Deep-Learning-Based Segmentation and Localization of White Matter Hyperintensities on Magnetic Resonance Images

Wenhao Jiang, Fengyu Lin, Jian Zhang, Taowei Zhan, Peng Cao & Silun Wang

Interdisciplinary Sciences: Computational Life Sciences

ISSN 1913-2751
Volume 12
Number 4

Interdiscip Sci Comput Life Sci (2020)
12:438-446
DOI 10.1007/s12539-020-00398-0

Springer

Your article is protected by copyright and all rights are held exclusively by International Association of Scientists in the Interdisciplinary Areas. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Author's personal copy

Interdisciplinary Sciences: Computational Life Sciences (2020) 12:438–446
https://doi.org/10.1007/s12539-020-00398-0

ORIGINAL RESEARCH ARTICLE

Deep-Learning-Based Segmentation and Localization of White Matter Hyperintensities on Magnetic Resonance Images

Wenhao Jiang[1,2,3] · Fengyu Lin[1] · Jian Zhang[1] · Taowei Zhan[1] · Peng Cao[2] · Silun Wang[1]

Received: 26 August 2019 / Revised: 16 June 2020 / Accepted: 24 September 2020 / Published online: 2 November 2020
© International Association of Scientists in the Interdisciplinary Areas 2020

Abstract

White matter magnetic resonance hyperintensities of presumed vascular origin, which could be widely observed in elderly people, and has significant importance in multiple neurological studies. Quantitative measurement usually relies heavily on manual or semi-automatic delineation and intuitive localization, which is time-consuming and observer-dependent. Current automatic quantification methods focus mainly on the segmentation, but the spatial distribution of lesions plays a vital role in clinical diagnosis. In this study, we implemented four segmentation algorithms and compared the performances quantitatively and qualitatively on two open-access datasets. The location-specific analysis was conducted sequentially on 213 clinical patients with cerebral ischemia and lacune. The experimental results suggest that our deep-learning-based model has the potential to be integrated into the clinical workflow.

Keywords White matter hyperintensities · Neuroimaging · MRI · Segmentation · Localization

1 Introduction

White matter hyperintensities (WMH) are abnormal brain regions those exhibit relatively higher intensity level on T2-weighted and fluid-attenuated inversion recovery (FLAIR) magnetic resonance images (MRIs). These regions are widely observed in brain MRI of aging individuals and patients with cognitive decline [1]. In clinical practice, it is difficult to visualize small vessels in vivo. Hence, the observation of WMH is widely adopted as markers to identify cerebral small vessel diseases (cSVD), which is a significant contributor to mixed dementia and stroke [2]. Nevertheless, diseases such as multiple sclerosis [3], bipolar disorders [4] and autoimmune diseases [5] are also tightly associated with WMH. Therefore, the quantification of WMH is crucial in pathological diagnosis.

The US National Institute of Neurological Disorders and Stroke and the Canadian Stroke Network have proposed neuroimaging consensus standards for SVD classification [6]. However, manual delineation of WMH is a time-consuming and observer-dependent task that is hard to be standardized. Experienced radiologists and neuroscientists are highly demanded to make a reliable diagnosis, which is becoming a barrier to clinical diagnosis. In this way, an automatic schema that localizes, delineates, and analyzes the localization-dependent WMH is highly demanded to ease the burden of clinical ex-perts, and provide aids in diagnosis.

Multiple studies have been carried out to localize and segment the WMH automatically or semi-automatically. A logistic-regression-embedded toolkit named lesion prediction algorithm (LPA) [7] is proposed to segment T2 hyperintense lesions on FLAIR images. A Lesion Growth Algorithm (LGA) [8] provided a semi-automatic approach to segment lesion areas on T1-weighted and FLAIR images, by manually setting the initial threshold and parameters of Markov random field. W2MHS [9] is proposed as a WMH segmentation toolkit, which is based on random forest and support vector machines. An open-source library for hyperintensities detection and accumulation was published in this work. A

Wenhao Jiang and Fengyu Lin contributed equally to the work.

✉ Silun Wang
lawrence.wang@szdrbrain.com

Peng Cao
caopeng1@hku.hk

[1] YIWEI Medical Technology Co., Ltd, Room 1001, MAI KE LONG Building, Nanshan, ShenZhen 518000, China

[2] Department of Diagnostic Radiology, The University of Hong Kong, Hong Kong 999077, SAR, China

[3] School of Computer Science, Harbin Institute of Technology, Harbin 150001, China

Author's personal copy

multi-atlas based detection and localization (MADL) [10] is carried out to parcellate the brain and identify WMH in individual parcels simultaneously. The segmentation was performed using a Bayes maximum a posteriori (MAP) estimation and thresholding. Recent years has also seen increasingly more works promoting the WMH segmentation by releasing public datasets. A public MR image dataset [11] of 30 MS patients with expert delineated labels was online for algorithm training and testing. A subset [12] with a total of 20 patients from the Alzheimer Disease Neuroimaging Initiative (ADNI) [13] was semi-automatically labelled using region growing algorithm and released for studying the WMH of patients with Alzheimer Disease (AD). The WMH Segmentation Challenge [14] was held in conjunction with the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2017). The challenge released 60 cases of MR images from three hospitals as a training set, with manually labeled WMH contours.

Among the most recent works based on the public datasets, deep learning has seen great potential in outperforming conventional computer vision-based strategies. A U-Net [15] based fully convolutional neural network (CNN) structure [16] was implemented on the MICCAI WMH dataset and ranked 1st in the challenge, with an average dice score of 80% on the validation set. A U-ResNet [17] algorithm was applied to segment WMH and distinguish stroke, taking advantage of both U-Net and ResNet [18]. CNN was also performed on the WMH segmentation of ADNI scans [13]. The results indicate that deep-learning-based method outperformed conventional machine learning methods by excluding artefacts and pathologies, which seemed like WMH.

Though the segmentation of WMH is well explored, the study of location dependent WMH is yet to be made. Location information of WMH is of great importance in diagnosis. For example, profuse patches of WMH in deep white matter (dWM) usually indicate ischemic, while periventricular caps and smooth halos are believed to be the regions of demyelination associated with subependymal gliosis and discontinuity of the ependymal lining [19]. Therefore, embedding location information into segmentation helps disease classification and diagnosis.

There is accumulated evidence that the multi-channel deep-learning-based algorithm DeepMedic [20] is efficient in brain lesion segmentation related tasks, such as brain tumors, stroke lesions, and traumatic brain injuries (TBI). However, due to the differences in lesion size and location, the capability of DeepMedic in WMH segmentation remains to be validated. In this study, we intent to facilitate the automatic segmentation of WMH by utilizing deep-learning algorithms like DeepMedic. Moreover, a Dice coefficient loss described in the V-Net paper [21] is adopted as the loss function to guide the training. Comparisons among the DeepMedic model and other state-of-the-art lesion segmentation models were made on two open-access datasets, namely the WMH segmentation challenge [14] training set and the MS patient dataset [11] both of which contains high-quality T1-weighted and FLAIR images and professionally labeled delineations. Sequentially, the trained model is implemented on the clinical cohort to segment the WMH. The cohort comprises a total of 213 sets of T1 and FLAIR images from patients who were diagnosed with cerebral ischemia and lacune. We further register the FLAIR images of each patient along with the segmented regions of the WMH into the same brains in Montreal Neurological Institute (MNI) coordinates and generate the overall lesion prevalence map regarding the WMH segmentation distribution. Finally, a Hammer atlas [22] is adopted to segment the coregistered images into 68 regions of interest (ROIs), where the location-specified WMH volume distribution are analyzed. The experimental WMH segmentation results suggest that our modified model outperformed other existing methods. The prevalence map reveals the vulnerability of each brain areas against different diseases. The ROI-specific correlation analysis indicates the inner-correlation among individual brain areas with regard to WMH load.

This work has its contributions in three ways.

1. A modified deep-learning-based convolutional neural network is proposed to segment WMH from FLAIR and T1 images.
2. Distribution of the WMH is studied to identify the vulnerable areas.
3. Disease-specific associations among the vulnerable areas are investigated via covariance analysis.

2 Materials

2.1 Public Datasets

Two public datasets, namely the MICCAI WMH segmentation challenge data [14] and the public MS patient data [11], were utilized for training and testing the WMH segmentation deep learning model.

For the challenge data, a total of 60 sets of T1 and FLAIR images were collected from three individual sites with manual WMH segmentation labels. The T1-weighted images were all 3D sequences acquired in the sagittal direction. Meanwhile, the FLAIR images were 3D or multi-slice 2D sequences, that were acquired in the sagittal or transversal direction. The other cohort of 30 MS patients with white matter lesions was scanned by a 3T Siemens Magnetom Trio MR system. For each patient, a 2D T1-weighted sequence, a 2D T2-weighted sequence, and a 3D FLAIR sequence were collected. In order to make the data consistent with

Author's personal copy

Table 1 Parameters of the public datasets

	Machine	Patients	T1			FLAIR		
			Dim.	Sampling (mm)	TR/TE (ms)	Dim.	Sampling (mm)	TR/TE (ms)
WMH challenge	3T Philips	20	3	1.00 × 1.00 × 1.00	7.9/4.5	2	0.96 × 0.95 × 3.00	11,000/125
	3T Simens	20	3	1.00 × 1.00 × 1.00	2300/1.9	2	1.00 × 1.00 × 3.00	9000/82
	3T GE	20	3	0.94 × 0.94 × 1.00	7.8/3.0	3	0.98 × 0.98 × 1.20	8000/126
MS patient data	3T Philips	30	2	0.42 × 0.42 × 3.30	2000/20	3	0.47 × 0.47 × 0.80	5000/392

Dim number of dimensions

Table 2 Demographic information of the clinical cohorts

Age	Ischemia	Lacune	Total
31–40	2 (0/2)	2 (1/1)	4 (1/3)
41–50	17 (11/6)	7 (6/1)	24 (17/7)
51–60	23 (8/15)	41 (28/13)	64 (36/28)
61–70	38 (15/23)	25 (18/7)	63 (33/30)
71–80	14 (8/6)	19 (12/7)	33 (20/13)
81–90	2 (1/1)	22 (11/11)	24 (12/12)
91–100	0	1(1/0)	1 (1/0)
Total	96 (43/53)	117 (77/40)	213 (120/93)

Two numbers inside the brackets indicate the number of male and female patients

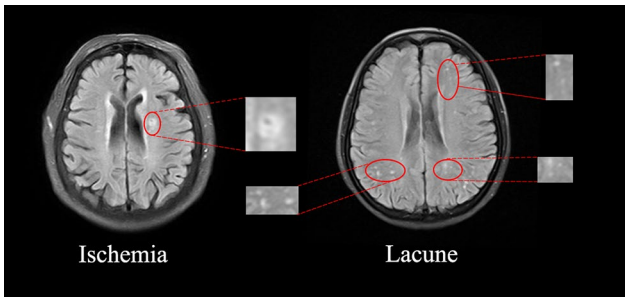


Fig. 1 Sample images of the clinical data. The circled areas indicate the white matter hyperintensities on the scan

the challenge dataset, only T1 and FLAIR sequences were used in training and testing of this WMH segmentation project. Details of the selected images from three sites are listed below in Table 1.

2.2 Clinical Datasets

The clinical dataset consists of retrospective T2-weighted FLAIR and T1-weighted scans from 213 subjects between 2014 and 2019. The personal information of the subjects was completely anonymous, and the data acquisition complies with the local ethical standards. The pathological manifestations of the cohort are ischemia and lacune. Table 2 shows the demographic information of this cohort.

2D MRI scans were obtained from two different scanners, namely SIEMENS Avanto and Philips Medical Systems. The MRI Image acquisition in this study is ongoing. The T1W images were acquired at 1.5T Siemens Avanto (repetition time (TR) = 550ms, echo time (TE) = 8.7 ms, flip angle = 90°, slice thickness = 5–6 mm) and 3T Philips Medical Systems (TR = 7500–8500 ms, TE = 92 ms, flip angle = 150°, slice thickness = 5–6 mm). The FLAIR images were acquired at 1.5T Siemens Avanto (TR = 250 ms, TE = 2.303 ms, flip angle = 75°, slice thickness = 6 mm) and 3T Philips Medical Systems (TR = 11,000 ms, TE = 121 ms, flip angle = 90°, slice thickness = 6 mm).

Sample images of the clinical data are provided in Fig. 1. Lesion areas are marked with red ellipses in the figure. The ischemia lesion presents itself as orbicular hyperintensity, while lacune lesions are more separated and relatively more substantial in size.

3 Methods

The proposed method can be divided into three steps. Firstly, all images uniformly went through pre-processing operations, followed by a segmentation stage where the WMH were delineated. Afterward, WMH localization was carried out by registration and mapping to an existing atlas. The pipeline of this work is depicted in Fig. 2.

3.1 Data Preprocessing

To enhance the quality of the cohort images, three steps were conducted sequentially in the data preprocessing stage. Firstly, N4 bias correction [23] was performed to correct global inhomogeneity. Secondly, coregistration was achieved based on an interpolation method [24]. Thirdly, image intensities were normalized using a z score normalization strategy. Meanwhile, only z score normalization operation was applied to the WMH segmentation challenge and MS patient datasets, where the images have already been elaborated. Data augmentation, which is considered to be efficient in

Author's personal copy

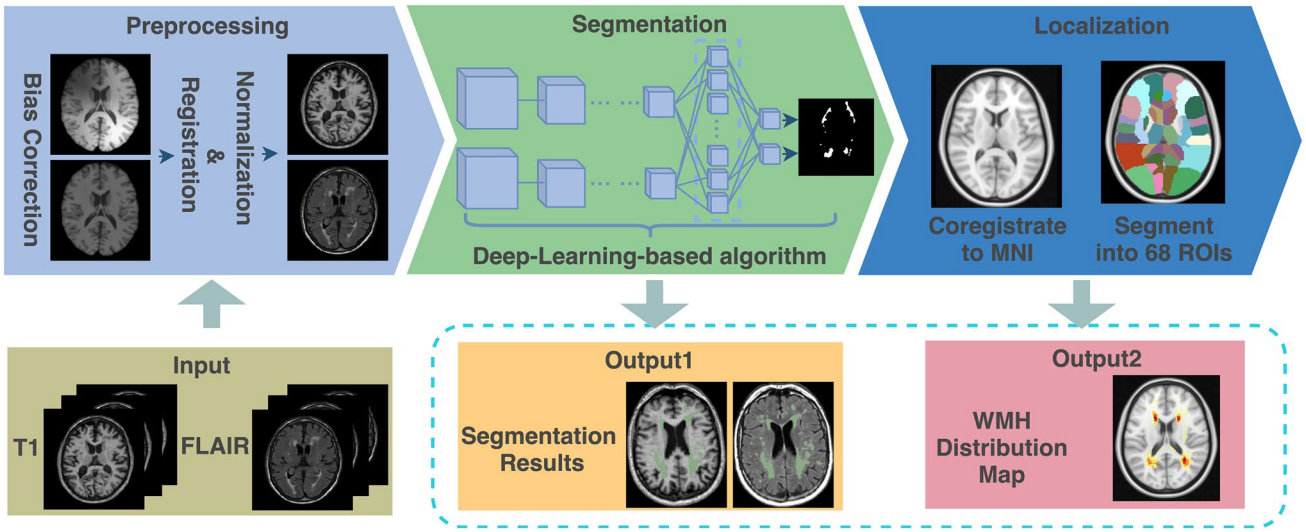


Fig. 2 The pipeline of the proposed segmentation and localization workflow. The method takes T1 and FLAIR images as inputs and outputs WMH segmentation on both T1 and FLAIR, as well as the distribution of WMH

reducing overfitting [25] was also operated to enrich the training set. Rotation (− 15°–15°), shearing (− 18°–18°) and scaling (0.9–1.1) were randomly performed on the two training datasets.

3.2 WMH Segmentation

Three deep-learning-based segmentation methods, namely DeepMedic [20], SYSU_Media [16], and Deep ResU-Net [26] were implemented and compared in this study. Details of these methods are elaborated as below.

DeepMedic The original multi-pathway convolutional layer structure of DeepMedic is adopted in this study. To incorporate both local and contextual information, the model is composed of multiple 3D convolution pathways, with one path extracting the spatial information from the original images and the others from down-sampled images. Specifically, each pathway includes 8 convolutional layers (with 3D 33 sized convolutional kernels), 2 fully connected layer (to combine normal and sub-sampled pathways), and 1 segmentation layer (final output). Dropout, L1 and L2 regularization are utilized to alleviate the overfitting problem. No pooling layer is needed in this model, as the layer has been proved to the cause of some spatial invariance undesirable for lesion segmentation [20]. Moreover, the relatively small size of WMH within the whole brain contributes to the unbalanced distribution of training samples. To tackle this problem, we tailored the loss function and introduced a soft dice coefficient loss [21], which considers the foreground and background separately. The soft dice coefficient is defined as Eq. 1, where y_{true} and y_{pred} denote the true and predicted value of pixel y and Loss_{sd} indicates the soft dice coefficient.

$$\text{Loss}_{\text{sd}} = \frac{2\sum_{\text{pix}} y_{\text{true}} y_{\text{pred}}}{\sum_{\text{pix}} y_{\text{true}}^2 + \sum_{\text{pix}} y_{\text{pred}}^2}. \quad (1)$$

SYSU_Media Similar to the DeepMedic, the SYSU_Media [16] model also used a fully convolutional neural net-work structure as the backbone. The network architecture is a 2D 19-layers deep convolutional network resembling U-Net. The improvement of this model is mainly reflected in bigger kernel size for the first two layers, which integrates different transformations. In addition, external preprocessing on top of the basic preprocessing is vital for SYSU_Media to guarantee uniformly sized inputs, where all the axial slices were automatically cropped or padded. Moreover, the first and last small proportion of the input images were discarded to avoid invariances from the skull and other irrelevant parts of the brain. The reason why 2D CNN was implemented in comparison with 3D CNN is that, although the 3D CNN structure take slice-wise information into consideration, resolutions vary across different scanners and could cause inconsistency. From this perspective, further comparisons between 2D and 3D models are required.

Deep ResU-Net The combination of residual block and U-Net has been proven effective on both generic images [26] and brain MRI [17]. Integrating the advantages of both U-Net and residual neural network, the ResU-Net model consists of three parts: encoder, bridge, and decoder, all of which are built with residual units. Two convolution blocks (including one batch normalization layer, one ReLU activation layer and one subsequent convolution layer) and an

Author's personal copy

identity mapping (connecting the input from the previous block and the output of the last convolution layer inside the block) were utilized in all of the three above-mentioned components. Comparing with the plain neural unit used in U-Net, the residual unit demonstrates its capability to alleviate the degradation problem in backward propagation. The implementation details could be found in the original Deep ResU-Net paper [26].

3.3 WMH Localization

To further study the association between the WMH load distribution and the occurrence of diseases, lesion localization is required. The lesion-symptom mapping toolkit regLSM [27] was used to register different brain images with the standard MNI template. In this study, the individual scans with segmented lesions are registered to a standard Montreal Neurological Institute (MNI)-152 template, based on which a WMH prevalence map is generated. The algorithms used in the regLSM are voxel-based LSM (SVR-LSM) and region of interest-based analyses (SVR-ROI) [27]. Sequentially, the Hammers atlas [22], which divides the brain into 68 ROIs, is utilized to analyze the distribution of WMH on the coregistered template and lesions.

4 Results and Discussion

4.1 WMH Segmentation

In the experiment part, all of the above-mentioned segmentation algorithms have been applied on two public datasets. The LPA algorithm, which is integrated into the LST, was set as the baseline. We compared the performance of the four methods qualitatively and quantitatively. Quantitative comparisons were made with regard to DSI (dice similarity index), recall, precision and the absolute log-transformed volume difference (lAVD). The DSI and lAVD is defined in Eqs. 2–3.

$$\mathrm{DSI} = \frac{2 \times (\text{Predicted Area} \cap \text{True Area})}{\text{Predicted Volume} + \text{True Volume}} \tag{2}$$

$$\mathrm{lAVD} = \left| log \frac{\text{Predicted Volume}}{\text{True Volume}} \right| \tag{3}$$

The segmentation performances of different models are summarized in Table 3. In total, 72 subjects were selected as training set, while 18 subjects were utilized as validation set. For the challenge dataset, since the 110 test samples are kept secret and currently not available, we split the 60 training images into a training set of 48 patients and a validation set of 12 patients. Both subsets were equally sampled from the three sites of the original data. For the MS dataset, 24 of the total 30 patients were randomly selected as the training set, while the other 6 as validation set. To illustrate the efficiency of soft dice loss function, we compared the outcomes by the models trained with the original cross entropy loss in DeepMedic with the exact same parameters (learning rate, epoch, etc.), shown in Table 3. As a result, utilizing soft dice loss subserve the WMH segmentation performance of the DeepMedic model.

Table 3 Performance comparison among different WMH segmentation models on the challenge and MS datasets

Model	DSI	Recall	Precision	lAVD
DeepMedic (soft dice)	**0.791**	0.799	**0.801**	**0.141**
DeepMedic (cross entropy)	0.725	0.778	0.714	0.279
SYSU Media	0.692	0.669	0.757	0.282
SYSU Media (ensemble)	0.789	**0.873**	0.733	0.219
Deep ResU-Net	0.732	0.752	0.756	0.293
LPA	0.615	0.755	0.555	0.402

Bold values indicate the significant difference in regarding of diagnostic performance

DSI Dice Similarity Index, *lAVD* absolute log-transformed volume difference

Figure 3 plotted the segmentation results of these four models on the public dataset. Note that, for the DeepMedic model, two down-sampling convolution pathways were implemented, with one pathway input being sub-sampled three, the other five times on the original images. The original 3D CNN structure of DeepMedic was used, with a soft dice coefficient as loss function. As for the SYSU_Media network, where 2D CNN is used instead of 3D, the input images were extracted from the original 3D data slice-by-slice with a unified cropping size of 200×200 pixels, siding with the original configures. Three U-Net-based models with randomly initialized parameters were trained separately with shuffled data and further averaged probability maps to generate an ensemble model. We compared segmentation performance of the SYSU_Media model with or without the ensemble in Table 3. Different from the original paper, in order to maintain the consistence of training data among the algorithms and the fairness of the comparison, the first and last 1/8 slices were not discarded. For the Deep ResU-Net model, six residual units were utilized in the encoding and the decoding pathways. The input shape of the model is set to be the same as SYSU_Media.

As have been shown in Table 3 and Fig. 3, the DeepMedic model significantly outperformed the SYSU_Media model, the Deep ResU-Net model and the LST baseline. One reason the DeepMedic model outperforming the others is the multi-convolutional-pathway schema that integrates the context and spatial information. Moreover, the 3D convolution in DeepMedic also contributes to the well

Author's personal copy

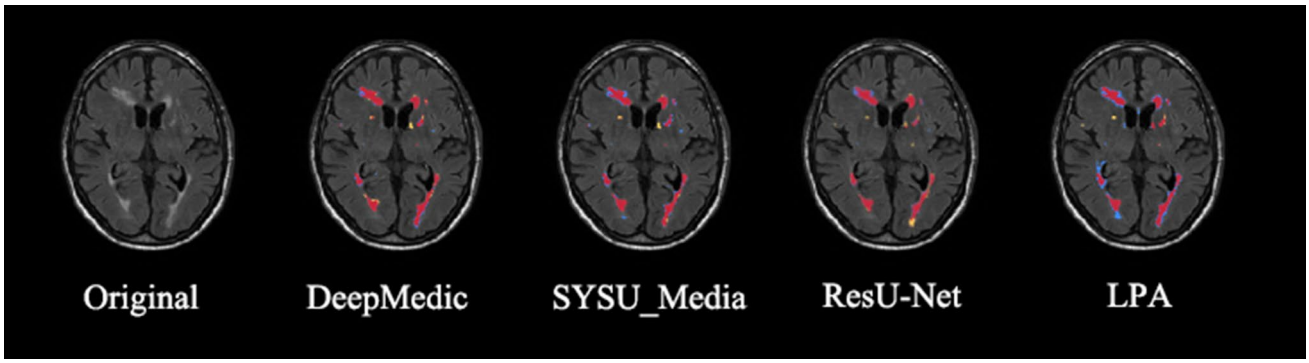


Fig. 3 The WMH segmentation results of the four algorithms. Red color indicates the segmented areas by the algorithm. Blue color means the WMH which appears on the manually labeled images but not on the predicted ones (false negative). Yellow areas are the predicted areas that did not appear on the manual labels (false positive)

capture of spatial features, amending the lack of slice-wise information caused by the conventional 2D convolution as in the SYSU_media and the Deep ResU-Net model. All these four algorithms behaved higher DSI on WMH segmentation challenge dataset comparing with the MS dataset. One possible explanation is the observer-dependence of the two manually labeled datasets could confuse the model. While the WMH segmentation challenge dataset takes the major proportion of training sets, the model tends to be more sensitive to the challenge data. To better illustrate the generalization capability of the champion model of this study, an extra leave-one-scanner-out (LOSO) experiment was conducted on the DeepMedic model. The data acquired with one random machine in the training set were selected as the test scans and remained unseen in the training stage, resulting in a DSI of 0.686, coincide with the previous experimental outcome. A rational explanation for the mild drop of DSI in this experiment was that by conducting LOSO, the number of training samples saw a major reduction, which is not sufficient enough comparing with the former attempts.

Aside from the comparison, a separate test was conducted on clinical cohort. The FLAIR images of five patients diagnosed with WMH were randomly selected from our clinical datasets and manually delineated by experienced radiologists. We tested the proposed model which has been previously trained on the two public datasets including the five images and got an average DSI value of 0.670, precision of 0.739, and recall of 0.640. A sample of the test results is demonstrated in Fig. 4a. Possible reasons for the mild degradation of segmentation performance are: the size and resolution differences between datasets, pathological differences among subjects (the test cohort were all diagnosed with cerebral ischemia or lacune), racial differences (the test subjects were all Asian while the public datasets picked up an amalgam of ethnicities).

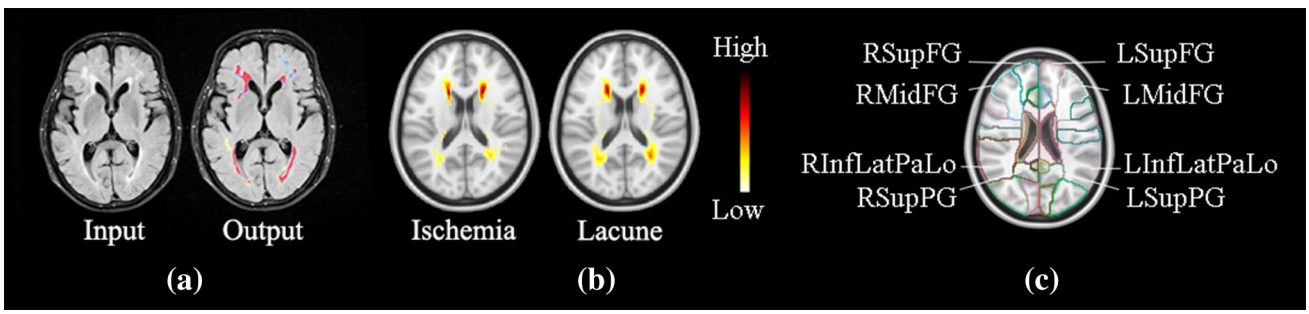


Fig. 4 **a** Original FLAIR image and segmentation result using DeepMedic on clinical cohort. Red color indicates the segmented areas by the algorithm. Blue color means the WMH which appears on the manually labeled images but not on the predicted ones (false negative). Yellow areas are the predicted areas that did not appear on the manual labels (false positive). **b** The lesion prevalence map of WMH. From left to right are the WMH load of: (1) patients with ischemia; (2) patients with lacune. The color indicates the rate of prevalence. Red indicates a high rate, while yellow indicates low. **c** A sample slice in the Hammer atlas. *RSupFG* right superior frontal gyrus, *LSupFG* left superior frontal gyrus, *RMidFG* right middle frontal gyrus, *LMidFG* left middle frontal gyrus, *RInfLatPalo* right inferior lateral parietal lobe, *LInfLatPalo* left inferior lateral parietal lobe, *RSupPG* right superior parietal gyrus, *LSupPG* left superior parietal gyrus

Author's personal copy

The implementation details of the DeepMedic model were described as below. An Adam optimizer with a learning rate of 0.001 was adopted to update the hyper-parameters in the training stage. Drop-out was operated with a rate of 0.5 to avoid over-fitting. The L1 (10^{-6}) and L2 (10^{-4}) regularization were also implemented to create sparsity and constrain the weight. For each iteration in the training stage, patches sized $37 \times 37 \times 37$ were fed to the model with a batch size of 10. The training and validation was carried out on a Linux workstation with Nvidia GTX 1080 Ti graphic processing unit.

Table 4 The brain volume and WMH load of ten brain regions

Region	Total vol.(mL)	Isc. WMH(mL)	Lac. WMH(mL)
LMFG	67.284	0.916	1.289
LSFG	85.367	0.658	0.862
LPTL	63.077	0.387	0.587
LSPG	60.885	0.412	0.775
LPG	49.730	0.401	0.730
RMFG	66.609	0.890	1.449
RSFG	82.371	0.578	0.683
RPTL	63.052	0.472	0.668
RSPG	60.713	0.433	0.970
RPG	49.422	0.426	0.852

LMFG left middle frontal gyrus, *LSFG* left superior frontal gyrus, *LPTL* left posterior temporal lobe, *LSPG* left superior parietal gyrus, *LPG* left precentral gyrus, *RMFG* right middle frontal gyrus, *RSFG* right superior frontal gyrus, *RPTL* right posterior temporal lobe, *RSPG* right superior parietal gyrus, *RPG* right precentral gyrus, *Isc* ischemia, *Lac* lacune

4.2 WMH Localization

The prevalence map of WMH obtained by the DeepMedic model on clinical data was generated based on the segmentation results. As shown in Fig. 4b, the distribution of WMH is mainly localized in the regions of frontal gyrus, whereas a small amount of WMH distributed in regions of superior parietal gyrus. WMH presents similar distribution in different kinds of lesion type, including ischemic lesions and lacune infarction. Figure 4c demonstrated a sample slice within the Hammers atlas [22], which divided the brain into 68 ROIs. An observation could be made from the figure, that the most prominent areas in the prevalence map fall in the middle frontal gyrus and the superior parietal gyrus area.

We then ranked all the 68 brain regions regarding the segmented WMH load and listed the top ten ROIs in Table 4. The table contains the total volume and the WMH load of the two populations for each brain region. From the table, we can conclude that the WMH load of lacune patients is higher than the ones with ischemia in all the listed ten areas.

The brain-wide correlation matrices on the ischemia patients and lacune patients were demonstrated in Fig. 5a. The *x*-axis and *y*-axis in the matrix indicate the individual brain regions. The sorting order of the matrix aligns with the indices of each region defined in the original atlas template.

The differential map subtracting the two matrices was provided in Fig. 5b. The two figures demonstrated the WMH correlation of each brain regions and their differential on different diseases. Several observations could be made based on the figure. Firstly, the high correlation coefficient between the frontal gyrus areas (around 60) and the superior parietal gyrus areas (around 30) could be witnessed on both ischemia and lacune matrices. The observation is in accordance with the outcomings from the above figure. Secondly, the correlation between

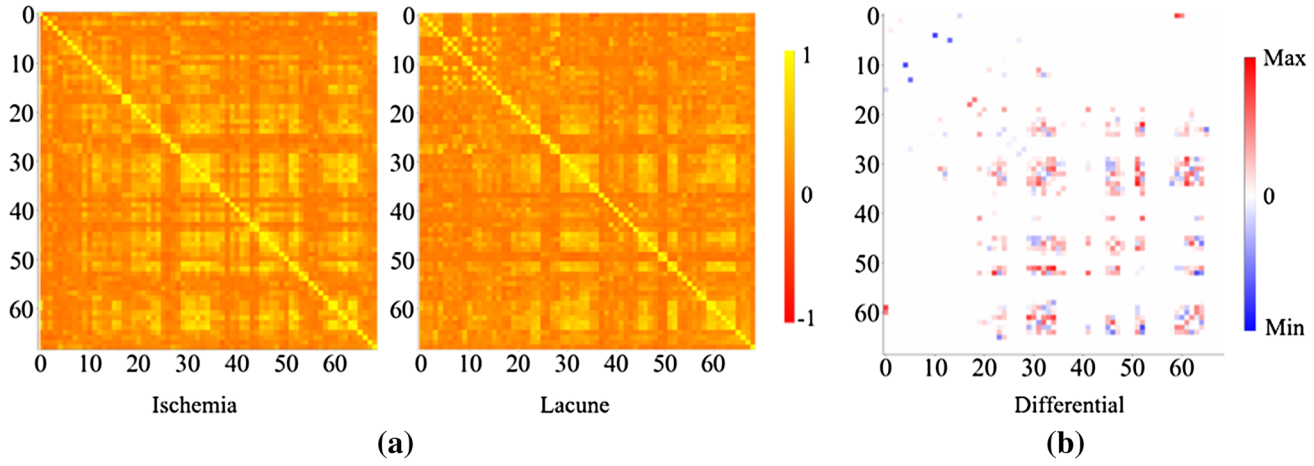


Fig. 5 **a** The brain-wide correlation matrices of patients with ischemia and lacune. Red indicates negative correlation while yellow indicates positive. **b** The differential map, subtracting the left matrices. Red and blue suggest positive and negative value, respectively

Author's personal copy

the amygdala and the superior temporal sulcus and gyrus is higher in lacune patients than in ischemia patients. This could be observed through the blue portion and relatively brighter color on the top left corner (around 10) of the differential map and lacune matrix. A feasible explanation is that social communication and cognitive disorders, which are highly associated with the above-mentioned areas, are more frequent among lacune patients. Besides, the wide occurrences of red and blue blobs indicate the difference in structural correlation between lacune patients and ischemia patients.

Note that, in order to maintain the significance in Fig. 5b, Bonferroni correction was applied to the correlation coefficient. To be specific, only the correlations with $p_\text{value} > 0.05/69$ (68 brain regions and background) were adopted.

5 Conclusion

In this study, comparisons of four lesion segmentation algorithms were made on the two accessible public datasets. Training and testing were conducted on two datasets with delineated WMH lesions. We creatively combined the Deep-Medic model with the soft dice coefficient loss function, resulting in the average DSI of 0.832 on the training set of WMH segmentation challenge, and 0.709 on the MS dataset. Location-specific analysis of two diseases, namely the ischemia and lacune, was carried out based on the WMH volume of individual brain areas. Associations among WMH load of different brain areas were found based on the localization results. Our outcomings proved the efficiency of the proposed segmentation and localization methods.

In future work, we intend to expand the scale of the in-house dataset by increasing the number of cohorts and extending the list of the investigated diseases. Another potential direction is to further classify diseases concerning the distribution of WMH distribution in each ROI. By encoding the spatial distribution of WMH, the classification model has the potential to make a reasonable diagnosis based on the prior knowledge and the training samples.

Acknowledgements This manuscript has been accepted for presentation in The Fourth CCF Bioinformatics Conference (CBC 2019).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R, Lindley RI, O'Brien JT, Barkhof F, Benavente OR et al (2013) Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol 12(8):822–838
2. Pantoni L (2010) Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. Lancet Neurol 9(7):689–701
3. Ge Y (2006) Multiple sclerosis: the role of MR imaging. Am J Neuroradiol 27(6):1165–1176
4. Kempton MJ, Geddes JR, Ettinger U, Williams SCR, Grasby PM (2008) Meta-analysis, database, and metaregression of 98 structural imaging studies in bipolar disorder. Arch Gen Psychiatry 65(9):1017–1032
5. Theodoridou A, Settas L (2006) Demyelination in rheumatic diseases. J Neurol Neurosurg Psychiatry 77(3):290–295
6. Hachinski V, Iadecola C, Petersen RC, Breteler MM, Nyenhuis DL, Black SE, Powers WJ, DeCarli C, Merino JG, Kalaria RN et al (2006) National institute of neurological disorders and stroke—Canadian stroke network vascular cognitive impairment harmonization standards. Stroke 37(9):2220–2241
7. Schmidt P (2017) Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging. Ph.D. thesis, lMU
8. Schmidt P, Gaser C, Arsic M, Buck D, Förschler A, Berthele A, Hoshi M, Ilg R, Schmid VJ, Zimmer C et al (2012) An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. Neuroimage 59(4):3774–3783
9. Ithapu V, Singh V, Lindner C, Austin BP, Hinrichs C, Carlsson CM, Bendlin BB, Johnson SC (2014) Extracting and summarizing white matter hyperintensities using supervised segmentation methods in Alzheimer's disease risk and aging studies. Hum Brain Mapp 35(8):4219–4235
10. Wu D, Albert M, Soldan A, Pettigrew C, Oishi K, Tomogane Y, Ye C, Ma T, Miller MI, Mori S (2019) Multi-atlas based detection and localization (MADL) for location-dependent quantification of white matter hyperintensities. Neuroimage Clin 22:101772
11. Lesjak Ž, Galimzianova A, Koren A, Lukin M, Pernuš F, Likar B, Špiclin Ž (2018) A novel public MR image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus. Neuroinformatics 16(1):51–63
12. Rachmadi MF, Vald'es-Hern'andez MC, Agan MLF, Perri C, Komura T, Alzheimer's Disease Neuroimaging Initiative et al (2018) Segmentation of white matter hyperintensities using convolutional neural networks with global spatial information in routine clinical brain MRI with none or mild vascular pathology. Comput Med Imaging Gr 66:28-43
13. Clifford JR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C et al (2008) The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging 27(4):685–691
14. Kuijf HJ, Biesbroek JM, de Bresser J, Heinen R, Andermatt S, Bento M, Berseth M, Belyaev M, Cardoso MJ, Casamitjana A et al (2019) Standardized assessment of automatic segmentation of white matter hyperintensities; results of the WMH segmentation challenge. IEEE Trans Med Imaging 38(11):2556–2568
15. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computerassisted intervention. Springer, pp 234–241
16. 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
17. Guerrero R, Qin C, Oktay O, Bowles C, Chen L, Joules R, Wolz R, Vald'es-Hernández MC, Dickie DA, Wardlaw J et al (2018) White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. Neuroimage Clin 17:918–934

Author's personal copy

18. He K, Zhang X, Ren S, Jian S (2016) Deep residual learning for image recognition. In: The IEEE conference on computer vision and pattern recognition (CVPR)
19. Kim KW, MacFall JR, Payne ME (2008) Classification of white matter lesions on magnetic resonance imaging in elderly persons. Biol Psychiatry 64(4):273–280
20. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B (2017) Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 36:61–78
21. Milletari F, Navab N, Ahmadi S-A (2016) Vnet: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). IEEE, pp 565–571
22. Hammers A, Allom R, Koepp MJ, Free SL, Myers R, Lemieux L, Mitchell TN, Brooks DJ, Duncan JS (2003) Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp 9(4):224–247
23. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC (2010) N4itk: improved n3 bias correction. IEEE Trans Med Imaging 29(6):1310
24. Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G (1995) Automated multimodality image registration based on information theory. Inf Process Med Imaging 3:263–274
25. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097-1105
26. Zhang Z, Liu Q, Wang Y (2018) Road extraction by deep residual u-net. IEEE Geosci Remote Sens Lett 15(5):749–753
27. Zhao L, Biesbroek JM, Shi L, Liu W, Kuijf HJ, Chu WWC, Jill MA, Lee RKL, Leung TWH, Lau AYL et al (2018) Strategic infarct location for post-stroke cognitive impairment: a multivariate lesion-symptom mapping study. J Cereb Blood Flow Metab 38(8):1299–1311