

QUANTIFYING WHITE MATTER HYPERINTENSITY AND BRAIN VOLUMES IN HETEROGENEOUS CLINICAL AND LOW-FIELD PORTABLE MRI

Pablo Laso^{1,2} and Stefano Cerri^{1,3} and Annabel Sorby-Adams¹ Jennifer Guo¹ Farrah Mateen¹
Philipp Goebel⁴ Jiaming Wu⁴ Peirong Liu¹ Hongwei Bran Li¹ Sean I. Young¹ Benjamin Billot⁵
Oula Puonti^{1,6} Gordon Sze⁷ Sam Payabavash⁷ Adam DeHavenon⁷ Kevin N. Sheth⁷
Matthew S. Rosen¹ John Kirsch¹ Nicola Strisciuglio² Jelmer M. Wolterink²
Arman Eshaghi⁴ Frederik Barkhof⁴ W. Taylor Kimberly¹ Juan Eugenio Iglesias^{1,4,5}

¹ Massachusetts General Hospital, Harvard Medical School, USA

² Electrical Engineering, Mathematics and Computer Science, University of Twente, The Netherlands

³ Pioneer Centre for AI, University of Copenhagen, Denmark

⁴ Centre for Medical Image Computing, University College London, UK

⁵ Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA

⁶ Danish Research Centre for Magnetic Resonance, Denmark ⁷ Yale School of Medicine, USA

ABSTRACT

Brain atrophy and white matter hyperintensity (WMH) are critical neuroimaging features for ascertaining brain injury in cerebrovascular disease and multiple sclerosis. Automated segmentation and quantification is desirable but existing methods require high-resolution MRI with good signal-to-noise ratio (SNR). This precludes application to clinical and low-field portable MRI (pMRI) scans, thus hampering large-scale tracking of atrophy and WMH progression, especially in underserved areas where pMRI has huge potential. Here we present a method that segments white matter hyperintensity and 36 brain regions from scans of any resolution and contrast (including pMRI) without retraining. We show results on eight public datasets and on a private dataset with paired high- and low-field scans (3T and 64mT), where we attain strong correlation between the WMH ($\rho=.85$) and hippocampal volumes ($\rho=.89$) estimated at both fields. Our method is publicly available as part of FreeSurfer, at: <http://surfer.nmr.mgh.harvard.edu/fswiki/WMH-SynthSeg>.

1. INTRODUCTION

White matter hyperintensity (WMH) on magnetic resonance imaging of the human brain is associated with stroke, cognitive decline, and cardiovascular disease. WMH is frequently detected in brain MRI scans in the general population with chronic disease such as hypertension. A recent observational study was performed in a safety net emergency setting evaluating adult patients with a vascular risk factor who were being evaluated for a non-stroke complaint. In this cohort, more than half of the subjects had WMH identified on portable, low field MRI [1]. In addition, WMH is a hallmark of multiple

sclerosis (MS), a disease that creates a demyelination process that may lead to disability [2]. The MS disease process is correlated with other neurodegeneration, leading to abnormally high atrophy rates in different brain regions [3]. Closer monitoring of WMH and atrophy is thus desirable at a larger scale.

Inexpensive portable MRI (pMRI) technology is becoming increasingly available for imaging WMH in the community at large scale. For example, the low-field (64mT) Swoop system (Hyperfine Inc) produces images that agree well with high-field counterparts when WMH are scored by a radiologist [1]. A crucial component of large-scale deployment is automated segmentation and quantification of WMH and brain regions, as manual identification and tracing of regions of interest (ROIs) in 3D is impractical and irreproducible.

Quantification of WMH and brain anatomy (including atrophy) is also very desirable in clinical MRI. As opposed to a research MRI, which is typically isotropic, clinical scans often comprise fewer slices acquired in 2D. These take less time for clinical review and are less susceptible to motion artifacts. Precise quantitative analysis of these scans would allow closer tracking of atrophy and WMH progression.

A large array of methods exist for segmenting brain anatomy and WMH. Representative classical methods include: FreeSurfer [4] and FSL [5] for brain ROIs; LST [6] and BIANCA [7] for WMH; or SAMSEG [8, 9], which segments both. Machine learning techniques, often using convolutional neural networks (CNNs), include: QuickNat [10] or FastSurfer [11], for brain ROIs; or [12, 13] for WMH. These methods are designed for conventional high-field MRI (1.5-3T), and often have requirements in terms of resolution (typically 1mm isotropic), pulse sequence (often T1-weighted for anatomy, FLAIR for WMH), or both. Therefore, they strug-

gle with the huge variability in orientation (axial, coronal, sagittal), resolution, and contrast of clinical MRI in real scenarios. This problem is exacerbated in pMRI, where the low field imposes limitations in signal-to-noise ratio (SNR) that are compensated with large voxel sizes, and where the geometry of the scanner often leads to severe signal loss away from its center. While domain adaptation [14] can mitigate these problems to some extent, a CNN that can handle any MRI contrast and resolution without retraining is highly desirable.

Here we present WMH-SynthSeg, a CNN that segments WMH and brain anatomy from scans of any resolution and contrast, including low-field pMRI. WMH-SynthSeg builds on our previous work on domain randomization [15, 16] to achieve such agnosticity. Compared with our previous method for simultaneous segmentation of WMH and anatomy [17], WMH-SynthSeg: (i) does not require retraining; (ii) uses a specific WMH model and a composite loss to improve sensitivity and specificity; (iii) adapts to low-field MRI; and (iv) uses multi-task learning for enhanced robustness. We show that, as a result, WMH-SynthSeg can robustly segment WMH and anatomy from clinical and pMRI.

2. METHODS

2.1. Synthetic training data

WMH-SynthSeg relies on a synthetic MRI generator similar to [16], which requires a training dataset with N 1mm isotropic T1-weighted (T1w) scans $\{I_n\}$ and corresponding 3D segmentations $\{S_n\}$; these are defined on the same 1mm isotropic grid and include labels for brain ROIs and WMH.

At every iteration during training: (i) a random pair (I_n, S_n) is selected; (ii) (I_n, S_n) are augmented non-linear deformation; (iii) a Gaussian mixture model conditioned on the labels is sampled independently at every voxel, with means and variances that are randomly sampled from uniform distributions – except for the WMH class (details below); (iv) the Gaussian image is corrupted by a random smooth bias field; (v) random orientation and resolution are simulated (via smoothing) to synthesize a lower resolution scan; and (vi) the low-resolution scan is upsampled to the original 1mm isotropic grid. This process generates: the upsampled synthetic scan I^{syn} , deformed segmentation S , deformed real image I , and bias field B . All these are defined on the original 1mm grid (see [16] for examples of synthetic images).

The generator has 4 key improvements compared with [16]: (i) The mean intensity of the WMH class is not distributed across the whole range 0-255. Instead, we simulate WMH in T2-like sequences (including FLAIR) and WM hypointensity in T1w-like sequences. This is done as follows: when the white matter (WM) mean is high (over 128), we constrain the WMH mean to be lower than the WM mean (T1w-like). Conversely, when the WM mean is below 128, we constrain the WMH mean to be greater than the WM mean (T2-like). (ii) The standard deviation of the noise (Gaussian variances)

and bias field strength is twice as large as in [16], to accommodate the lower SNR and stronger signal losses of pMRI.

(iii) The generator produces not only I^{syn} but also a deformed image I and a bias field B that will be used as regression targets by the CNN in a multi-task learning setting. This boosts the robustness of the CNN as shown in the experiments.

(iv) The sampling scheme for the random resolution covers a wider spectrum of acquisitions. 25% of the time, we generate 1mm isotropic images, to support high-resolution scans. Another 25% we generate clinical scans of random orientation with 1mm in-plane resolution and random slice spacing between 2.5mm and 8.5mm. 25% of the scans mimic the resolution of the stock sequences that the Hyperfine Swoop ships with (axial with ~ 1.5 mm in plane and 5mm spacing). The final 25% simulates more isotropic scans acquired at low field, with random voxel sizes between 2-5 mm in every direction.

2.2. Model architecture and training

WMH-SynthSeg uses a 3D U-net [18] with five levels, 64 feature maps per level, and group normalization [19]. Each level has two convolutions (kernel size: 3x3x3) followed by ReLU activations. The final layer has $L+2$ channels: the first L correspond to the labels and are fed to a softmax layer to produce soft segmentations; the last two correspond to the predicted bias field and high-resolution T1w intensities.

Training uses the Adam optimizer to minimize a loss function consisting of four terms with equal weight: the cross-entropy and Dice scores between the predicted and ground truth segmentations; the average ℓ_1 error of the predicted T1w intensities (normalized such that the median intensity of the WM is 1); and the ℓ_1 error of the predicted bias field (in logarithmic scale):

$$\mathcal{L} = CE(S, \hat{S}) - AvDice(S, \hat{S}) + |I - \hat{I}| + |\log B - \log \hat{B}|,$$

where \hat{S} , \hat{I} , and \hat{B} are the predictions for the segmentation, T1w intensities, and bias field, respectively.

We note that, while training with Dice may be more common in segmentation, combining it with cross-entropy has two advantages. First, it provides a more informative gradient in the first iterations of training, when the gradient of the Dice loss is rather flat. And second, it explicitly penalizes false positives in scans without WMH – in which the Dice score for the WMH is zero independently of the prediction. In addition, including I and B in the loss increases the robustness of the method, as shown by the experiments below.

At test time, the input scan is resampled to 1mm isotropic resolution and fed to the CNN. Test-time augmentation is performed by left-right flipping the image, flipping the output back, and averaging with the non-flipped version. The first L channels of the output yield the final segmentation; the outputs corresponding to the bias field and the T1w intensities are a potentially useful by-product, but are disregarded here.

We train the CNN with PyTorch using 160^3 voxel patches. The validation loss typically converges in $\sim 10^5$ iterations.

3. EXPERIMENTS AND RESULTS

3.1. Datasets

We used nine different datasets in our experiments, some just for training (“Tr”), some for testing (“Te”), and some for both using cross validation (“Tr/Te”).

HCP [20] (Tr): 897 1mm isotropic scans of young subjects from the Human Connectome Project. We used FreeSurfer to automatically segment the anatomy into 36 ROIs.

ADNI [21] (Tr): 1148 1mm isotropic scans from the ADNI. We used FreeSurfer to segment the anatomy and WMH.

GE3T (Tr/Te): 20 cases with 1mm isotropic T1w and 1x1x3mm axial FLAIRs. This is a subset of the WMH segmentation challenge [22]. We combined the automated FreeSurfer segmentation of the T1w with the manual delineations available for the FLAIRs into a single ground truth segmentation.

Singapore (Tr/Te): another subset of the challenge with 20 cases from a separate site (same MRI acquisitions and labels).

Utrecht (Tr/Te): another subset with 20 cases from a third site.

ISBI [23] (Tr/Te): 15 1mm isotropic T1w scans (segmented with FreeSurfer) and 1x1x2mm axial FLAIRs with manually traced WMH (merged with the anatomy into one label map).

FLI-IAM [24] (Tr/Te): T1w and FLAIR scans from 15 cases with varying resolution but all close to 1mm isotropic. Consensus WMH tracings are available from 7 raters, which we merged with the FreeSurfer segmentations of the T1w scans.

ADHD [25] (Te): 20 1mm isotropic T1w scans from typically developing control children and adolescents and no WMH.

MGH (Te): 12 MS patients from our hospital (MGH) with 1mm T1w and FLAIR, as well as pMRI axial T1w and FLAIR (in-plane resolution: 1.6-1.8mm; slice spacing: 5-6mm).

3.2. Competing methods

We compare our method with: (i) SAMSEG [8, 9], which is a Bayesian method that is adaptive to MRI contrast, and is (to our best knowledge) the only existing method that can readily segment anatomy and WMH from scans acquired with any pulse sequence; and (ii) LST-LPA [26], which yields great performance on FLAIR acquisitions but does not work on other MRI contrasts. We also consider two ablations of our method to assess the importance of its components: a version with just Dice in the loss (similar to [17] but with domain randomization), and a version without the prior on the mean of the WMH class. We note that LST and SAMSEG operate at the native resolution of the scan, whereas WMH-SynthSeg always produces a 1mm isotropic segmentation.

3.3. Experimental setup

We analyze the performance of our proposed method WMH-SynthSeg with three different experiments. The first experiment assesses the performance of the method directly with Dice scores. We first trained WMH-SynthSeg using GE3T

and Singapore (using 15 scans for validation), and tested on ISBI, FLI-IAM, and Utrecht. We then reversed the roles to obtain Dice scores for GE3T and Singapore. We note that HCP and ADNI were also part of the training dataset in both folds. We note that training inputs are all synthetic and that the real images are only used as regression targets.

The second experiment assesses false positive rates (FPR) using young healthy controls from the ADHD dataset. Since WMH is not expected in these scans, we can use the estimated WMH loads as a proxy for FPR. The model in this experiment is trained with all the datasets from the first experiment.

The third experiment assesses the ability of the methods to segment pMRI data, using the same model as in the second experiment. We used the FreeSurfer segmentations of the high-field 1mm T1w scans as ground truth for the anatomy, and the LST segmentations of the high-field 1mm FLAIRs as ground truth for the WMH. Since accurate co-registration of low- and high-field scans is difficult due to nonlinear geometric distortions, we use the correlation between the ground truth and estimated ROI volumes to assess performance.

3.4. Results

Table 1 shows the average Dice across the high-field datasets in the first experiment, for the WMH and for 23 representative brain ROIs: brainstem, and left/right cortex, WM, hippocampus, amygdala, thalamus, caudate, pallidum, putamen, accumbens, and cerebellum cortex and WM (we exclude less reliable ROIs, e.g., accumbens). WMH-SynthSeg outperforms the competing methods across the board. The ablations show that cross-entropy and multi-task learning have a moderate positive impact on the segmentation of anatomy, whereas the prior on mean of WMH component greatly boosts the performance of the WMH segmentation. In absolute terms, our new method yields competitive Dice scores for anatomy (Dice=.85 for isotropic T1w) and WMH (Dice=.62 in FLAIR, higher than SAMSEG and LST). We also highlight its capability to produce useful WMH segmentations from the T1w, with Dice scores as high as those of the competing methods in FLAIR.

Method	T1w		FLAIR	
	Anat	WMH	Anat	WMH
LST (LPA)	N/A	N/A	N/A	0.57
SAMSEG	0.81	0.46	0.72	0.56
WMH-SynthSeg (NoWMH-noCE-noMTL)	0.83	0.47	0.76	0.53
WMH-SynthSeg (NoWMH)	0.85	0.47	0.78	0.54
WMH-SynthSeg (full)	0.85	0.55	0.79	0.62

Table 1: Average Dice scores for anatomy (averaged over 23 ROIs) and WMH, on high-field T1w and FLAIR scans. NoWMH-noCE-noMTL is the ablation without prior on the WMH mean, cross-entropy term in the loss, or multi-task learning (i.e., similar to [17]). NoWMH is the ablation without the prior on the mean of the WMH intensities.

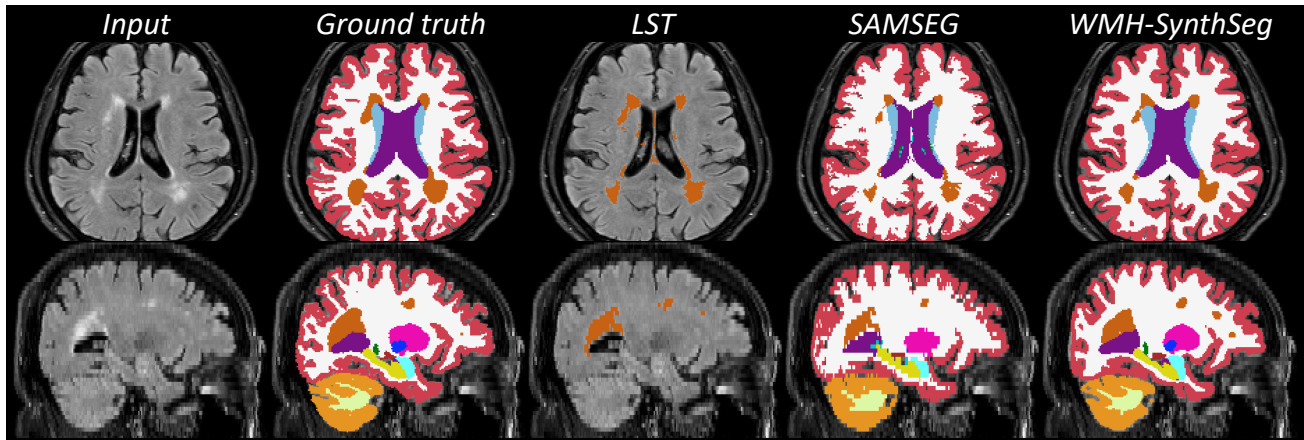


Fig. 1: Input, ground truth, and automated segmentations of a sample high-field scan from the Singapore dataset. The top row shows the high-resolution axial view; the bottom row shows a lower resolution orthogonal view (in sagittal orientation).

Method	T1w		FLAIR	
	Hippo	WMH	Hippo	WMH
LST (LPA)	N/A	N/A	N/A	-0.33
SAMSEG	0.71	0.63	0.69	0.64
WMH-SynthSeg (full)	0.89	0.75	0.86	0.85

Table 2: Correlation between ground truth volumetric measurements obtained from high-field (FreeSurfer from T1w for anatomy, LST from FLAIR for WMH) and from automated segmentations of the pMRI (MGH dataset). The hippocampal volumes (“Hippo”) are left-right averaged.

Figure 1 shows a qualitative comparison on a FLAIR scan from the Singapore dataset, both in the high-resolution axial plane, and in a lower resolution orthogonal view (sagittal). LST produces crisp segmentations of the WMH at native resolution, but with many false positives around the septum pellucidum (between the ventricles). SAMSEG, which also operates at native resolution, struggles with partial voluming (e.g., for the cortex) and often undersegments WMH. Our method, on the other hand, produces *isotropic* segmentations that are accurate for both anatomy and WMH.

In the FPR experiment with young controls, our method produces on average 950 mm^3 . This is a low value comparable to that produced by SAMSEG (877 mm^3); we note that LST is not compatible with the ADHD dataset as it has T1w contrast. The ablated versions show increases to $1,150 \text{ mm}^3$ (without the WM mean prior) and $1,850 \text{ mm}^3$ (without the prior or multi-task learning), highlighting the contribution of these components to the accuracy of the algorithm.

Finally, Table 2 shows the correlations between the volumetric measurements derived from the high-field scans (ground truth) and the pMRI, for the WMH and for a representative brain ROI (the hippocampus, which is tightly connected with aging and many brain diseases, e.g., dementias). LST completely fails at low field, as it was not designed for

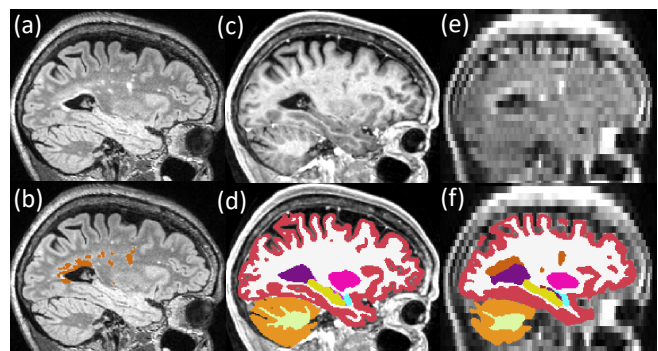


Fig. 2: (a) High-field 1mm isotropic FLAIR from MGH dataset. (b) LST segmentation, used as ground truth for WMH. (c) High-field 1mm T1w. (d) FreeSurfer segmentation of (c), used for ground truth for anatomy. (e) pMRI of the same subject at $2 \times 2 \times 5.8 \text{ mm}$ axial resolution. (f) WMH-SynthSeg segmentation. We note that, despite affine alignment of the high-field images to the pMRI, the anatomy on the slices is slightly different due to nonlinear distortion.

it. Being contrast agnostic, SAMSEG yields fairly strong correlations (between .63 and .71). WMH-SynthSeg produces very strong correlations (12-21 points higher than SAMSEG). This is attributed to its excellent ability to adapt to low-field images, which is qualitatively exemplified in Figure 2.

4. CONCLUSION

We have presented the first method that can simultaneously segment brain ROIs and WMH in scans of any resolution and contrast, including pMRI. Future work will include realistic modeling of WMH and evaluation on pMRI from larger cohorts. WMH-SynthSeg is publicly available and has potential in analyzing pMRI acquired in medically underserved areas.

5. ACKNOWLEDGMENTS

Supported by a grant from the Jack Satter Foundation and by NIH grants RF1MH123195, R01AG070988, R01EB031114, UM1MH130981, RF1AG080371, and R01NS112161.

6. REFERENCES

- [1] A de Havenon, NR Parasuram, et al., “Identification of white matter hyperintensities in routine emergency department visits using portable bedside magnetic resonance imaging,” *J of the American Heart Association*, vol. 12, no. 11, pp. e029242, 2023.
- [2] R Dobson and G Giovannoni, “Multiple sclerosis—a review,” *European J Neurology*, vol. 26, pp. 27–40, 2019.
- [3] E Fisher, J Lee, et al., “Gray matter atrophy in multiple sclerosis: a longitudinal study,” *Annals of Neurology*, vol. 64, no. 3, pp. 255–265, 2008.
- [4] B Fischl, D Salat, et al., “Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain,” *Neuron*, vol. 33, pp. 341–355, 2002.
- [5] B Patenaude, SM Smith, et al., “A bayesian model of shape and appearance for subcortical brain segmentation,” *Neuroimage*, vol. 56, no. 3, pp. 907–922, 2011.
- [6] P Schmidt, C Gaser, et al., “The LST toolbox for lesion segmentation and quantification,” in *Computer Methods in Biomechanics and Biomedical Engineering*, 2012, vol. 16, pp. 196–200.
- [7] L Griffanti, G Zamboni, et al., “Bianca (brain intensity abnormality classification algorithm): A new tool for automated segmentation of white matter hyperintensities,” *NeuroImage*, vol. 141, pp. 191–205, 2016.
- [8] O Puonti, JE Iglesias, et al., “Fast and sequence-adaptive whole-brain segmentation using parametric bayesian modeling,” *NeuroImage*, vol. 143, pp. 235–249, 2016.
- [9] S Cerri, O Puonti, et al., “A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in MS,” *NeuroImage*, vol. 225, pp. 117471, 2021.
- [10] A Roy, S Conjeti, et al., “QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy,” *NeuroIm*, vol. 186, pp. 713–727, 2019.
- [11] L Henschel, S Conjeti, et al., “Fastsurfer – a fast and accurate deep learning based neuroimaging pipeline,” *NeuroImage*, vol. 219, pp. 117012, 2020.
- [12] T Brosch, LYW Tang, et al., “Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to MS lesion segmentation,” *IEEE Trans Med Im*, vol. 35, no. 5, pp. 1229–1239, 2016.
- [13] M Ghafoorian, N Karssemeijer, et al., “Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities,” *Scientific Reports*, vol. 7, no. 1, pp. 5110, 2017.
- [14] M Wang and W Deng, “Deep visual domain adaptation: A survey,” *Neurocomputing*, vol. 312, pp. 135–53, 2018.
- [15] B Billot, DN Greve, et al., “SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining,” *Med Im Anal*, vol. 86, pp. 102789, 2023.
- [16] Iglesias JE, B Billot, et al., “SynthSR: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry,” *Science Advances*, vol. 9, no. 5, pp. eadd3607, 2023.
- [17] B Billot, S Cerri, et al., “Joint segmentation of multiple sclerosis lesions and brain anatomy in MRI scans of any contrast and resolution with CNNs,” in *ISBI*. IEEE, 2021, pp. 1971–1974.
- [18] O Ronneberger, P Fischer, et al., “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*. Springer, 2015, vol. 18, pp. 234–241.
- [19] Y Wu and K He, “Group normalization,” in *ECCV*, 2018, pp. 3–19.
- [20] DC Van Essen, SM Smith, et al., “The WU-Minn human connectome project: an overview,” *Neuroimage*, vol. 80, pp. 62–79, 2013.
- [21] CR Jack Jr, MA Bernstein, et al., “The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods,” *J of MRI*, vol. 27, no. 4, pp. 685–691, 2008.
- [22] HJ Kuijf, JM Biesbroek, et al., “Standardized assessment of automatic segmentation of white matter hyperintensities; results of the wmh segmentation challenge,” *IEEE Trans Med Im*, 2019.
- [23] A Carass, S Roy, et al., “Longitudinal MS lesion segmentation data,” *Data in brief*, vol. 12, pp. 46–50, 2017.
- [24] O Commowick, A Istace, et al., “Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure,” *Scientific Reports*, vol. 8, no. 1, pp. 13650, 2018.
- [25] P Bellec, C Chu, et al., “The neuro bureau ADHD-200 repository,” *NeuroImage*, vol. 144, pp. 275–286, 2017.
- [26] P Schmidt, *Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging*, Ph.D. thesis, LMU, 2017.