Received: 3 December 2021 | Revised: 25 August 2022 | Accepted: 7 September 2022

DOI: 10.1002/hbm.26109

RESEARCH ARTICLE

WILEY

Semisupervised white matter hyperintensities segmentation on MRI

Fan Huang[1] | Peng Xia[1] | Varut Vardhanabhuti[1] | Sai-Kam Hui[2] | Kui-Kai Lau[3,4] | Henry Ka-Fung Mak[1] | Peng Cao[1]

[1]Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China

[2]Department of Rehabilitation Science, The Hong Kong Polytechnic University, Hong Kong, China

[3]Department of Medicine, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China

[4]The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China

Correspondence
Peng Cao, Department of Diagnostic Radiology, LKS Faculty of Medicine, Room 406, Block K, Queen Mary Hospital, Pokfulam Road, Hong Kong, SAR, China.
Email: caopeng1@hku.hk

Funding information
Health and Medical Research Fund, Grant/Award Numbers: 06172916, 07182706; University Research Committee, University of Hong Kong (HKU), Grant/Award Number: 201905159001

Abstract

This study proposed a semisupervised loss function named level-set loss (LSLoss) for cerebral white matter hyperintensities (WMHs) segmentation on fluid-attenuated inversion recovery images. The training procedure did not require manually labeled WMH masks. Our image preprocessing steps included biased field correction, skull stripping, and white matter segmentation. With the proposed LSLoss, we trained a V-Net using the MRI images from both local and public databases. Local databases were the small vessel disease cohort (HKU-SVD, $n = 360$) and the multiple sclerosis cohort (HKU-MS, $n = 20$) from our institutional imaging center. Public databases were the Medical Image Computing Computer-assisted Intervention (MICCAI) WMH challenge database (MICCAI-WMH, $n = 60$) and the normal control cohort of the Alzheimer's Disease Neuroimaging Initiative database (ADNI-CN, $n = 15$). We achieved an overall dice similarity coefficient (DSC) of 0.81 on the HKU-SVD testing set ($n = 20$), DSC = 0.77 on the HKU-MS testing set ($n = 5$), and DSC = 0.78 on MICCAI-WMH testing set ($n = 30$). The segmentation results obtained by our semisupervised V-Net were comparable with the supervised methods and outperformed the unsupervised methods in the literature.

KEYWORDS
brain MRI, convolutional neural networks, deep learning, segmentation, semisupervised learning, small vessel diseases, white matter hyperintensities

1 | INTRODUCTION

White matter hyperintensities (WMHs) are common in patients with cerebral small vessel disease (SVD) and also demyelinating disorders such as multiple sclerosis (MS; Wardlaw et al., 2013, 2015). They were usually measured using T2-weighted (T2W) or fluid-attenuated inversion recovery (FLAIR) sequences on magnetic resonance imaging (MRI; Wardlaw et al., 2013). Automatic WMH volume segmentation in FLAIR images could benefit the clinical management of neurovascular and neurodegenerative diseases, serving as a quantitative method for the assessment in a large patient population (Carass et al., 2017; Driscoll et al., 2009; Li et al., 2014; Lladó et al., 2012; Moeskops et al., 2015; Thambisetty et al., 2010).

WMH segmentation techniques could be categorized into supervised and unsupervised methods. Supervised WMH segmentation methods were based on training machine learning classifiers for the segmentation. Conventional supervised WMH segmentation methods included k-nearest neighbors (Anbeek et al., 2004; Steenwijk

et al., 2013), support vector machine (Ferrari et al., 2003), and random decision forest (Akselrod-Ballin et al., 2009; Kamber et al., 1995). Recent deep learning techniques were based on the convolutional neural network (CNN) and fully convolutional network (FCN), which learned nonhandcrafted features from manual labels (Brosch et al., 2016; Moeskops et al., 2018; Valverde et al., 2017). FCN achieved state-of-the-art performance on multiple segmentation challenges (Carass et al., 2017; Guerrero et al., 2018; Kamnitsas et al., 2017; Kuijf et al., 2019). But, most of the deep neural networks required enormous labels for the model training, which could be challenging due to the time and cost issues in the manual annotation (Atlason et al., 2019). Furthermore, the human annotation may suffer from intrarater and interrater variabilities (Grimaud et al., 1996). Recent efforts of deep learning methodology focused on image synthesis techniques using generative adversarial networks to ease the needs of training data (Bowles et al., 2018; Frid-Adar et al., 2018).

For unsupervised segmentation methods, WMH areas were segmented from normal-appearing brain tissues based on their intensity features in T1-weighted (T1w) and FLAIR images (Lladó et al., 2012). In T1w, edema-associated WMH appeared as hypointense regions in contrast with normal white matter (WM; Bastin et al., 2002). In FLAIR, WMH appeared as hyperintense regions in contrast with normal WM. Thus, unsupervised segmentation methods first categorized the voxels on T1w images into gray matter (GM), WM, and cerebrospinal fluid (CSF) by intensity-based probabilistic models, such as class conditional probability density function (Ithapu et al., 2014; Khayati et al., 2008), the stochastic model estimated from health brain MRI images (Van Leemput et al., 2001), and atlas-based k-nearest neighbors (De Boer et al., 2009; Wu, Albert, et al., 2019). The GM, WM, and CSF masks were then applied to the coregistered FLAIR image of the same patient to find voxel-wise outliers, which were attributed to the WMH (De Boer et al., 2009; Jain et al., 2015; Schmidt et al., 2012; Sudre et al., 2015; Van Leemput et al., 2001).

A few deep learning studies focused on semisupervised training for FCN models, which were conventionally used in supervised learning. The SegAE network proposed by Atlason et al. was an unsupervised deep learning approach for WMH segmentation (Atlason et al., 2019). The network utilized a convolutional autoencoder to model the intensities of multiple MRI scans and reconstruct these images as weighted sums for WMH segmentation. In another study, Kim and Ye (2019) proposed a loss function based on Mumford-Shah function that can be used in deep learning-based image segmentation with or without labeled data. The authors showed the Mumford-Shah loss was trainable through error back-propagation (Kim & Ye, 2019). They examined the loss function with the public databases of natural images and medical images and showed that the Mumford-Shah loss could be synergistically combined with the supervised loss term to improve the segmentation performance (Kim & Ye, 2019). Especially, they achieved an overall Dice coefficient of 0.88 on the brain tumor BRATS database (Menze et al., 2014).

In this article, we proposed a level-set loss function to train a segmentation network without supervision for WM hyper-intensity segmentation on FLAIR images, which was adapted from the concept of the Mumford-Shah loss function by Kim et al. The level-set loss was inspired by traditional level-set algorithms, which defined an energy term for an enclosed contour (named level-set contour). When the energy term iteratively decreased (via gradient descent method) to the minimum, the level-set contour evolved toward the desired foreground. Finally, it stopped at the foreground boundary, segmenting the desired target. Therefore, the level-set loss (LSLoss) allows the training of a segmentation network without actual ground truth labels. In this study, we used the level-set loss function for WMH segmentation, which is one of the pathological patterns found in FLAIR images from patients with cerebral SVD. To guarantee the network recognized WMH as the foreground, not other normal brain tissues like the GM, we predefined four parameters regarding the WMH appearance on FLAIR images. Our contributions included: (1) we developed four loss terms, including the overall foreground loss, the overall background loss, the region-of-interest (ROI) loss, and the divergence loss for WMH segmentation; (2) we trained a segmentation network (a 3D V-Net) in an unsupervised manner where no ground truth labels were required; and (3) we evaluated the LSLoss with multicenter datasets ($n = 455$), including the images of both SVD and MS cohorts.

2 | MATERIALS AND METHODS

This study included both local ($n = 380$) and public datasets ($n = 75$) for the network training and evaluation. The local MRI datasets were from our own institutional MRI unit. They included image data from two retrospective studies on small vessel disease cohort (HKU-SVD) and multiple sclerosis cohort (-MS). We also collected the image data from the WMH segmentation challenge database (Kuijf et al., 2019), which was organized by the International Conference on Medical Image Computing Computer-assisted Intervention (MICCAI) 2017. In addition, we included 15 cases of healthy subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database for training our model.

2.1 | The HKU-SVD dataset

We have previously established a large cohort of 1076 consecutive Chinese ischemic stroke patients with MRI performed at our institutional MRI Unit during 2008–2014. Details of this cohort, including the imaging protocols, can be found in our previous publications (Lau et al., 2017, 2018). For the purpose of this study, we focused on the 413/1076 patients who had an ischemic stroke classified due to the SVD score according to the TOAST (Trial of Org 10172 in Acute Stroke Treatment) classification (Adams et al., 1993). After excluding 53 patients who had incomplete imaging data (either missing T1w or FLAIR images), imaging data from 360 ischemic patients due to SVD were used in this study. Their MRI images (both T1w and FLAIR) were retrieved from the database. In this database, 237 cases were axial FLAIR images, and 123 were coronal FLAIR images. The MRI scanner parameters and the image details were summarized in Tables 1 and 2, respectively. The data

TABLE 1 An overview of the MRI scanning parameters for images from different databases in this study

		T1w		MPRAGE			FLAIR		
Medical center	Scanner	TR (ms)	TE (ms)	TR (ms)	TE (ms)	TI (ms)	TR (ms)	TE (ms)	TI (ms)
Local MRI unit	3 T Philips Achieva	4.79	2.44	2000	20	800	11,000	120	2800
UMC Utrecht	3 T Philips Achieva	7.9	4.5	–	–	–	11,000	125	2800
NUHS Singapore	3 T Siemens TrioTim	–	–	2300	1.9	900	9000	82	2500
VU Amsterdam	3 T GE Signa HDxt	7.8	3.0	–	–	–	8000	126	2340
ADNI-CN	–	–	–	–	–	–	–	–	–

Abbreviations: ADNI-CN, Alzheimer's Disease Neuroimaging Initiative-control normal; FLAIR, fluid-attenuated inversion recovery; MRI, magnetic resonance imaging; T1w, T1-weighted; TE, echo time; TR, repitation time.

utilization was approved by the research ethics committee (Institutional Review Board [IRB]), the University of Hong Kong. Therefore, this dataset was referred to as the HKU-SVD dataset.

2.2 | The HKU-MS dataset

We also collected the data from the MS cohort ($n = 20$) in another study (Qian et al., 2016). In brief, this retrospective study aimed to investigate if diffusion and diffusional kurtosis metrics could well aid in discriminating MS and neuromyelitis optica. The patients' scans were performed from 2010 to 2014 at our institutional MRI unit. All patients were scanned using the same standard-of-care protocol used for the HKU-SVD dataset. All MRI images were acquired using axial slice orientation. The data utilization was approved by the research ethics committee (IRB), the University of Hong Kong. The MRI scanner parameters and the image detail were summarized in Tables 1 and 2, respectively. This dataset was referred to as the HKU-MS dataset.

2.3 | The MICCAI WMH challenge dataset

As mentioned in the MICCAI official webpage, the brain data of the MICCAI 2017 challenge aimed to provide a benchmark for automatic segmentation of WMH of presumed vascular origin (Kuijf et al., 2019). The training set was publicly available, including 60 cases from three different MRI centers located in Amsterdam (the Netherlands), Utrecht (the Netherlands), and Singapore, which were referred to as the MICCAI-AMS, MICCAI-UTR, and MICCAI-SGP databases. The MRI parameters and the image detail were summarized in Tables 1 and 2, respectively.

2.4 | The ADNI dataset

The ADNI database was a large and publicly available database for developing improved methods for clinical trials (http://adni.loni.usc.edu/). We collected both T1w and FLAIR images of 15 healthy subjects from the "control normal" (CN) group (Mueller et al., 2005; Weiner et al., 2010). These images were non-WMH presence cases and included in the training phase to improve the network generalizability. This database was referred to as ADNI-CN dataset.

2.5 | Level-set loss for WMH segmentation

The study by Kim and Ye (2019). showed that a Mumford-Shah functional could be directly minimized through error back-propagation, as the SoftMax layer in CNN can be used as a differentiable approximation of the characteristic function. In this study, we introduced how we adapted this concept to formulate a semisupervised LSLoss function for WMH segmentation in FLAIR images. The loss terms of the proposed LSLoss were developed based on (1) WMH had a typically brighter appearance than other brain tissues on FLAIR, and (2) they were within the WM. We first clustered the brain tissues into WM, GM, and CSF as described in the "Data preparation" subsection. The resulted WM binary mask was then used as a ROI mask.

Inside the ROI, the image was separated into a foreground area $(1-H(\phi))$ and a background area $H(\phi)$, where the $H(.)$ is the Heaviside step function, and ϕ is the signed-distance function ($\phi(x) = 0$ is the enclosed contour for segmentation). In this study, we used a Tanh layer as the network output to generate ϕ for the LSLoss training. Based on this, the loss terms were defined as:

$$\text{LSLoss} = \omega_{\text{fore}} L_{\text{fore}} + \omega_{\text{back}} L_{\text{back}} + \omega_{\text{ROI}} L_{\text{ROI}} + \omega_{\text{div}} \,\text{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right).$$

It contains four loss terms: (1) the overall image intensity difference between the foreground area and an estimated foreground value μ_f (rescaled by σ_f) inside the ROI; (2) the overall difference ROI mask; (3) the area outside the ROI mask $((1-H(\phi))-\text{ROI})^2(1-\text{ROI})$; and (4) a divergence between the background area and an estimated background value μ_b (rescaled by σ_b) inside the term to maintain the segmentation as an enclosed contour with smooth boundary. The regions corresponding to the first three region-based loss terms were shown in Figure 1. The loss terms were formulated respectively as:

$$L_{\text{fore}} = \frac{\int_\Omega \left(\frac{I-\mu_f}{\sigma_f}\right)^2 (1-H(\phi))\,\text{ROI}}{\int_\Omega (1-H(\phi))\,\text{ROI}}$$

TABLE 2 The scanner and image details of the T1w and FLAIR in this study

Medical center	MR scanner	Cohort	Number of subjects (labeled cases)	T1w/MPRAGE		FLAIR			
				Image size	Voxel size	Type	No.	Image size	Voxel size
Local MRI unit	3 T Philips Achieva	SVD	360 (20)	512 × 512 × 25–800 × 800 × 25	0.3 × 0.3 × 5.5–0.45 × 0.45 × 5.5	Axial	237	512 × 512 × 50–768 × 768 × 60	0.3 × 0.3 × 2.5–0.45 × 0.45 × 2.5
						Coronal	123	512 × 512 × 25–768 × 768 × 30	0.3 × 0.3 × 5.5–0.45 × 0.45 × 5.5
		MS	20 (20)	256 × 256 × 170	1 × 1 × 1.2	Axial	20	512 × 512 × 330	0.5 × 0.5 × 0.55
UMC Utrecht	3 T Philips Achieva	SVD	20 (20)	256 × 256 × 192	1 × 1 × 1	Axial	20	240 × 240 × 48	0.96 × 0.95 × 3
NUHS Singapore	3 T Siemens TrioTim	SVD	20 (20)	256 × 256 × 192	1 × 1 × 1	Axial	20	256 × 232 × 48	1 × 1 × 3
VU Amsterdam	3 T GE Signa HDxt	SVD	20 (20)	256 × 256 × 176	0.94 × 0.94 × 1	Sagittal	20	256 × 132 × 83	0.98 × 0.98 × 1.2
ADNI	–	Control	15 (–)	240 × 256 × 171–256 × 256 × 211	1 × 1 × 1–1 × 1 × 1.2	–	–	256 × 256 × 141–256 × 256 × 182	1 × 1 × 1.2

Note: Most of the data (340 subjects) from our own database were without WMH labels. The data from UMC Utrecht, NUHS Singapore, and VU Amsterdam had WMH ground truth labels.
Abbreviations: FLAIR, fluid-attenuated inversion recovery; SVD, small vessel disease; T1w, T1-weighted; WMH, white matter hyperintensities.

$$L_{\text{back}} = \frac{\int_{\Omega}\left(\frac{I-\mu_b}{\sigma_b}\right)^2 H(\phi)\,\text{ROI}}{\int_{\Omega} H(\phi)\,\text{ROI}}$$

$$L_{\text{ROI}} = \frac{\int_{\Omega}((1-H(\phi))-\text{ROI})^2(1-\text{ROI})}{\int_{\Omega} 1-\text{ROI}}$$

$$\text{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) = \frac{\partial}{\partial x}\left(\frac{\phi_x}{|\nabla\phi|}\right) + \frac{\partial}{\partial y}\left(\frac{\phi_y}{|\nabla\phi|}\right) + \frac{\partial}{\partial z}\left(\frac{\phi_z}{|\nabla\phi|}\right)$$

The parameters, ω_{fore}, ω_{back}, ω_{ROI}, and ω_{div}, are the weightings for the loss terms. In this study, we used $\omega_{\text{fore}} = 0.5$, $\omega_{\text{back}} = 0.2$, $\omega_{\text{ROI}} = 0.2$, and $\omega_{\text{div}} = 0.1$ for the training. The values were determined by recursively searching the global minimum of LSLoss using data with ground truth labels.

By training a network via error back-propagation, the LSLoss reached its minimum, in which the foreground $(1-H(\phi))$ segmented the bright area (i.e., WMH) on FLAIR, the background $H(\phi)$ segmented the normal-appearing WM and the area outside the ROI mask was not interested. The final WMH segmentation was obtained as $(1-H(\phi))$, where $H(.)$ is the Heaviside step function, and ϕ is the output of the Tanh layer of the segmentation network.

2.6 | WMH segmentation neural network

In this study, we examined the LSLoss with the segmentation network V-Net (Milletari et al., 2016; Ronneberger et al., 2015). A schematic of the V-Net used in this study is shown in Figure 2. It should be noted that the V-Net had dual outputs: one was a Tanh layer for the LSLoss, and another was a logistic sigmoid (Logit) layer for the cross-entropy loss (CELoss). The Tanh layer was trained with the LSLoss for WMH segmentation, for which no ground truth labels were needed. The Logit layer was trained with a regular CELoss for WM mask segmentation, for which the FSL-produced WM mask (as introduced in the Section 2.7) was used as the ground truth labels. After training, the V-Net could perform both WM and WMH segmentations simultaneously. For deploying the model, the WM mask by the FSL toolbox was no longer needed. The final WMH segmentation was obtained by multiplying the binarized WM mask with the predicted WMH mask.

2.7 | Data preparation

For the images from the HKU-SVD, HKU-MS, and ADNI-CN datasets, we applied the same preprocessing pipeline as that had been done for the MICCAI-WMH database (Kuijf et al., 2019). First of all, since the datasets consisted of both axial and coronal slices, we swapped the dimension of the original FLAIR volumes such that the thickest slices

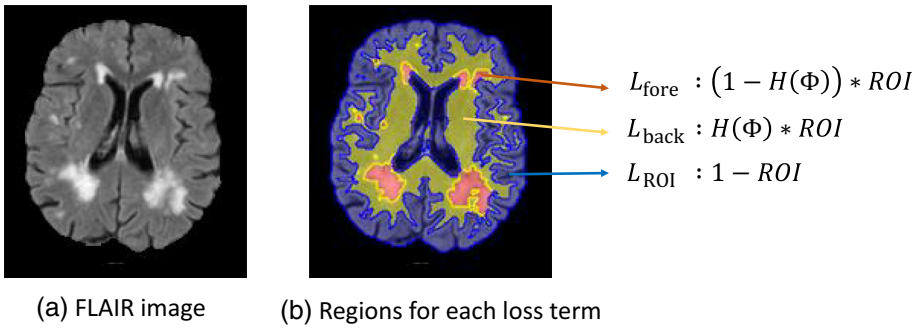

FIGURE 1 A demonstration of brain regions where the three level-set loss terms measured: (1) L_{fore}, (2) L_{back}, and (3) L_{ROI}. Region-of-interest (ROI) was the white matter (WM) area, and $H(\Phi)$ was the normal-appearing WM area. The terms L_{fore} and L_{back} limited the foreground loss $([I-\mu_f]/\sigma_f)^2$ and background loss $([I-\mu_b]/\sigma_b)^2$ to be calculated within the ROI. L_{ROI} forced $H(1-\Phi)$, that is, the WMH area, to be 0 within (1−ROI), that is, outside the ROI

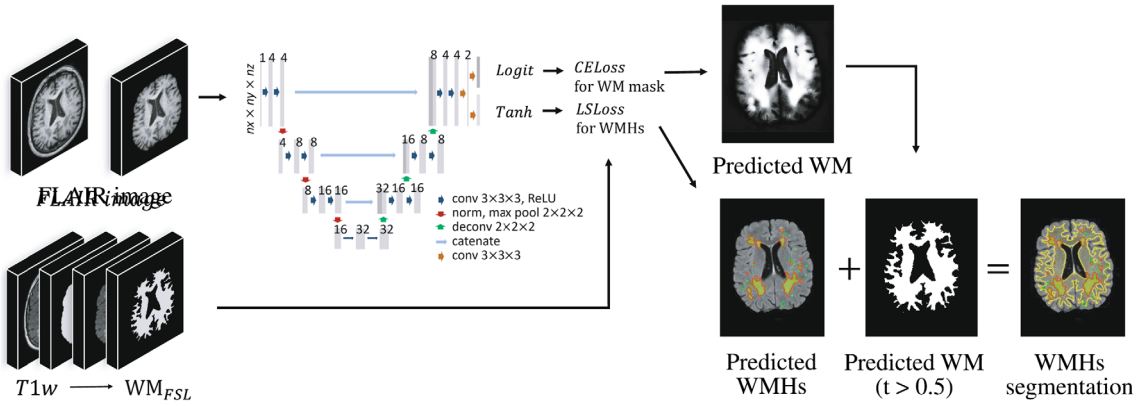

FIGURE 2 The network architecture in this study. We trained a V-Net with dual loss functions: One was the proposed level-set loss (LSLoss) for white matter hyperintensity (WMH) segmentation, and another was the cross-entropy loss (CELoss) for WM segmentation. The final WMH segmentation results were obtained by multiplying the binarized WM mask with the predicted WMH map. The WM mask was obtained using the FSL toolbox, which was only required during training

were moved to the last image dimension. Then the volumes were preprocessed using SPM12 to correct bias field inhomogeneities (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) (Ashburner & Friston, 2005). Afterward, the T1w images were coregistered to the bias-field corrected FLAIR volumes of the same subject using Elastix 4.8 (https://elastix.lumc.nl/) (Klein et al., 2009; Shamonin et al., 2014). The parameter file for registration was obtained from the challenge website (https://wmh.isi.uu.nl/). After the registration, we obtained the brain tissue volumes by skull stripping using the BET tool of FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) (Jenkinson et al., 2005; Smith, 2002), and clustered the brain tissues into three classes: CSF, GM, and WM, using the FAST tool in FSL (Zhang et al., 2000). At last, the extracted WM masks were applied on the co-registered FLAIR volumes for WMH segmentation network training. For evaluation, we created testing sets for the HKU-SVD and HKU-MS data, respectively. We randomly selected 20 cases from the HKU-SVD dataset and 5 cases from the HKU-MS dataset and manually annotated the WMH regions using the ITKSnap 3.8.0 (http://www.itksnap.org) (Yushkevich et al., 2006). For the HKU-SVD testing set, 13 cases were of axial slices, and 7 were of coronal slices. An experienced neurologist from our department reviewed the annotation results.

Since the WMH has various sizes and shapes, the image intensity was the main feature for segmentation. On FLAIR images, CSF effects were suppressed with intensities close to 0.0. GM was slightly brighter than WM due to the T2 weighting. WMH was much brighter than both WM and GM. We normalized the FLAIR images to let CSF have an intensity of 0.0. WM intensity was set to 1.0. Such that if WMH were presented in the image, their intensity values would be above 1.0. It allowed us to determine the user-defined parameters μ_f and μ_b. The image normalization was done as follows:

$$I_N = \frac{I_{\text{FLAIR}}(x) - \mu_{\text{WM}}}{\mu_{\text{WM}}}$$

where μ_{WM} was the average intensity in the WM mask. After the normalization, I_N was rescaled by:

$$I_{NR} = \frac{I_N - \min}{\text{mean}_{\text{GM}}}$$

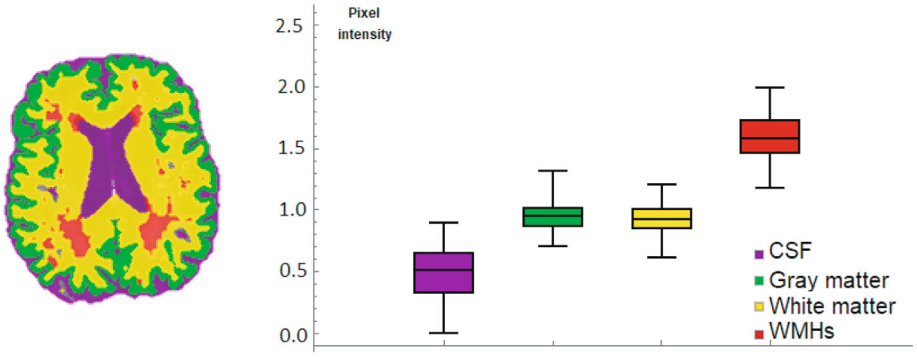

FIGURE 3 After the preprocessing, the intensities of cerebrospinal fluid (CSF), white matter (WM), gray matter (GM), and white matter hyperintensity (WMH) were shown in the box plots. In this example, the average intensities of CSF, WM, and WMH were 0.5, 1.0, and 1.5, respectively.

where min is the minimum value of I_N (in most cases min $= -1$), and mean_{GM} the average intensity of GM mask. Figure 3 showed the box plots for CSF, WM, GM, and WMH, respectively, after the described normalization. In this study, we set $\mu_f = 1.5$ and $\mu_b = 1.0$. Both the σ_f and σ_b were set to 0.5, which were estimated using 10 randomly selected cases from the MICCAI-WMH dataset using the provided WMH labels.

2.8 | Network training

A GTX2080 Ti GPU was used to train the V-Net for 20 epochs with a learning rate of 0.01 using the Adam optimizer (Kingma & Ba, 2014), and a batch size of 15 for training and 5 for validation. The network architecture and the level-set loss function were implemented by the deep learning package of Wolfram Mathematica 12.0 (https://www.wolfram.com/).

2.9 | Evaluation

In this study, the WMH segmentation results were evaluated in 3D using the evaluation metrics proposed on the MICCAI WMH challenge, where the evaluation python script was downloaded from the challenge website. In brief, we evaluated our method with dice similarity coefficient (DSC), Haussdorf distance (95 percentile), absolute volume difference (AVD), true positive rate (TPR), and F1-score. The details of the evaluation metrics are listed below:

- Dice similarity coefficient

The DSC measures the overlap between the ground truth mask and the predicted mask. It is calculated via $\frac{2TP}{2TP+FP+FN}$ where the TP, FP, and FN are the pixel-wise true positives, false positives, and false negatives. A higher DSC (the maximum is 1) indicates a better segmentation.

- 95 percentile Hausdorff distance (H95)

The Hausdorff distance measures the longest distance traveling from a point in one set to a point in another set. It is defined as $d_H(T,P) = \max\{\sup_{t \in T} \inf_{p \in P} d(t,p)\}$ is the Euclidean distance between point t and p. Since max is sensitive to outliers, we calculated the 95-percentile value instead of the maximum for the evaluation as suggested in the work by Kuijf et al. (2019).

- Absolute volume difference

The AVD is calculated via $\frac{|V_T - V_P|}{V_T}$, where VT and VP are the volume (width $\times$ height $\times$ depth mm^3) of the ground truth WMH mask and the predicted WMH mask, respectively. A lower AVD indicates a better segmentation.

- Per-WMH TPR

The TPR (referred as sensitivity or recall in the literature) is a measurement for the number of correctly detected WMH volumes. It is calculated via $\frac{TP}{TP+FN}$, where TP and FN are the true positive and false negative.

- Per-WMH F1-score

The F1-score is calculated from the per-WMH precision, and TPR of the WMH detection, where the precision is $\frac{TP}{TP+FP}$ and the TPR was described above. A higher F1-score indicates better precision and TPR, and the lowest possible value is 0 if either the precision or the TPR is zero.

2.10 | Data analysis for higher resolution images

Our first study evaluated the proposed LSLoss using the MRI images from the institutional MR imaging center: the HKU-SVD database. It consisted of 360 cases retrieved from the SVD cohort of our previous study. These were high-resolution (HR) FLAIR images with spatial resolutions ranging from 0.3 $\times$ 0.3 $\times$ 2.5 mm^3/pixel to 0.45 $\times$ 0.45 $\times$ 5.5 mm^3/pixel. We slightly rescaled the images to the same resolution 0.4 $\times$ 0.4 $\times$ 3 mm^3/pixel. Afterward, we cropped/padded the rescaled images to 256 $\times$ 256 $\times$ 64 (width $\times$ height $\times$ depth). We randomly selected 20 cases and manually annotated the WMH for evaluation. The remaining 340 cases were used for training. We calculated the DSC, HD95, AVD, TPR, and

F1-score on the 20 testing cases using (1) the raw WMH segmentation (i.e., no WM masks were applied), (2) using the FSL toolbox generated WM mask (denoted as WM_{FSL}), and (3) using the WM mask obtained from the trained V-Net (denoted as WM_{VNet}), respectively. Afterward, we further rescaled the HKU-SVD images to a lower resolution of $1 \times 1 \times 3$ mm^3/pixel to examine the effect of changing spatial image resolution on the segmentation performance, as we will resize these images for a cross-centers study.

2.11 | Data analysis for lower resolution images

Our second study evaluated the method with the cross-center MRI images consisting of 455 cases. These were from different sources, including our institutional imaging center (i.e., HKU-SVD and HKU-MS) and publicly available databases (i.e., MICCAI-WMH and ADNI-CN). More than three different MRI scanners acquired the images: 3 T Philips Achieva, 3 T Siemens TrioTim, and 3 T GE Signa HDxt with various scanning parameters (as summarized in Table 1). Moreover, the MRI images were from multiple disease cohorts, including SVD, MS, and CN cases. All the images were reoriented into the LPI (L-R within P-A within I-S) orientation and rescaled to the same spatial resolution $1 \times 1 \times 3$ mm^3/pixel. Afterwards, we cropped/padded the rescaled images to $128 \times 128 \times 64$ (width $\times$ height $\times$ depth). We used 400 cases for training and 55 cases with WMH labels for testing. We calculated the DSC, HD95, AVD, TPR, and F1-scores on the 55 testing cases. The evaluated WMH segmentations were obtained by multiplying the raw WMH outputs, and the corresponding predicted WM masks WMV Net. For comparison, we also summarized the literature results in the lower part of the Table 4, including the results of the techniques proposed by Atlason et al. (2019), Kamnitsas et al. (2017), Li et al. (2018), Viteri et al. (2021), Wu, Zhang, et al. (2019), Zhang et al. (2018), Zhao et al. (2018), and Zhou et al. (2020). Since the methods developed by Kamnitsas, Zhang, and Zhou were proposed earlier than the MICCAI WMH challenge, the methodologies were reimplemented and evaluated on the WMH dataset in a previous study (Jiang et al., 2020).

3 | RESULTS

The results of the first study, which compared the segmentation performances of the proposed method on both HR and low-resolution (LR) image inputs, were summarized in Table 3. On the HR images, we achieved an average DSC of 0.83 ± 0.05 (mean ± SD), an average TPR of 0.9 ± 0.10, and an average F1-scores of 0.92 ± 0.07. Meanwhile, on the LR images, the DSC, TPR, and F1-scores were 0.76 ± 0.1, 0.87 ± 0.19, and 0.88 ± 0.13, respectively. In Table 3, WMH segmentation on lower-resolution images was worse than that of higher-resolution images, especially in the per-WMH TPR. Figure 4 showed two cases with both HR and LR inputs. In the higher-resolution cases, small WMH segments (highlighted by the yellow circles) were corrected

TABLE 3 Segmentation performance of the proposed method on HR and LR images

Network and loss	Training: (dataset: no.)	Testing: (dataset: no.)	FLAIR voxel size (mm^3)	ROI mask	DSC (mean ± SD)	HD95 (mm) (mean ± SD)	AVD (mm3) (mean ± SD)	TPR (mean ± SD)	F1Scores (mean ± SD)
V-Net + LSLoss	HKU-SVD (HR): 340	HKU-SVD (HR): 20	$0.4 \times 0.4 \times 3$	No	0.80 ± 0.05	3.63 ± 2.75	14.08 ± 9.89	0.90 ± 0.11	0.95 ± 0.06
				WM_{FSL}	0.82 ± 0.05	3.30 ± 2.36	11.21 ± 7.46	0.91 ± 0.11	0.92 ± 0.06
				WM_{VNet}	0.83 ± 0.05	3.33 ± 2.45	11.46 ± 8.97	0.90 ± 0.10	0.92 ± 0.07
	HKU-SVD (LR): 340	HKU-SVD (LR): 20	$1 \times 1 \times 3$	No	0.74 ± 0.1	8.44 ± 16.41	25.72 ± 17.98	0.86 ± 0.2	0.90 ± 0.15
				WM_{FSL}	0.75 ± 0.1	8.37 ± 16.43	25.49 ± 16.25	0.86 ± 0.2	0.88 ± 0.14
				WM_{VNet}	0.76 ± 0.1	8.16 ± 15.83	24.83 ± 18.51	0.87 ± 0.19	0.88 ± 0.13

Abbreviations: AVD, absolute volume difference; DSC, dice similarity coefficient; FLAIR, fluid-attenuated inversion recovery; HR, high resolution; LR, low resolution; LSLoss, level-set loss; ROI, region-of-interest; SVD, small vessel disease; TPR, true positive rate.

TABLE 4 Performance of the proposed method with V-Net that was trained using local and public datasets

Authors	Network types	Training: (dataset: no.)	Testing: (dataset: no.)	DSC (mean ± SD)	HD95 (mean ± SD)	AVD (mean ± SD)	TPR (mean ± SD)	F1Scores (mean ± SD)
This study	V-Net + LSLoss (V-Net$_{LS}$)	HKU-SVD (LR): 340						
		HKU-MS (LR): 15	HKU-SVD (LR):20	0.81 ± 0.06	5.95 ± 9.32	19.02 ± 17.37	0.85 ± 0.19	0.87 ± 0.12
		MICCAI-AMS: 10	HKU-MS (LR): 5	0.77 ± 0.11	6.14 ± 4.07	25.49 ± 18.41	0.70 ± 0.24	0.68 ± 0.19
		MICCAI-UTR: 10	MICCAI-AMS: 10	0.74 ± 0.13	7.04 ± 6.16	55.03 ± 74.26	0.86 ± 0.13	0.68 ± 0.14
		MICCAI-SGP: 10	MICCAI-UTR: 10	0.76 ± 0.12	7.64 ± 7.37	28.66 ± 25.44	0.72 ± 0.12	0.7 ± 0.09
		ADNI-CN: 15	MICCAI-SGP: 10	0.83 ± 0.05	3.22 ± 0.87	13.70 ± 9.30	0.82 ± 0.15	0.71 ± 0.1
		Total: 400						
This study	V-Net + LSLoss[a]	MICCAI WMH training set	MICCAI WMH testing set	0.75 ± 0.11	6.04 ± 5.41	30.6 ± 38.64	0.61 ± 0.19	0.72 ± 0.15
Viteri et al. (2021)	U-Net[b]			0.78	6.24	28.29	0.78	0.74
Li et al. (2018)	SYSU Media[b]			0.80	6.30	21.87	0.84	0.76
Kamnitsas et al. (2017)	DeepMedic[b]			0.79	—	32.7	0.80	0.80
Wu, Albert, et al. (2019) and Wu, Zhang, et al. (2019)	U-Net with skip connection[b]			0.78	7.35	28.23	0.82	0.71
Zhou et al. (2020)	U-Net + CRF + Spatial[b]			0.78	3.69	22.86	0.77	0.67
Zhang et al. (2018)	Deep ResU-Net[b]			0.73	—	30.42	0.75	0.75
Zhao et al. (2018)	LPA[c]			0.62	—	44.22	0.76	0.64
Atlason et al. (2019)	SegAE[a]			0.63	21.93	39.63	0.3	0.37

Note: The upper part of the table summarized the results of the V-net trained on all the data without using any ground truth labels. The lower part summarized the results of the V-net trained on the MICCAI challenge dataset (5-fold cross-validation) and the literature summary on the same dataset.

Abbreviations: AVD, absolute volume difference; DSC, dice similarity coefficient; HR, high resolution; LR, low resolution; LSLoss, level-set loss; MICCAI, Medical Image Computing Computer-assisted Intervention; SVD, small vessel disease; TPR, true positive rate; WMH, white matter hyperintensities.

[a]Semisupervised method.

[b]Supervised method.

[c]Unsupervised method.

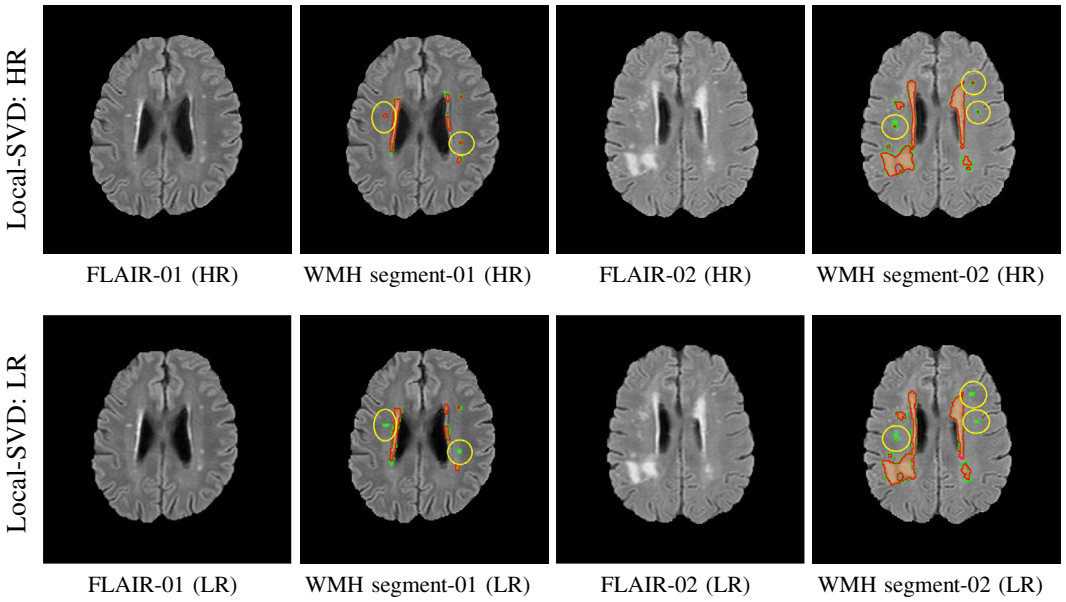

FIGURE 4 Segmentation results on high-resolution (HR) images were compared with those on low-resolution (LR) images. Green areas indicated the white matter hyperintensity (WMH) ground truth labels, and the red areas showed the WMH segmentation. The yellow circles highlighted the missed WMH on LR images. The HR images were with image resolution of 512 × 512 × 60, and the LR images were with image resolution of 256 × 256 × 48. FLAIR, fluid-attenuated inversion recovery; SVD, small vessel disease

TABLE 5 The comparison of segmentation performance between the proposed LSLoss and CELoss on MICCAI dataset

Methods	MICCAI dataset	DSC (mean ± SD)	HD95 (mean ± SD)	AVD (mean ± SD)	TPR (mean ± SD)	F1Scores (mean ± SD)
V-Net + LSLoss	AMS	0.74 ± 0.10	12.78 ± 10.78	21.49 ± 30.10	0.64 ± 0.11	0.71 ± 0.05
	UTR	0.76 ± 0.12	7.64 ± 7.37	28.66 ± 25.44	0.72 ± 0.12	0.70 ± 0.09
	SGP	0.83 ± 0.05	3.22 ± 0.87	13.70 ± 9.30	0.82 ± 0.15	0.71 ± 0.10
V-Net + CELoss	AMS	0.70 ± 0.12	16.95 ± 12.88	20.57 ± 22.10	0.25 ± 0.09	0.39 ± 0.10
	UTR	0.76 ± 0.08	14.42 ± 11.43	21.72 ± 11.68	0.08 ± 0.03	0.15 ± 0.05
	SGP	0.75 ± 0.09	10.80 ± 4.28	27.85 ± 14.92	0.10 ± 0.05	0.18 ± 0.08

Abbreviations: AVD, absolute volume difference; CELoss, cross-entropy loss; DSC, dice similarity coefficient; LSLoss, level-set loss; MICCAI, Medical Image Computing Computer-assisted Intervention; TPR, true positive rate.

detected. In contrast, they were missed in the lower-resolution results.

The results of the second study were summarized in the upper part of Table 4, in which we trained a stand-alone 3D V-Net using both local and public datasets with the proposed LSLoss (denoted as V-Net$_{LS}$). The V-Net$_{LS}$ achieved the average DSC of 0.81 and 0.77 on the HKU-SVD ($n = 20$) and HKU-MS ($n = 5$) images, respectively. For the MICCAI WMH database, V-Net$_{LS}$ achieved 0.74, 0.76, and 0.83 on the MICCAI-AMS, MICCAI-UTR, and MICCAI-SGP, respectively. We compared the segmentation results using a V-Net with CELoss on the MICCAI WMH database (as shown in Table 5; Jiang et al., 2020). The overall DSCs were 0.70, 0.76, and 0.75 on the MICCAI-AMS, MICCAI-UTR, and MICCAI-SGP databases, and the cross-center validation was only 0.71 and 0.63 on the HKU-SVD and HKU-MS images. A p-value ($p < .05$) of a paired t-test indicated that the difference between V-Net$_{LS}$ and V-Net$_{CE}$ was significant. In Table 3, the results were obtained by a network trained with a single-center dataset only (i.e., the HKU-SVD cohort). In Table 4, the segmentations were obtained by a network trained with cross-center datasets. Results on cross-center datasets were generally better than that on single-center datasets, with DSC improved from 0.76 to 0.81, HD95 improved from 8.16 to 5.95, and AVD from 24.83 to 19.02. In the lower part of Table 4, we summarized the segmentation performances on the MICCAI WMH dataset obtained by the proposed V-Net trained with LSLoss (5-fold cross-validation), in which we achieved average DSC, HD95, AVD, TPR, and F1score of 0.75, 6.04, 30.6, 0.61, and 0.72, respectively. We also summarized other supervised/unsupervised/semisupervised approaches proposed in the literature for comparison, including five deep learning-based, one unsupervised, and one semisupervised method.

Figure 5 demonstrated a few examples of the original 2D FLAIR images and the segmentation of the V-Net trained with the proposed LSLoss on images from the HKU-SVD, HKU-MS, and MICCAI

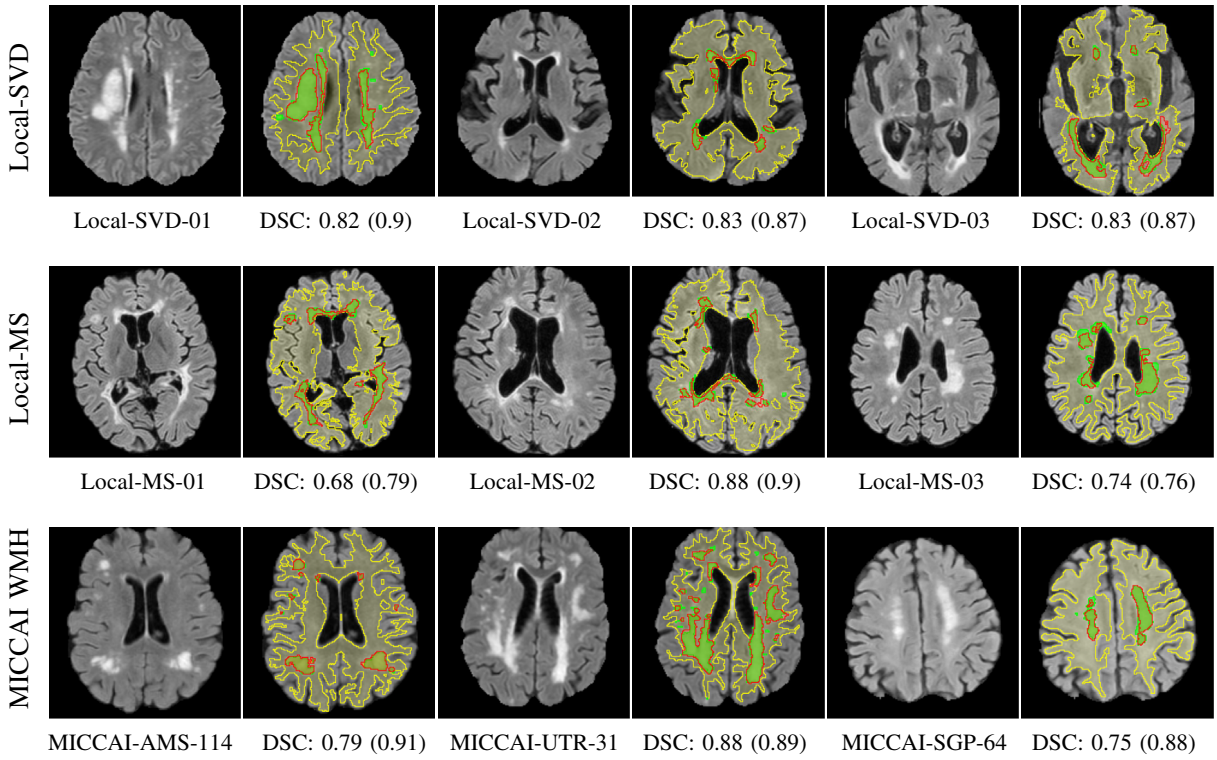

FIGURE 5 The white matter hyperintensity (WMH) were segmented by the V-Net trained with the proposed level-set loss. The images were from the small vessel disease cohort (HKU-SVD), multiple sclerosis cohort (HKU-MS), and Medical Image Computing Computer-assisted Intervention (MICCAI) WMH database. We showed both the fluid-attenuated inversion recovery (FLAIR) images and the corresponding WMH segmentations. On the segmentation results, the yellow masks indicated the V-Net predicted WM mask, the green masks indicated the ground truth labels, and the red masks indicated the V-Net$_{LS}$ segmentations. Image names were given under the FLAIR images, and the corresponding dice similarity coefficients (DSCs), denoted as "3D-DSC (2D-DSC of the current slice)," were given under the segmentation subfigures

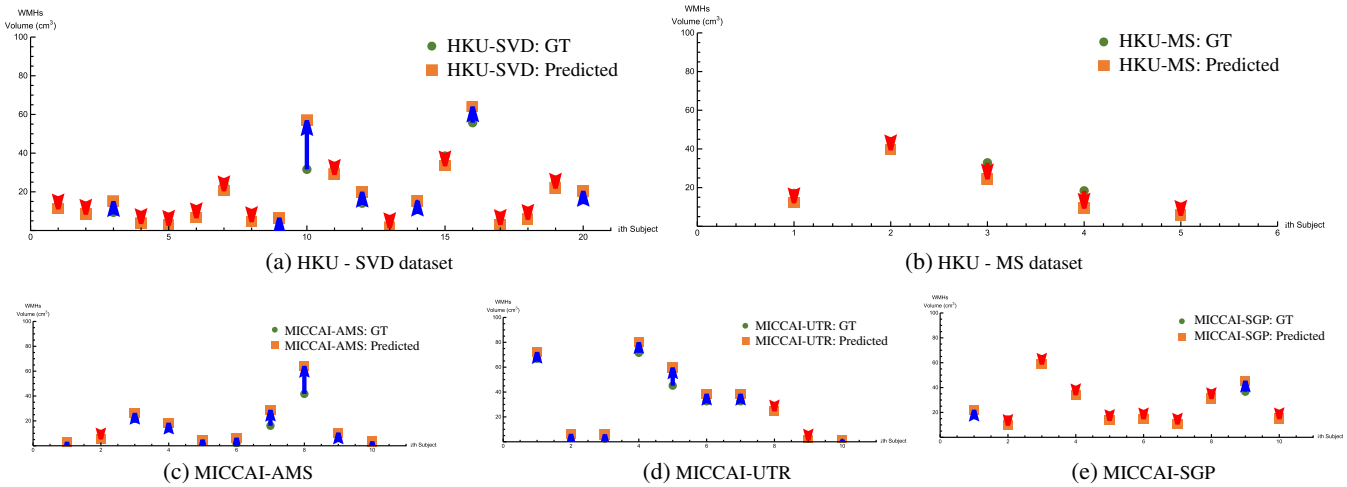

FIGURE 6 The scatter plots for comparing the predicted total white matter hyperintensity (WMH) volumes with the manual WMH volumes on the testing images from the small vessel disease cohort (HKU-SVD), multiple sclerosis cohort (HKU-MS), and Medical Image Computing Computer-assisted Intervention (MICCAI) WMH databases. (a,b) The plots for HKU-SVD and HKU-MS. (c–e) The plots for MICCAI WMH database. The blue arrows indicated over-estimated WMHs sizes (predicted WMHs volume > GT WMHs volume), and the red arrows indicated under-estimated WMHs sizes (predicted WMHs volume < GT WMHs volume)

WMH databases. The scatter plots in Figure 6 compared the predicted total WMH volumes to the manual WMH volumes (i.e., AVD). The WMH volumes from our method were in good agreement with that of the ground truth labels. In Figure 7, we demonstrated a few examples which gave high AVD in each dataset.

4 | DISCUSSION

In the literature, the supervised WMH segmentation methods achieved Dice coefficients between 0.62 and 0.80, as summarized in Table 4. Those studies were mostly performed on the MICCAI database, which had a relatively small sample size, that is, ~70 images. The current study included 455 images that were more than the sample sizes in previous studies in Table 4. We achieved the Dice coefficients of 0.81 on LR images (which matched the resolution of MICCAI datasets) and 0.83 on HR images. Though demonstrated on the MICCAI database, such a semisupervised method could be applied on other public brain MRI databases, for example, ADNI (Weiner et al., 2010), UK Biobank (Miller et al., 2016), and human connectome project (Van Essen et al., 2013), without the need of manual labels for WMH.

Level-set methods had been developed for many years for medical image segmentation (Li et al., 2010,2011 Wang et al., 2009; Zhang et al., 2015). In this study, the proposed LSLoss required the mean and standard deviation of foreground/background that were measured from 10 labeled images from the MICCAI datasets. WMH had uncertain number, shape, and volume, so we only utilized the foreground/background pixel intensities and the divergence of the level-set ϕ, which maintained the smoothness of the segmentation. Meanwhile, many level-set constraints could be used for segmentation, such as the foreground/background pixel intensities, morphological shapes, area/perimeter restrictions, and so forth. The proposed level-set loss can support those level-set constraints.

In this study, we attempted to use prior knowledge of the WMH to determine the four network parameters (μ_f, σ_f, μ_b, and σ_b), thus, ground truth labels are not necessary for training. It allows us to train a generic network model using a large amount of MR data collected from various centers. On the other hand, as proposed by Atlason et al. (2019), regular segmentation loss functions such as CELoss and DICE loss can be used if ground truth data are partially available. The combination of LSLoss and these regular segmentation losses allows the so-called transfer learning, which takes existing pretrained network architecture trained on the existing dataset (typically over a million) and then fine-tuned on specific data with human annotations (Chen et al., 2019).

In general semantic segmentation tasks, the four loss parameters μ_f, σ_f, μ_b, and σ_b can be set as learnable network parameters during training. But for WMHs segmentation, we fixed $\mu_f = 1.5$, $\sigma_f = 1.0$, $\mu_b = 1.0$, and $\sigma_b = 1.0$. The values were determined based on 10 randomly selected MICCAI images (from the training set with ground truth labels provided). According to our experiments, lower μ_f or higher μ_b might result in false-positive segmentation on health cases, as the network wrongly detect GM as WMH.

In Figure 7, we showed a few examples with high AVD, that is, the worst segmentation cases as indicated by the AVD scatter plots (Figure 6). The WMH of SVD cases (i.e., HKU-SVD and MICCAI WMH datasets) were slightly over-segmented. The predictions included ambiguous regions near the very bright WMH. In contrast, observers labeled those ambiguous regions as normal tissues. On the other hand, the WMH of MS cases (i.e., HKU- MS dataset) was under-segmented by the proposed method. We only included 15 MS cases for training, which might have caused the network to under-fit the MS datasets. The solution to this could be to include more MS cases in the training set. In Figure 8, V-Net$_{CE}$ missed some small WMH while V-Net$_{LS}$ detected them. We reasoned that small WMH was difficult to be detected by V-Net$_{CE}$ because the CE loss was not optimal

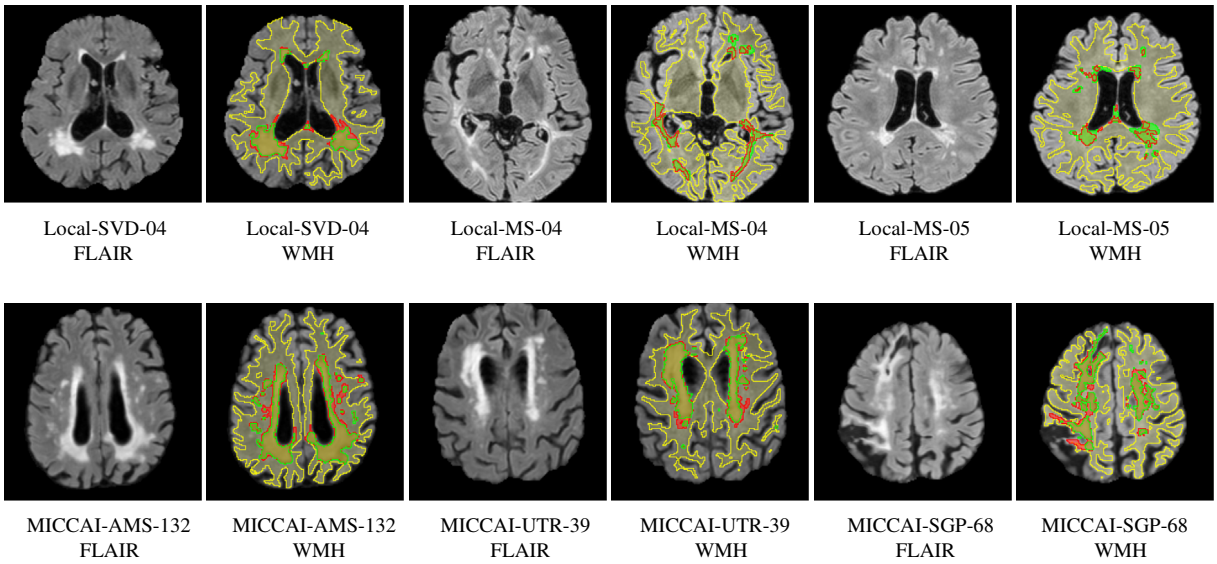

FIGURE 7 The images with high absolute volume difference segmentation in each database, which showed the worst cases for white matter hyperintensity (WMH) segmentation. The WMH were over-segmented on cerebral small vessel disease cases, and undersegmented on multiple sclerosis cases. However, most of WMHs were detected in the slices presented. The yellow masks were the WM$_{VNet}$, the green areas were the ground truth labels for WMHs, and the red areas were the WMH segmentation from V-Net$_{LS}$

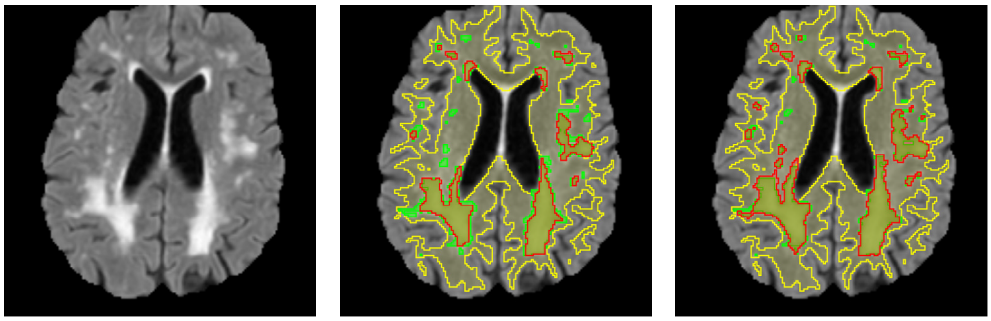

FIGURE 8 The comparison of white matter hyperintensity (WMH) segmentation obtained by V-Net$_{CE}$ (trained with cross-entropy loss) and V-Net$_{LS}$ (trained with level-set loss). Small WMHs were mostly detected by VNet$_{LS}$ (six individual small lesions were missed), while they were missed by V-Net$_{CE}$ (12 were missed VNet$_{LS}$ [6 individual small lesions were missed], while they were missed by V-Net$_{CE}$ (12 were missed)

for small WMH. For V-Net$_{LS}$, since the LSLoss could utilize as much training data as possible, small WMH were correctly detected.

Our experiment applied the ROI mask on the raw network segmentation to remove false positives outside the WM region. In Table 3, we reported the WMH segmentation performance of using no WM mask, WM$_{FSL}$ mask, and WM$_{VNet}$ mask. The overall DSCs calculated with WM$_{FSL}$ and WM$_{VNet}$ mask were generally better than the results using no WM masks. This was because the raw WMH segmentation included some false positives outside the WM region. The GM tissue on FLAIR images could have a similar intensity as WMH and could be a confounding factor in our intensity-based approach. However, those false positives were eliminated by an additional WM mask. Another solution might be adding more control cases (such as the ADNI-CN datasets in this study) to the training set to improve the network generalizability for normal brain tissue. Since this study mainly focused on detecting WMH in patients with SVD (360 cases), we limited the number of health cases (15 cases) during the training.

The second study included 20 brain data of MS patients when training the V-Net$_{LS}$. Some MS-specific WMH were with ovoid appearance (named Dawson's fingers) and located in various locations such as adjacent to the temporal lobe, cortex, and corpus callosum areas near the ventricle (Thompson et al., 2018). While the WMH of SVD had the presumed vascular origin and bilateral distribution (Wardlaw et al., 2013). The V-Net$_{LS}$ achieved overall DSC = 0.77, HD95 = 6.14, AVD = 25.49, TPR = 0.70, and F1score = 0.67. The DSC of WMH segmentation on MS cases was comparable to other testing cases of HKU-SVD, while the TPR and F1score were lower. This was because some MS-specific WMH were missed, as we only included 15 MS data in the training set. This issue could be resolved by including more MS data in the training set.

This study showed that the proposed method had good accuracy in WMH detection and segmentation. It does not require ground truth labels for training, such that data of multicenter, various MR scanners, and different disease cohorts can be used for the training. Besides WMH, it can be further extended to the detection and segmentation for other SVD subtypes. WMH, lacunes, cerebral microbleeds, and visible perivascular spaces (PVS) were four closely correlated cerebral vascular pathologies of SVD (Pantoni, 2010; Wardlaw et al., 2013). Recent studies suggested that a total SVD score that summarized the presence of all these SVD subtypes might capture the overall effect of SVD on the brain better than by considering them individually (Lau et al., 2017; Staals et al., 2014). Lacunes were frequently seen in elderly patients with no symptoms and were associated with an increased risk of stroke, gait impairment, and dementia (Santos et al., 2009). Lacunes with presumed vascular origin generally had a central CSF-like hypointense with a surrounding rim of WMH (Wardlaw et al., 2013). As a result, the LSLoss function can be designed accordingly, where the foreground intensity μ_f be a hypointense value, the background μ_b be a hyperintense value, and the ROI mask could be a WMH mask obtained by the V-Net$_{LS}$. PVS were the extensions of the extracerebral fluid space around arteries, arterioles, veins, and venules (Bouvy et al., 2016). They were defined as the fluid-filled spaces that follow the typical course as it goes through the GM or WM (Wardlaw et al., 2013). The spaces had a similar appearance as CSF on FLAIR and T2W images. Since they followed the course of penetrating vessels, they usually have elongated shapes parallel to the vessels. Thus, a novel loss term that describes this type of morphological shape can be added to the LSLoss function to train a PVS detection network. In summary, the proposed LSLoss can be used to assess the total SVD scores, and other neurological diseases that require the quantitative analysis on the cerebral vascular pathologies of SVD.

5 | CONCLUSION

In conclusion, we developed a semisupervised loss function for WMH segmentation. We used 400 unlabeled MRI images for model training. The overall performance on the public database outperformed the semisupervised and unsupervised techniques. Our study demonstrated the feasibility of using other semisupervised learning in WMH segmentation.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

The MRI data used in this study were from multiple sources, including our own institutional MRI unit and public available datasets. The HKU-SVD data are available on request from one of the coauthors, Dr Kui-Kai Lau (gkklau@hku.hk). The data are not publicly available due to the restriction of local research ethics committee (Institutional Review Board, IRB). The HKU-MS data are available on request from one of the co-authors, Dr Ka-Fung Mak (makkf@hku.hk). The data are not publicly available due to the restriction of local research ethics committee (Institutional Review Board, IRB). The MICCAI WMH Challenge (2017) dataset are publicly available in https://wmh.isi.uu.nl/. The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset are publicly available in http://adni.loni.usc.edu/.

ORCID

Peng Xia https://orcid.org/0000-0003-3113-7408
Henry Ka-Fung Mak https://orcid.org/0000-0002-2007-0650
Peng Cao https://orcid.org/0000-0002-1116-1171

REFERENCES

Adams, H. P., Jr., Bendixen, B. H., Kappelle, L. J., Biller, J., Love, B. B., Gordon, D. L., & Marsh, E., 3rd. (1993). Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of org 10172 in acute stroke treatment. *Stroke*, *24*(1), 35–41.

Akselrod-Ballin, A., Galun, M., Gomori, J. M., Filippi, M., Valsasina, P., Basri, R., & Brandt, A. (2009). Automatic segmentation and classification of multiple sclerosis in multichannel MRI. *IEEE Transactions on Biomedical Engineering*, *56*(10), 2461–2469.

Anbeek, P., Vincken, K. L., van Osch, M. J., Bisschops, R. H., & van der Grond, J. (2004). Automatic segmentation of different-sized white matter lesions by voxel probability estimation. *Medical Image Analysis*, *8*(3), 205–215.

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851.

Atlason, H. E., Love, A., Sigurdsson, S., Gudnason, V., & Ellingsen, L. M. (2019). SegAE: Unsupervised white matter lesion segmentation from brain MRIs using a CNN autoencoder. *NeuroImage: Clinical*, *24*, 102085.

Bastin, M. E., Sinha, S., Whittle, I. R., & Wardlaw, J. M. (2002). Measurements of water diffusion and T1 values in peritumoural oedematous brain. *Neuroreport*, *13*(10), 1335–1340.

Bouvy, W. H., Zwanenburg, J. J., Reinink, R., Wisse, L. E., Luijten, P. R., Kappelle, L. J., Geerlings, M. I., Biessels, G. J., & Ultra Vascular Cognitive Impairment (VCI) Study Group. (2016). Perivascular spaces on 7 tesla brain MRI are related to markers of small vessel disease but not to age or cardiovascular risk factors. *Journal of Cerebral Blood Flow & Metabolism*, *36*(10), 1708–1717.

Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D. A., Hernández, M. V., Wardlaw, J., & Rueckert, D. (2018). Gan augmentation: Augmenting training data using generative adversarial networks. arXiv preprint arXiv:1810.10863.

Brosch, T., Tang, L. Y., Yoo, Y., Li, D. K., Traboulsee, A., & Tam, R. (2016). Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Transactions on Medical Imaging*, *35*(5), 1229–1239.

Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., & Sudre, C. H. (2017). Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. *NeuroImage*, *148*, 77–102.

Chen, S., Ma, K., & Zheng, Y. (2019). Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625.

De Boer, R., Vrooman, H. A., Van Der Lijn, F., Vernooij, M. W., Ikram, M. A., Van Der Lugt, A., Breteler, M. M., & Niessen, W. J. (2009). White matter lesion extension to automatic brain tissue segmentation on MRI. *NeuroImage*, *45*(4), 1151–1161.

Driscoll, I., Davatzikos, C., An, Y., Wu, X., Shen, D., Kraut, M., & Resnick, S. (2009). Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology*, *72*(22), 1906–1913.

Ferrari, R. J., Wei, X., Zhang, Y., Scott, J. N., & Mitchell, J. R. (2003). Segmentation of multiple sclerosis lesions using support vector machines. Proceedings of SPIE - The International Society for Optical Engineering, 5032. https://doi.org/10.1117/12.481377

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved liver lesion classification. 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018).

Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, G., Plummer, D., Tofts, P., McDonald, W., & Miller, D. (1996). Quantification of MRI lesion load in multiple sclerosis: A comparison of three computer-assisted techniques. *Magnetic Resonance Imaging*, *14*(5), 495–505.

Guerrero, R., Qin, C., Oktay, O., Bowles, C., Chen, L., Joules, R., Wolz, R., Valdés-Hernández, M. D. C., Dickie, D. A., & Wardlaw, J. (2018). White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. *NeuroImage: Clinical*, *17*, 918–934.

Ithapu, V., Singh, V., Lindner, C., Austin, B. P., Hinrichs, C., Carlsson, C. M., Bendlin, B. B., & Johnson, S. C. (2014). Extracting and summarizing white matter hyperintensities using supervised segmentation methods in Alzheimer's disease risk and aging studies. *Human Brain Mapping*, *35*(8), 4219–4235.

Jain, S., Sima, D. M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., De Mey, J., Barkhof, F., Steenwijk, M. D., & Daams, M. (2015). Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage: Clinical*, *8*, 367–375.

Jenkinson, M., Pechaud, M., & Smith, S. (2005). BET2: MR-based estimation of brain, skull and scalp surfaces. Eleventh annual meeting of the organization for human brain mapping.

Jiang, W., Lin, F., Zhang, J., Zhan, T., Cao, P., & Wang, S. (2020). Deep-learning-based segmentation and localization of white matter hyperintensities on magnetic resonance images. *Interdisciplinary Sciences: Computational Life Sciences*, *12*(4), 438–446.

Kamber, M., Shinghal, R., Collins, D. L., Francis, G. S., & Evans, A. C. (1995). Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images. *IEEE Transactions on Medical Imaging*, *14*(3), 442–453.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, *36*, 61–78.

Khayati, R., Vafadust, M., Towhidkhah, F., & Nabavi, M. (2008). Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. *Computers in Biology and Medicine*, *38*(3), 379–390.

Kim, B., & Ye, J. C. (2019). Mumford–Shah loss functional for image segmentation with deep learning. *IEEE Transactions on Image Processing*, *29*, 1856–1866.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Klein, S., Staring, M., Murphy, K., Viergever, M. A., & Pluim, J. P. (2009). Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, *29*(1), 196–205.

Kuijf, H. J., Biesbroek, J. M., De Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M. J., & Casamitjana, A.

(2019). Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge. *IEEE Transactions on Medical Imaging*, *38*(11), 2556–2568.

Lau, K. K., Li, L., Schulz, U., Simoni, M., Chan, K. H., Ho, S. L., Cheung, R. T. F., Küker, W., Mak, H. K. F., & Rothwell, P. M. (2017). Total small vessel disease score and risk of recurrent stroke: Validation in 2 large cohorts. *Neurology*, *88*(24), 2260–2267.

Lau, K. K., Lovelock, C. E., Li, L., Simoni, M., Gutnikov, S., Küker, W., Mak, H. K. F., & Rothwell, P. M. (2018). Antiplatelet treatment after transient ischemic attack and ischemic stroke in patients with cerebral microbleeds in 2 large cohorts and an updated systematic review. *Stroke*, *49*(6), 1434–1442.

Li, C., Huang, R., Ding, Z., Gatenby, J. C., Metaxas, D. N., & Gore, J. C. (2011). A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. *IEEE Transactions on Image Processing*, *20*(7), 2007–2016.

Li, C., Xu, C., Gui, C., & Fox, M. D. (2010). Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, *19*(12), 3243–3254.

Li, G., Wang, L., Shi, F., Lyall, A. E., Lin, W., Gilmore, J. H., & Shen, D. (2014). Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. *Journal of Neuroscience*, *34*(12), 4228–4238.

Li, H., Jiang, G., Zhang, J., Wang, R., Wang, Z., Zheng, W.-S., & Menze, B. (2018). Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. *NeuroImage*, *183*, 650–665.

Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J. C., Quiles, A., Valls, L., Ramió-Torrentà, L., & Rovira, À. (2012). Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Information Sciences*, *186*(1), 164–185.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., & Wiest, R. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993–2024.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., & Andersson, J. L. (2016). Multimodal population brain imaging in the UK biobank prospective epidemiological study. *Nature Neuroscience*, *19*(11), 1523–1536.

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 fourth international conference on 3D vision (3DV).

Moeskops, P., Benders, M. J., Kersbergen, K. J., Groenendaal, F., de Vries, L. S., Viergever, M. A., & Išgum, I. (2015). Development of cortical morphology evaluated with longitudinal MR brain images of preterm infants. *PLoS One*, *10*(7), e0131552.

Moeskops, P., de Bresser, J., Kuijf, H. J., Mendrik, A. M., Biessels, G. J., Pluim, J. P., & Išgum, I. (2018). Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI. *NeuroImage: Clinical*, *17*, 251–262.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., Trojanowski, J. Q., Toga, A. W., & Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's & Dementia*, *1*(1), 55–66.

Pantoni, L. (2010). Cerebral small vessel disease: From pathogenesis and clinical characteristics to therapeutic challenges. *The Lancet Neurology*, *9*(7), 689–701.

Qian, W., Chan, K. H., Hui, E. S., Lee, C. Y., Hu, Y., & Mak, H. K. F. (2016). Application of diffusional kurtosis imaging to detect occult brain damage in multiple sclerosis and neuromyelitis optica. *NMR in Biomedicine*, *29*(11), 1536–1545.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. International conference on medical image computing and computer-assisted intervention.

Santos, M., Gold, G., Kövari, E., Herrmann, F. R., Bozikas, V. P., Bouras, C., & Giannakopoulos, P. (2009). Differential impact of lacunes and microvascular lesions on poststroke depression. *Stroke*, *40*(11), 3557–3562.

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., & Zimmer, C. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage*, *59*(4), 3774–3783.

Shamonin, D. P., Bron, E. E., Lelieveldt, B. P., Smits, M., Klein, S., & Staring, M. (2014). Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Frontiers in Neuroinformatics*, *7*, 50.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155.

Staals, J., Makin, S. D., Doubal, F. N., Dennis, M. S., & Wardlaw, J. M. (2014). Stroke subtype, vascular risk factors, and total MRI brain small-vessel disease burden. *Neurology*, *83*(14), 1228–1234.

Steenwijk, M. D., Pouwels, P. J., Daams, M., van Dalen, J. W., Caan, M. W., Richard, E., Barkhof, F., & Vrenken, H. (2013). Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). *NeuroImage: Clinical*, *3*, 462–469.

Sudre, C. H., Cardoso, M. J., Bouvy, W. H., Biessels, G. J., Barnes, J., & Ourselin, S. (2015). Bayesian model selection for pathological neuroimaging data applied to white matter lesion segmentation. *IEEE Transactions on Medical Imaging*, *34*(10), 2079–2102.

Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J. L., & Resnick, S. M. (2010). Longitudinal changes in cortical thickness associated with normal aging. *NeuroImage*, *52*(4), 1215–1223.

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., & Freedman, M. S. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, *17*(2), 162–173.

Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., Oliver, A., & Lladó, X. (2017). Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage*, *155*, 159–168.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & WU-MInn HCP Consortium. (2013). The WU-Minn human connectome project: An overview. *NeuroImage*, *80*, 62–79.

Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., & Suetens, P. (2001). Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, *20*(8), 677–688.

Viteri, J. A., Loayza, F. R., Peláez, E., & Layedra, F. (2021). Automatic brain white matter hypertinsities segmentation using deep learning techniques. HEALTHINF.

Wang, L., Li, C., Sun, Q., Xia, D., & Kao, C.-Y. (2009). Active contours driven by local and global intensity fitting energy with application to brain MR image segmentation. *Computerized Medical Imaging and Graphics*, *33*(7), 520–531.

Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R. I., O'Brien, J. T., Barkhof, F., & Benavente, O. R. (2013). Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *The Lancet Neurology*, *12*(8), 822–838.

Wardlaw, J. M., Valdés Hernández, M. C., & Muñoz-Maniega, S. (2015). What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. *Journal of the American Heart Association*, *4*(6), e001140.

Weiner, M. W., Aisen, P. S., Jack, C. R., Jr., Jagust, W. J., Trojanowski, J. Q., Shaw, L., Saykin, A. J., Morris, J. C., Cairns, N., & Beckett, L. A. (2010). The Alzheimer's disease neuroimaging initiative: Progress report and future plans. *Alzheimer's & Dementia*, *6*(3), 202–211.e207.

Wu, D., Albert, M., Soldan, A., Pettigrew, C., Oishi, K., Tomogane, Y., Ye, C., Ma, T., Miller, M. I., & Mori, S. (2019). Multi-atlas based detection and

localization (MADL) for location-dependent quantification of white matter hyperintensities. *NeuroImage: Clinical*, *22*, 101772.

Wu, J., Zhang, Y., Wang, K., & Tang, X. (2019). Skip connection U-net for white matter hyperintensities segmentation from MRI. *IEEE Access*, *7*, 155194–155202.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, *31*(3), 1116–1128.

Zhang, K., Zhang, L., Lam, K.-M., & Zhang, D. (2015). A level set approach to image segmentation with intensity inhomogeneity. *IEEE Transactions on Cybernetics*, *46*(2), 546–557.

Zhang, Y., Brady, J. M., & Smith, S. (2000). Hidden Markov random field model for segmentation of brain MR image. Medical Imaging 2000: Image Processing.

Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, *15*(5), 749–753.

Zhao, L., Biesbroek, J. M., Shi, L., Liu, W., Kuijf, H. J., Chu, W. W., Abrigo, J. M., Lee, R. K., Leung, T. W., & Lau, A. Y. (2018). Strategic infarct location for post-stroke cognitive impairment: A multivariate lesion-symptom mapping study. *Journal of Cerebral Blood Flow & Metabolism*, *38*(8), 1299–1311.

Zhou, P., Liang, L., Guo, X., Lv, H., Wang, T., & Ma, T. (2020). U-net combined with CRF and anatomical based spatial features to segment white matter hyperintensities. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).

How to cite this article: Huang, F., Xia, P., Vardhanabhuti, V., Hui, S.-K., Lau, K.-K., Ka-Fung Mak, H., & Cao, P. (2023). Semisupervised white matter hyperintensities segmentation on MRI. *Human Brain Mapping*, *44*(4), 1344–1358. https://doi.org/10.1002/hbm.26109