Ponnada Narayana, PhD; DABR
Diagnostic and Interventional Imaging, Vice-Chair for Research and the Director of Magnetic Resonance Research

Effect of Training Size on Deep-Learning-Based Neural Tissue Segmentation of MRI in Multiple Sclerosis

Dr. Narayana is a Professor of Diagnostic and Interventional Imaging, Vice-Chair for Research and the Director of Magnetic Resonance Research at McGovern Medical School at The University of Texas Health Science Center at Houston (UTHealth). He also holds an Endowed Chair in Biomedical Engineering at UTHealth.

He served on numerous NIH Study Sections. He served on the Editorial Board of a number of major journals.

Dr. Narayana's major research interests include Quantitative Magnetic Resonance of Central Nervous System, Development of Advanced Magnetic Resonance Techniques, and Image Processing with an emphasis on automatic analysis. His current research interests include patient-adaptive, real time MRI, and application of AI to image analysis.

Dr. Narayana directs a state-of-the-art 3T whole body MRI scanner and 7T animal-dedicated MRI scanner.

He has been continuously funded by NIH for the last 35+ years. He directs the MRI Analysis Center that is involved in analyzing multi-center MRI data.

He served as thesis advisor for more than 18 MS, PhD, and MD/PhD students. He trained more than 20 post-doctoral fellows and mentored 10 junior faculty members. He has authored/coauthored 300 publications in peer-reviewed journals and numerous book chapters. He is a Diplomate of the American Board of Radiology in Radiological Physics.

Abstract: **Purpose**: To determine the required training size for a desired accuracy in brain MRI segmentation in multiple sclerosis (MS) using deep learning (DL).

Methods: Magnetic resonance images on 1008 clinically definite relapsing-remitting MS patients who participated in a multi-center, double-blinded, phase III clinical trial

were included in this study. Images were acquired on 1.5T and 3T scanners manufactured by GE, Philips, and Siemens. Images were acquired using dual turbo spin echo, FLAIR, and T1-weighted turbo spin echo sequences. Images segmented using an automated analysis pipeline1,2 and validated by two neuroimaging experts served as the ground truth. A DL model, based on a fully convolutional neural network3, was trained separately using 16 different training sizes. The segmentation accuracy as a function of the training size was determined. These data were fitted to the learning curve4, based on an inverse power law, for estimating the required training size for desired accuracy. The performance of the network was evaluated by calculating the Dice similarity coefficient (DSC), and lesion true-positive and false-positive rates.

Results: The DSC for lesions showed much stronger dependency on the sample size than gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). When the training size was increased from 10 to 800 the DSC values varied from 0.00 to 0.86 ± 0.016 for T2 lesions, 0.87 ± 009 to 0.94 ± 0.004 for GM, 0.86 ± 0.08 to 0.94 ± 0.005 for WM, and 0.91 ± 0.009 to 0.96 ± 0.003 for CSF. The training size needed for a given accuracy strongly depended on the lesion volume. For lesion volumes of 500 µl, an accuracy of 0.8 was achieved with a small training set of 30. In contrast, even with a training size of 800, the accuracy was only 0.5 for lesion size smaller than 70 µl.

Conclusion: Excellent segmentation was achieved with a training size as small as 10 image volumes for GM, WM, and CSF. In contrast, a training size of at least 50 image volumes was necessary for adequate lesion segmentation. The inverse power law dependence allows prediction of the minimum training size needed for a given DSC target.

References:

1. Sajja BR, Datta S, He R, et al. Unified approach for multiple sclerosis lesion segmentation on brain MRI. Ann Biomed Eng. 2006;34(1):142–151.
2. Datta S, Narayana PA. A comprehensive approach to the segmentation of multichannel three-dimensional MR brain images in multiple sclerosis. NeuroImage Clin. Elsevier; 2013;2(1):184–196.
3. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444.
4. Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH. Predicting sample size required for classification performance. BMC Med Inform Decis Mak. 2012