the Leukoraiosis And Disability (LADIS) study, a European 5th Framework project investigating the impact of ARWMC on the transition to disability. ARWMC ROIs were hand-drawn by a single expert, and these were used for training and testing of ROIs determined automatically by an artificial neural network (ANN) model.

Three of the standardized MR sequences within the LADIS protocol were used for segmentation (MPRAGE, T2SE and FLAIR). The specification of the sequences was somewhat flexible in order to meet the abilities of the various MR scanner types (all 1.5T) used at each centre. ***T2w fast spin echo***: TR/TE 4000-6000/100-120 ms, NEX=2. ***Fluid-attenuated inversion recovery (FLAIR)***: TR/TE = 6000-10000/100-140ms, TI 2000-2400ms, slice gap 0.5mm, 19-28 axial slices, NEX=2. For the 2D sequences resolution was 1 x 1 x 5 – 7.5 mm^3, slice gap 0.5mm, 19-28 axial slices. ***T1w MPRAGE***: TR/TE = 10-25/4-7 ms, flip-angle 15-30°, $1mm^3$ isotropic voxels.

In order to achieve a minimum quality level for this study, the following inclusion criteria were used per subject: availability of all three modalities, MPRAGE GM:WM contrast-to-noise ratio (CNR) > 70dB (Dyrby et al, 2005), > 10 subjects per centre. This reduced the dataset to 224 subjects from 6 centres, distributed thus [40 45 45 37 42 15]. Pre-processing of the data included: intra-subject intensity correction of FLAIR and T2 images, co-registration (6 degrees of freedom) and re-slicing to the FLAIR image space with the resolution of the MPRAGE, RF inhomogeneity correction via N3, and automatic generation of a brain mask via SPM2. Finally, inter-subject intensity standardization was applied for each modality to suppress centre-specific scanner variations[2].

The ANN is a supervised modelling method, implemented with an iterative train-and-test protocol, and executed on a per-pixel basis. We used a fully-connected feed-forward ANN with a single layer of hidden units and 30 inputs (3 modalities, each with their respective 3x3 neighbourhood). Further spatial context was applied using a relative location measure of x, y, and z within an automatically-generated brain mask. The output comprised of four probability maps – GM, WM, CSF and ARWMC, but only the latter is discussed here.

Examples for the modelling set (min. 8000 voxels per tissue class) were carefully drawn from 6 subjects (4 centres) to ensure inclusion of information regarding the spatial and partial-volume effects. These examples were split 70:30 into training and testing sets. The ANN was initiated with 80 hidden units, trained for 2500 iterations and then pruned in order to achieve generalizability. The ARWMC probability threshold was optimised with respect to overlap between ANN and manual ROI's, and a value of 90% was found to be optimal.

Results

The figures show the results of the correlation between the volumes of the hand-drawn ROIs and those determined by the ANN.

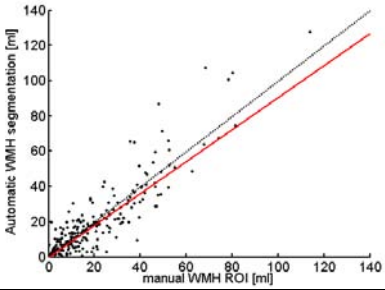

Figure 1: The ANN ROI volume (y-axis) vs. manual ROI volume (x-axis) from 224 subjects (dots), black: identity line, red line: linear regression found as volROI=0.9*volANN

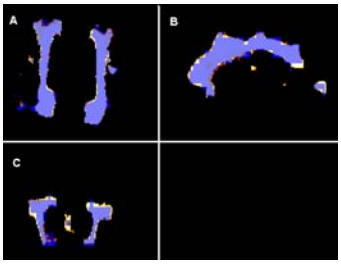

Figure 2: Tri-planar overlay image of (blue) ROI re-sliced to a higher resolution and (white) the smooth ANN segmentation result (probability>90%). A) axial- B) sagittal- and C) coronal view.

Correlation coefficients from each of the 6 centres (ANN ROI volume to manual ROI volume) were: [0.97 0.97 0.87 0.88 0.73 0.98]. The correlation of the combined set was 0.89. Centres 1 and 2 had two subjects in the training set, which may explain their better performance. Overall, there is a weak tendency for the ANN segmentation to be more conservative. The low correlation observed in Centre 5 is mainly caused by deviating FLAIR quality. In other cases a systematic bias was present in some anatomical regions (e.g. septum pellucidum, where lesion were not drawn manually) or regarding lacunar infarcts.

Conclusion

In this multi-centre investigation of ARWMC segmentation a high correlation of 0.89 (ranging to 0.97 for some centres) between manual and automated segmentation was reached. The cause of discrepancies fell into three categories: 1. higher resolution in ANN than manual segmentation, 2. systematic observer bias, 3. technical causes such as contrast variability. Improvements may be gained from representing all centres in the training set, or adding anatomical expert knowledge. The results demonstrate the generalisability of ANN methods, and underline the importance of rigorous standardisation of MRI quality in multi-centre studies.

References

1. Bishop, CM, Neural Networks for Pattern Recognition, Oxford University Press, 1995.
2. Dyrby, T. and Liptrot, M., Standardizing MR image intensity in multi-centre studies, ISMRM, Miami USA, 2005