Computer Methods and Programs in Biomedicine 183 (2020) 105065

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Two-step deep neural network for segmentation of deep white matter hyperintensities in migraineurs

Jisu Hong[a,b], Bo-yong Park[a,b], Mi Ji Lee[c], Chin-Sang Chung[c], Jihoon Cha[d], Hyunjin Park[b,e,*]

[a] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea
[b] Center for Neuroscience Imaging Research, Institute for Basic Science (IBS), Suwon 16419, South Korea
[c] Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, South Korea
[d] Department of Radiology, Severance Hospital, Research Institute of Radiological Science, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[e] School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon 16419, South Korea

ARTICLE INFO

Article history:
Received 24 April 2019
Revised 12 August 2019
Accepted 4 September 2019

Keywords:
Deep white matter hyperintensity
Segmentation
Deep neural network
Migraine

ABSTRACT

Background and Objective: Patients with migraine show an increased presence of white matter hyperintensities (WMHs), especially deep WMHs. Segmentation of small, deep WMHs is a critical issue in managing migraine care. Here, we aim to develop a novel approach to segmenting deep WMHs using deep neural networks based on the U-Net.

Methods: 148 non-elderly subjects with migraine were recruited for this study. Our model consists of two networks: the first identifies potential deep WMH candidates, and the second reduces the false positives within the candidates. The first network for initial segmentation includes four down-sampling layers and four up-sampling layers to sort the candidates. The second network for false positive reduction uses a smaller field-of-view and depth than the first network to increase utilization of local information.

Results: Our proposed model segments deep WMHs with a high true positive rate of 0.88, a low false discovery rate of 0.13, and F_1 score of 0.88 tested with ten-fold cross-validation. Our model was automatic and performed better than existing models based on conventional machine learning.

Conclusion: We developed a novel segmentation framework tailored for deep WMHs using U-Net. Our algorithm is open-access to promote future research in quantifying deep WMHs and might contribute to the effective management of WMHs in migraineurs.

1. Introduction

Migraine is one of the primary headache disorders, affecting 20% of the worldwide population. It is associated with nausea and vomiting, and conspicuous sensory, motor, and mood disturbances [1]. Patients with migraine often show an increased presence of white matter hyperintensities (WMHs) [2,3]. WMHs appear in T2-weighted and fluid-attenuated inversion recovery (FLAIR) magnetic resonance (MR) images [4]. WMHs are further classified into periventricular and deep WMHs, and both types show distinct risk factors and clinical implications [5]. Studies have reported that periventricular WMHs are associated with a decline in cognitive function and cerebral blood flow, and deep WMHs are of hypoxic/ischemic origin linked with a higher incidence of migraine [6–9]. These suggest the potential use of WMHs as biomarkers to improve therapy planning in migraine patients. In this study, we focus on analytical tools to quantify deep WMHs related to migraine.

Quantifying deep WMHs involves performing segmentation using MR images. The current gold standard is manual segmentation, which is time-consuming and subject to bias. There are several existing automatic segmentation algorithms for WMHs, including deep WMHs [10–12]. Griffanti et al. designed an automatic WMH quantification algorithm, but its performance was inadequate for small and deep WMHs, which are commonly found in young adults. Hulsey et al. used a simple intensity-based thresholding technique for WMH segmentation, but it was not effective for WMHs with low contrast, leading to high false negatives (FNs). Recently, Park et al. designed a fully automated deep WMH detection pipeline using traditional machine learning approaches and deep learning-inspired features. The algorithm performed well on detecting deep WMHs, but it required prior knowledge for the preprocessing steps to obtain the white matter (WM) mask.

The deep neural network (DNN), a method of deep learning, is a disruptive technology that improves on traditional machine learning approaches. Deep learning approaches can learn

* Corresponding author at: School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon 16419, South Korea.
E-mail address: hyunjinp@skku.edu (H. Park).

https://doi.org/10.1016/j.cmpb.2019.105065
0169-2607/

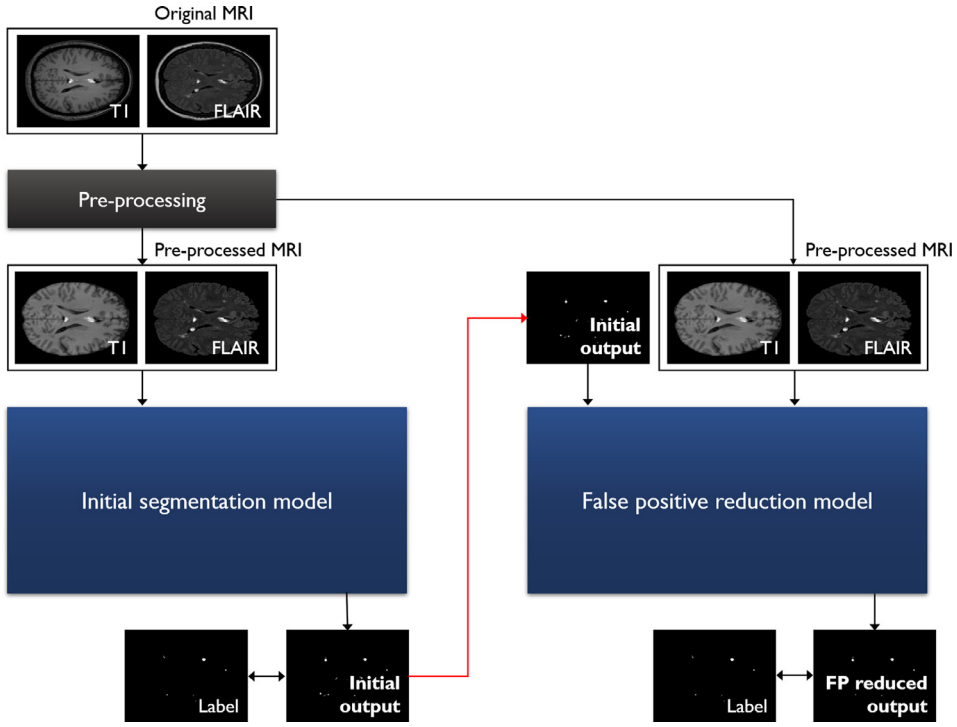

Fig. 1. Flowchart of the proposed model.

complex hierarchical information relevant to segmentation tasks, and they have achieved great success in many segmentation tasks for various medical imaging modalities [13–15]. Deep learning approaches can work in an end-to-end fashion (i.e., from raw input to desired output). Thus, they may not require elaborate preprocessing steps to refine the input data [16,17]. There are no deep learning-based approaches specifically for deep WMHs to the best of our knowledge, but there are deep learning studies dealing with large periventricular WMHs. A convolutional neural network (CNN) model was applied to segment WMHs using the spatial information of WMH locations [18]. This study used prior knowledge of WMH locations, but they did not fully consider neighborhood voxels, as they adopted one voxel per patch. This may cause bias in the segmentation results. Li et al. applied a U-Net based network with the ensemble approach but the performance of segmenting small and deep WMHs was low due to insufficient validation [19].

Motivated by recent successes in deep learning, we propose a new U-Net-based segmentation approach for deep WMHs in this study. We designed two U-Net-based DNN models: one for initial segmentation, and the other for false positive reduction. Our contributions were as follows. Deep WMHs have been segmented using conventional machine learning approaches and they have not performed well because deep WMHs are small and difficult to distinguish from artifacts. We propose a deep learning segmentation approach for deep WMHs to improve the conventional machine learning approaches. Our approach is specific to migraine brain as it was trained using data from the migraine brain.

2. Materials and methods

The overall flow of the proposed approach is given in Fig. 1. There are two U-net-based network models. The first network model identifies potential WMH candidates, and the second network model reduces the false positives (FPs) within the candidates.

2.1. Participants and imaging data

This study was approved by the Institutional Review Board (IRB) of Samsung Medical Center. Written consent was waived by the IRB. The magnetic resonance imaging (MRI) data of patients diagnosed with migraine at the Samsung Medical Center headache clinic between 2015 and 2017 were used in this study. Two headache specialists (MJL and C-SC) made the diagnosis of migraine according to the International Classification of Headache Disorders-3rd edition beta version (ICHD-3 beta) [20]. We included patients with migraine without aura, migraine with typical aura, and chronic migraine. We considered 233 non-elderly patients aged younger than 66 who voluntarily underwent brain MRI during the study period. 67 subjects whose MRI data included heavy motion-related artifacts were excluded as the artifacts made it difficult to define WMHs. 18 subjects without deep WMHs were excluded. Finally, 148 subjects were enrolled in the study.

The T1-weighted and FLAIR MRI scans were acquired using a 3 Tesla MR scanner (Achieva, Philips Medical Systems, Best, Netherlands). The T1-weighted MRI scans were acquired with the following imaging parameters: repetition time (TR) = 9.9 ms; echo time (TE) = 4.6 ms; field of view (FOV) = 240×240 mm^2; acquisition matrix = 480×480 pixels; and slice thickness = 1 mm with 360 slices. The FLAIR data were obtained using the following imaging parameters: TR = 11,000 ms; TE = 125 ms; inversion time = 2800 ms; FOV = 240×240 mm^2; acquisition matrix = 512×512 pixels; and slice thickness = 2 mm with 80 slices.

2.2. Manual annotation of WMHs

The manual annotation of the deep WMHs was performed individually by two specialists (MJL with 9 years of experience in clinical neurology and JC with 11 years of experience in neuroradiology) on 2D slices of the FLAIR image. Note that the manual annotation is the current gold standard for deep WMHs. The intraclass correlation coefficient between the two raters was 0.994 (95%

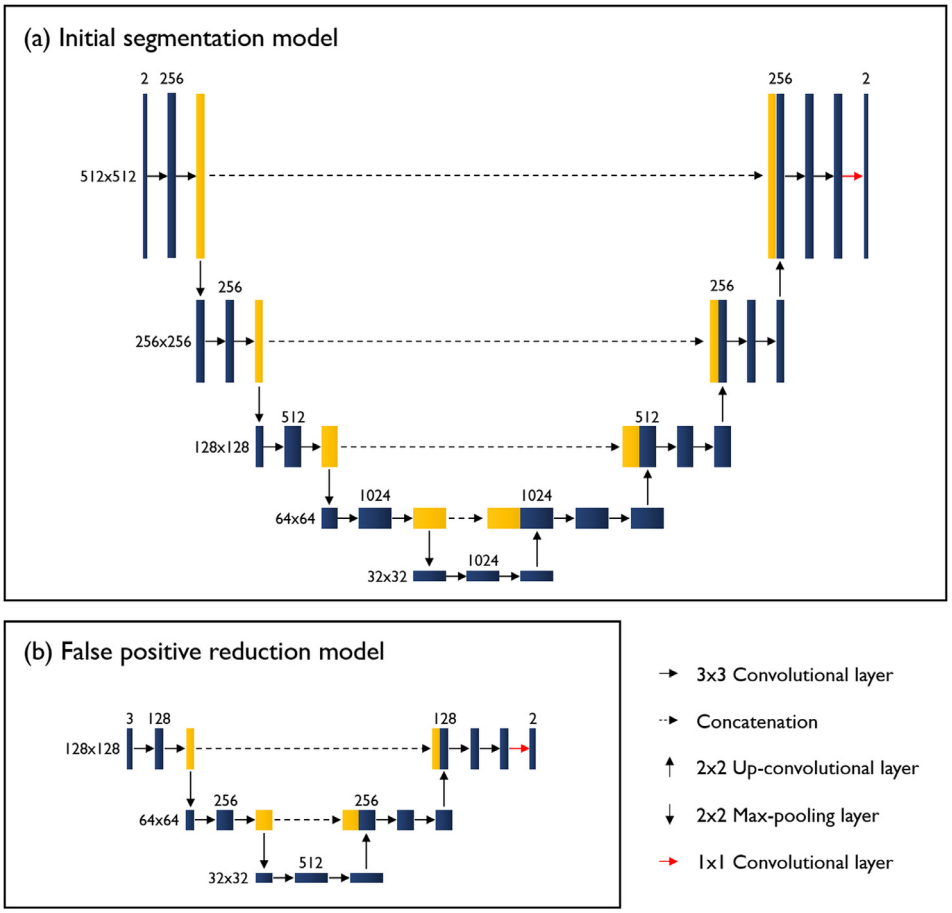

Fig. 2. Detailed structures of the two proposed network models. (a) Initial segmentation model and (b) FP reduction model.

confidence interval between 0.968 and 0.999) for the number of WMHs for each subject. The annotation of the first specialist was used as the ground truth. Detailed methods for defining manual annotations can be found in the literature [12]. In brief, a deep WMH was defined as a round- or oval-shaped FLAIR hyperintensity of variable size in the U-fiber or subcortical WM, which could be discrete or confluent, and showed T1 iso or hypo-intensity [21].

2.3. Preprocessing

The MRI data were preprocessed using the following tools: AFNI and FSL [22,23]. The T1-weighted and FLAIR data were reoriented in the right-posterior-inferior direction. The T1-weighted data were registered onto the FLAIR data using a rigid body transformation. The registration accuracy and further details are given in supplementary material. [24,25] The effect of magnetic field inhomogeneity for both the T1-weighted and FLAIR data was corrected. We also removed the skull from the MRI data. The preprocessing steps were the standard steps commonly performed in MRI imaging studies. Our preprocessing for MRI data was rather minimal compared to the one that requires elaborate WM masks [12].

2.4. Background of the U-net network

The U-net network is based on a CNN that includes compressing and expanding components [26]. The compressing component repeatedly applies convolutional layers followed by activation layers and pooling layers to encode the input data as latent variables. The expanding component restores the original dimension of the input data by up-sampling the latent variables. Each up-sampling layer is concatenated with the corresponding level in the down-sampling layer. Activation layers also follow the up-sampling layers. The final prediction layer is reached after a 1×1 convolutional layer is applied.

2.5. Proposed network for initial segmentation

Our proposed network for initial segmentation is based on the U-net and consists of four down-sampling layers and four up-sampling layers (Fig. 2). The input channel accommodates two images; T1-weighted and FLAIR images. Each down-sampling layer includes a 3×3 convolutional layer, a rectified linear unit (ReLU) activation layer, and a batch normalization layer. Each layer in the compressing stage uses 256, 256, 512, 1024, and 1024 kernels. Each layer in the expanding stage uses 1024, 512, 256, and 256 kernels, and is followed by the ReLU activation layer. After the activation layer, the layer of each expanding stage is concatenated with the one from the matching compressing stage (i.e., yellow box in Fig. 2), which leads to a two-fold increase in the number of channels. The final classification layer is obtained by applying a 1×1 kernel to the last layer of the expanding stage, and then applying the softmax function.

The input to the network is the 2D slice of the 3D MRI data. We did not consider using all the 3D data as input because the deep WMHs are defined on a single slice due to the large gap between slices. We maintained an equal ratio of slices with WMHs to those

Table 1
Detailed network parameters of the two proposed network models.

Architecture	Parameter	Value
Initial segmentation	Batch size	4
	Kernel size (conv.)	3×3
	Activation function (conv.)	ReLU
	Classification layer	ReLU + soft-max
	Cost function	Cross-entropy
	Optimizer	AdadeltaOptimizer
	Learning rate	0.5 (reduced by 0.9 every 25 epochs)
False positive reduction	Batch size	64
	Kernel size (conv.)	3×3
	Activation function (conv.)	ReLU
	Classification layer	ReLU + soft-max
	Cost function	Cross-entropy
	Optimizer	AdadeltaOptimizer
	Learning rate	0.5 (reduced by 0.9 every 25 epochs)

Conv.: convolutional layer.

without WMHs in the training. Multiple slices per patient might be selected during this step. The AdadeltaOptimizer was used as the training optimizer [27]. The initial reading rate was set to 0.5 and was reduced to 0.9 times the original rate every 25 epochs. A total of 250 epochs were used, and cross-entropy was used as the cost function. The detailed network parameters are listed in Table 1.

2.6. Proposed network for false positive reduction

The first network produced the potential WMH candidates, and we further refined the initial segmentation results by reducing the FPs. Our proposed network for FP reduction is also based on the U-net and consists of two down-sampling layers and two up-sampling layers. Each down-sampling layer includes a 3×3 convolutional layer, a ReLU activation layer, and a batch normalization layer. Each layer in the compressing stage uses 128, 256, and 512 kernels. Each layer in the expanding stage uses 256 and 128 kernels, and is followed by the ReLU activation layer. After the activation layer, the layer of each expanding stage concatenates with the one from the matched compressing stage (i.e., yellow box in Fig. 2), which leads to a two-fold increase in the number of channels. The final classification layer is obtained by applying a 1×1 kernel to the last layer of the expanding stage and then applying the soft-max function. The proposed network for FP reduction is shown in Fig. 2.

The input to the second network is the 2D patch of size 128×128 of the 3D MRI data, with an equal ratio of patches with WMH candidates and patches without WMH candidates for the training. The AdadeltaOptimizer was used as the training optimizer. The initial reading rate was set to 0.5 and was reduced by 0.9 times the original rate every 25 epochs. A total of 750 epochs were used, and cross-entropy was used as the cost function. The detailed network parameters are listed in Table 1.

2.7. Cluster refinement of WMH

The output of the FP reduction network was refined by the size of the detected WMH clusters. The effective diameter of a cluster was computed as the cubic root of the volume of the cluster multiplied by $3/\pi$. Clusters whose effective diameter was less than 2.4 mm were removed because small WMH clusters were considered insignificant for migraine studies [9].

2.8. Evaluation

Ten-fold cross-validation was used. All data were randomly divided into ten equally sized groups. Nine groups were used as a training set and the remaining group was used as a test set. The procedure was repeated ten times, leaving a different group out each time. The quality of segmentation was measured by comparing the segmented deep WMH clusters with the manual annotations. The number of true positives (TPs), FPs, and FNs were calculated. If the detected cluster showed spatial overlap with the manual annotation, it was deemed a success considering the small size of deep WMHs. The true positive rate [TPR = TP / (TP + FN)], the false discovery rate [FDR = FP / (TP + FP)], and the F_1 score [F1 = 2 * ((1 - FDR) * TPR) / ((1 - FDR) + TPR)] were obtained. Each leave-out test iteration led to one set of performance measures, and thus we reported the averaged TPR, FDR, and F_1 score over ten leave-out iterations.

2.9. Comparison with other methods

We compared our approach with two DNN models employing different network structures. The two compared DNN models also employed a two-stage approach with initial segmentation and FP reduction networks. The first comparison DNN, referred to as DL-COMP1, adopted a U-net based architecture with two down-sampling and two up-sampling layers for the initial segmentation. The DL-COMP1 network is similar to ours, with lower model capacity. The second comparison DNN, referred to as DL-COMP2, was based on a simple CNN model containing two convolutional layers and no max-pooling layers for the initial segmentation. The DL-COMP2 was adopted to see if the well-known CNNs were adequate for segmenting deep WMHs. The FP reduction network was kept the same as our proposed approach for both comparison DNN models. The detailed network parameters for the two compared DNN models are listed in Table 2.

Three openly accessible non-deep learning software algorithms, DEep White matter hyperintensity Segmentation framework (DEWS), Brain Intensity AbNormality Classification Algorithm (BIANCA), and Lesion-TOADS of MIPAV software, were compared with our approach [12,28,29]. DEWS is a recent algorithm dedicated to segmenting deep WMHs. It detected deep WMHs based on intensity and size information. An exquisite WM mask was constructed, and deep learning-inspired features related to size and texture were used for the FP reduction. BIANCA is a conventional machine learning algorithm for large periventricular WMH segmentation using supervised learning. It uses a k-nearest neighbor algorithm with intensity values and spatial locations as features and produces a probability map of WMHs. The final output was obtained after thresholding and binarizing the probability map. Lesion-TOADS is also used for large periventricular WMH segmentation and it uses a manually annotated atlas combined with a fuzzy classification algorithm. The algorithm reduced the FP based on the structural distances from

Table 2
Detailed network parameters of the two DNN comparison models.

Model	Parameter	Value
DL-COMP1	Base architecture	U-Net
	Batch size	8
	Kernel size (conv.)	3×3
	Activation function (conv.)	ReLU
	Classification layer	ReLU + soft-max
	Cost function	Cross-entropy
	Optimizer	AdadeltaOptimizer
	Learning rate	0.5 (reduced by 0.9 every 25 epochs)
DL-COMP2	Base architecture	CNNs
	Batch size	8
	Kernel size (conv.)	3×3
	Activation function (conv.)	ReLU
	Classification layer	ReLU + soft-max
	Cost function	Cross-entropy
	Optimizer	AdadeltaOptimizer
	Learning rate	0.5 (reduced by 0.9 every 25 epochs)

Table 3
Demographics of the enrolled subjects.

Clinical information	Subjects ($n = 148$)
Mean age (SD)	44.4 (12.40)
Females	82
Migraine (with aura)	13
Migraine (without aura)	88
Migraine (chronic)	30
Headache days (monthly)	10

Table 4
Comparison of results of the proposed network with other approaches.

	TPR	FDR	F_1 score
Proposed model	0.87	0.10	0.88
DL-COMP1	0.47	0.04	0.63
DL-COMP2	0.00	0.00	0.00
DEWS	0.80	0.04	0.87
BIANCA	0.02	0.98	0.02
Lesion-TOADS	0.76	0.98	0.04

the border of gray matter and the ventricles to obtain the final output.

2.10. Computational environments

Our approach and the comparison DNN models were trained on four NVIDIA Titan XP GPUs with 12 GB GDDR5 RAM using cuda 8.0 and cuDNN 6.0 with TensorFlow 1.12.0 library based on the Python 2.7.12 environment [30,31]. Statistical analysis was performed using MATLAB 2017b [32].

3. Results

3.1. Demographics of the subjects

Detailed demographic information of the enrolled subjects can be found in the literature [12]; this is summarized in Table 3.

3.2. Detection results for deep WMH clusters

The mean TPR, FDR, and F_1 score over the ten leave-out iterations from the initial segmentation network were 0.88, 0.13, and 0.87, respectively. The mean TPR, FDR, and F_1 score measured after the FP reduction network were 0.87, 0.10, and 0.88, respectively. Both results were measured after applying the same cluster refinement algorithm. Fig. 3 shows the representative outputs of two networks and corresponding manual annotations.

We also measured how our approach performed with respect to the size of the WMHs. In general, larger WMHs were easy to segment, and smaller WMHs were difficult to segment. We applied different threshold levels from 0 mm to 3.6 mm for effective diameter to the output of the proposed model. We then measured the TPR and FDR of our algorithm to segment WMHs that were larger than the threshold. Fig. 4 shows the plot of the TPR and FDR with respect to the various WMH diameters. The TPR monotonically increased while FDR decreased as the effective diameter increased (Fig. 4).

3.3. Comparison with other approaches

Our approach was compared with five approaches: two DNN models and three approaches based on conventional machine learning. The DL-COMP1 model showed a mean TPR of 0.47, a mean FDR of 0.04, and a F_1 score of 0.63. No WMHs were detected with the DL-COMP2 model. The DEWS had a mean TPR of 0.80, a mean FDR of 0.04, and a F_1 score of 0.87. The BIANCA had a mean TPR of 0.02, a mean FDR of 0.98, and a F_1 score of 0.02. The Lesion-TOADS showed a mean TPR of 0.76, a mean FDR of 0.98, and a F_1 score of 0.04. The results of a representative comparison are shown in Fig. 5 and Table 4.

3.4. Code availability

The code used to implement our approach and limited anonymized imaging data are available from a software sharing platform (https://github.com/jisu-hong/deepwmh).

4. Discussion

Most of the existing segmentation algorithms for WMHs focus on large periventricular WMHs and do not focus specifically on deep WMHs. These algorithms did not perform well when applied to deep WMHs. This is because deep WMHs have distinct characteristics compared with periventricular WMHs. Thus, a dedicated segmentation algorithm is necessary to manage deep WMHs, which are small, discrete, and difficult to distinguish from MR artifacts.

We proposed a two-stage approach for deep WMH segmentation using DNNs. Our approach required fewer preprocessing steps and did not use any prior information. The first network for initial segmentation captured the global features by using slice-wise data.

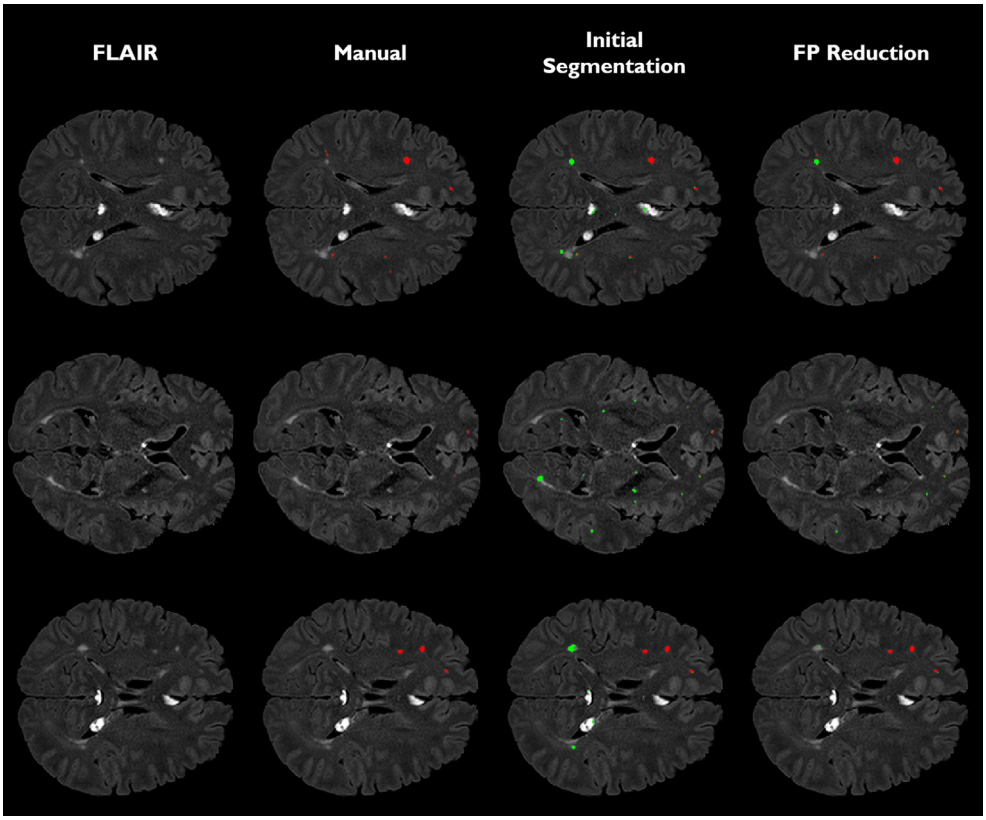

Fig. 3. Comparison between the initial output and the FP reduced output for a representative case. The columns are FLAIR image, manual annotation, initial segmentation results, and the results of FP reduction (from left to right). The red clusters are the TP WMH clusters, and the green clusters are the FP clusters. The rows are different axial slices for the representative cases.

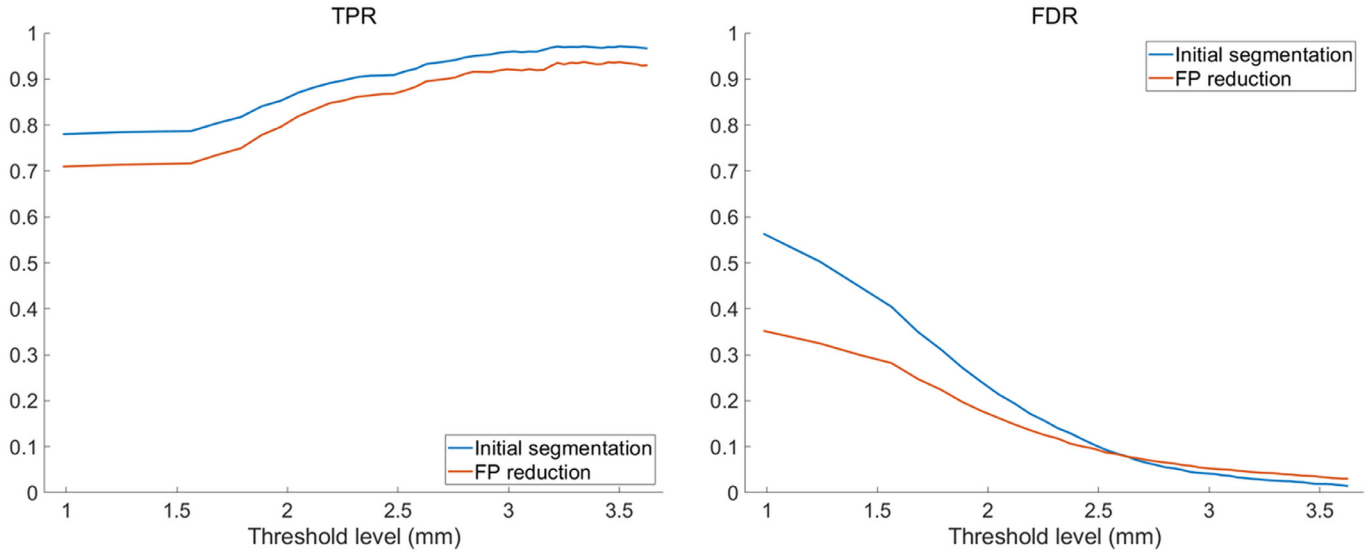

Fig. 4. TPR and FDR of the proposed network models with respect to the different WMH cluster size threshold.

The second model for FP reduction used patch-wise data to focus on local characteristics. To the best of our knowledge, our study is the first one to leverage recent advances in deep learning to perform deep WMH segmentation.

Many studies have successfully adopted the U-net for various segmentation tasks [33–35]. However, the U-net uses multiple down-sampling layers during encoding, which can cause the loss of local detailed information. Our FP reduction model tackled this problem by using a downsized FOV. The layers after the four max-pooling processes in the initial segmentation model and those after the two max-pooling processes in the FP reduction model both have a size of 32×32. Both hidden layers have the same matrix size, but the initial segmentation model has four max-pooling layers of raw data, which results in a compression ratio of 1:256, while those from the FP reduction model achieve a compression ratio of 1:16. This difference allows the FP reduction model to focus on detailed information and effectively reduces errors inherited from the initial segmentation model. We tried to combine the two

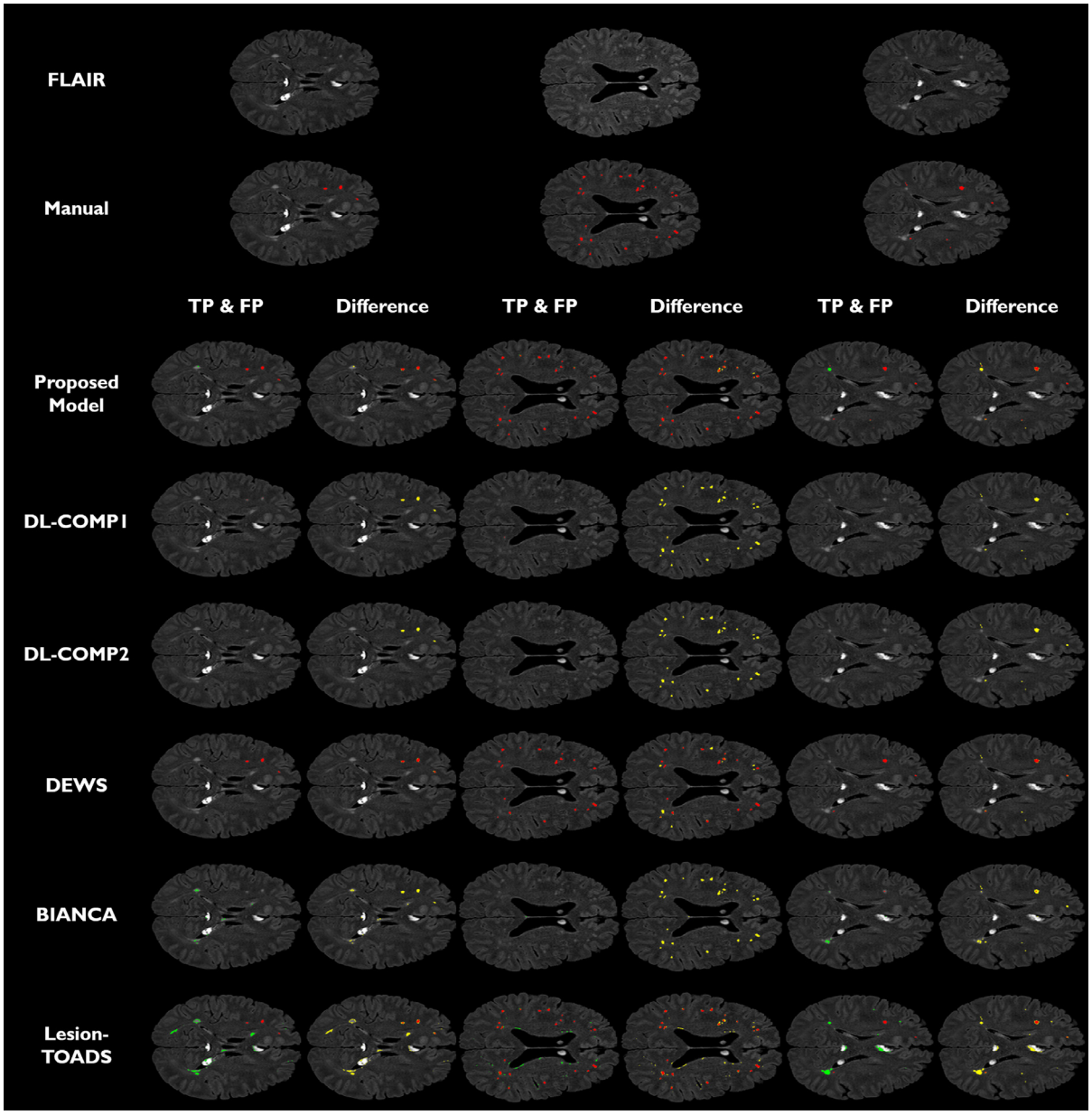

Fig. 5. Segmentation results for a representative case with our proposed model and comparison models. The rows represent FLAIR image, manual annotations, results of our proposed model, and other comparison models (from top to bottom). The columns in the first two rows are different axial slices for the representative case. The columns in rows 3 to 8 have segmentation results along with the difference image between manual annotation and various approaches. The red clusters are the TP WMH clusters, the green clusters are the FP clusters, and the yellow clusters are errors between manual annotation and various approaches.

networks into one large network, but it led to worse performance than having two separate networks. One possible reason could be that the FP reduction network focuses on patches with potential FPs, while the combined network would focus on the whole data and hence would be less sensitive to FPs.

Our approach was compared with existing methods. The TPR reflects the portion of detected WMHs from all WMHs, and the FDR reflects the portion of wrong detections among predicted WMHs. BIANCA had a very low mean TPR of 0.02 due to the low overall detection capability, and most of the detected objects were identified as FPs, which resulted in a mean FDR of 0.98. Lesion-TOADS detected many clusters, but they were mainly periventricular WMHs and artifacts. This caused both the TPR and FDR of the approach to be high (a mean TPR of 0.76, a mean FDR of 0.98, and a F_1 score of 0.04). DEWS had a mean TPR of 0.80, a mean FDR of 0.04, and a F_1 score of 0.87. It had worse TPR and better FDR than our model. The algorithm performed well but required many preprocessing steps driven by prior knowledge to compute the WM masks. Such preprocessing steps require manual tuning of the parameters, and thus the algorithm's generalizability to datasets other than the ones tested could be troublesome. In contrast, our proposed model included minimal preprocessing steps and suffered less in this regard. In the case of DL-COMP1, both the TPR and FDR were lower than those of the proposed model (a mean TPR of 0.47, a mean FDR of 0.04, and a F_1 score of 0.63). These results imply that the overall detection was low, but the detected lesions were actual deep WMHs. It might also imply that reducing the model capacity of our approach could lead to decreased performance. We also compared the results of our proposed network with a similar model that had 6 down-sampling layers in its

initial segmentation network. Using more layers did not improve the overall performance (a mean TPR of 0.59, a mean FPR of 0.02, and a F_1 score of 0.74). This might imply that our choice of the number of down-sampling layers was appropriate and extra complexity could be harmful for segmenting deep WMHs. In the case of DL-COMP2, no clusters were detected. This might imply that simple CNNs are not suitable for segmentation tasks, as many earlier studies have shown [36–38]. The proposed model was able to detect deep WMHs well with high TPR and low FDR, as shown in Fig. 5.

We additionally explored the use of different parameters for the FP reduction model. The proposed FP reduction model used an input patch of 128×128. We explored whether using a smaller input patch of 16×16 could improve segmentation performance. Other parameters of the FP reduction model were fixed, and the initial segmentation model was the same. The segmentation performance had a mean TPR of 0.67 and a mean FDR of 0.07. The FDR decreased, but the TPR decreased as well compared with the proposed model. This indicated that deep WMH clusters were reasonably detected, but the details of the segmented regions were of lesser quality than with the proposed model. This indicated that using a smaller FOV for FP reduction results in a trade-off between TPR and FDR.

We measured how our approach performed with respect to the size of the WMHs. We applied different threshold levels from 0 mm to 3.6 mm for effective diameter to the results. The TPR increased while FDR decreased as the effective diameter increased (Fig. 4). At threshold level above 2.5 mm, the FDR values of the initial segmentation network were not different from those of using two networks. One possible reason could be that the deep WMHs are likely to be small and the initial WMH clusters from the initial segmentation network tend to be small clusters as well (Fig. 3). The FP reduction network removed the FPs from these candidates. For threshold levels equivalent to small clusters, the WMH clusters are more likely to contain FPs compared to large clusters. The FP reduction network could be more effective for smaller clusters as there is more chance of FPs present.

For evaluation, we adopted the cluster-wise overlap rather than the dice score as the performance index. If the size of the detected object is small, the dice score is highly sensitive to small differences. In this study, we used multi-channel input data that included the FLAIR image and registered T1-weighted image onto the FLAIR image. The image registration procedure contained small errors of misalignment, which made the dice score unstable, although the manual inspection of the segmentation seemed acceptable. Thus, we adopted the cluster-wise overlap as the performance index to reduce the instability, which allowed some movement of pixels between the detected and the ground truth WMH clusters.

Our proposed model showed good performance for TPR, but the FDR was not ideal. This is partly because the deep WMHs are difficult to distinguish from the MR artifacts. Some deep WMHs share similar appearance profile compared to artifacts caused by magnetic field susceptibility and blood flow, and thus distinguishing between the two could be difficult [39]. Magnetic field susceptibility causes artifacts near the amygdaloidal nucleus and anterior temporal pole and blood flow near the sinuses and main artery are known to cause hyper-intense clusters that look similar to WMHs. In those cases, expert annotation is performed considering the neighborhood information.

The imaging data we used were obtained from a routine clinical setting, and thus the image quality could have been compromised to allow efficient image acquisition. Our imaging data and the associated MR artifacts reflect what is common in clinical practice, and future improvements of image quality might improve the performance of our proposed model. We used manually annotated labels as the ground truth within the supervised learning framework. The manual annotations may suffer from intra- and inter-observer variability, which negatively affects segmentation performance. Another source of variability is the potential misalignment introduced during the image registration procedure between T1 and FLAIR data. We used the U-Net to segment deep WMHs. Many promising deep learning techniques are being developed and using them might improve the segmentation of deep WMHs. This is left for future work.

5. Conclusions

Discrete and small-sized deep WMHs are difficult to segment using the current automatic approaches. Our proposed method leveraged DNN to segment deep WMHs well and may contribute to the effective management of WMHs in migraineurs.

Declaration of Competing Interest

None.

Acknowledgments

This work was supported by the Institute for Basic Science (grant number IBS-R015-D1), the NRF (National Research Foundation of Korea, grant numbers NRF-2019R1H1A2079721, NRF-2017R1A2B2009086, and NRF-2017R1A2B4007254), the MIST (Ministry of Science and ICT) of Korea under the ITRC (Information Technology Research Center) support program (grant number IITP-2019-2018-0-01798) supervised by the IITP (Institute for Information & communication Technology Promotion), and the IITP grant funded by the Korean government under the AI Graduate School Support Program (grant number 2019-0-00421).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2019.105065.

References

[1] T. Kurth, A.C. Winter, A.H. Eliassen, R. Dushkes, K.J. Mukamal, E.B. Rimm, W.C. Willett, J.E. Manson, K.M. Rexrode, Migraine and risk of cardiovascular disease in women: prospective cohort study, BMJ (2016), doi:10.1136/bmj.i2610.

[2] M.C. Kruit, M.A. Van Buchem, P.A.M. Hofman, J.T.N. Bakkers, G.M. Terwindt, M.D. Ferrari, L.J. Launer, Migraine as a risk factor for subclinical brain lesions, J. Am. Med. Assoc. (2004), doi:10.1001/jama.291.4.427.

[3] S. Erdélyi-Bõtor, M. Aradi, D.O. Kamson, N. Kovács, G. Perlaki, G. Orsi, S.A. Nagy, A. Schwarcz, T. Dõczi, S. Komoly, G. Deli, A. Trauninger, Z. Pfund, Changes of migraine-related white matter hyperintensities after 3 years: a longitudinal MRI study, Headache (2015), doi:10.1111/head.12459.

[4] J. Lin, D. Wang, L. Lan, Y. Fan, Multiple factors involved in the pathogenesis of white matter lesions, Biomed. Res. Int. (2017), doi:10.1155/2017/9372050.

[5] F. Fazekas, J.B. Chawluk, A. Alavi, MR signal abnormalities at 1.5 t in Alzheimer's dementia and normal aging, Am. J. Roentgenol. 149 (2) (1987) 351–356 Aug., doi:10.2214/ajr.149.2.351.

[6] S.W. Seo, J.M. Lee, K. Im, J.S. Park, S.H. Kim, S.T. Kim, H.J. Ahn, J. Chin, H.K. Cheong, M.W. Weiner, D.L. Na, Cortical thinning related to periventricular and deep white matter hyperintensities, Neurobiol. Aging (2012), doi:10.1016/j.neurobiolaging.2010.12.003.

[7] V.H. ten Dam, D.M.J. van den Heuvel, A.J.M. de Craen, E.L.E.M. Bollen, H.M. Murray, R.G.J. Westendorp, G.J. Blauw, M.A. van Buchem, Decline in total cerebral blood flow is linked with increase in periventricular but not deep white matter hyperintensities, Radiology (2007), doi:10.1148/radiol.2431052111.

[8] E.J. Van Dijk, N.D. Prins, H.A. Vrooman, A. Hofman, P.J. Koudstaal, M.M.B. Breteler, Progression of cerebral small vessel disease in relation to risk factors and cognitive consequences: rotterdam scan study, Stroke (2008), doi:10.1161/STROKEAHA.107.513176.

[9] K.W. Kim, J.R. MacFall, M.E. Payne, Classification of white matter lesions on magnetic resonance imaging in elderly persons, Biol. Psychiatry (2008), doi:10.1016/j.biopsych.2008.03.024.

[10] L. Griffanti, M. Jenkinson, S. Suri, E. Zsoldos, A. Mahmood, N. Filippini, C.E. Sexton, A. Topiwala, C. Allan, M. Kivimäki, A. Singh-Manoux, K.P. Ebmeier, C.E. Mackay, G. Zamboni, Classification and characterization of periventricular and deep white matter hyperintensities on MRI: a study in older adults, Neuroimage (2018), doi:10.1016/j.neuroimage.2017.03.024.

[11] K.M. Hulsey, M. Gupta, K.S. King, R.M. Peshock, A.R. Whittemore, R.W. McColl, Automated quantification of white matter disease extent at 3 T: comparison with volumetric readings, J. Magn. Reson. Imaging (2012), doi:10.1002/jmri.23659.

[12] B.Y. Park, M.J. Lee, S. hak Lee, J. Cha, C.S. Chung, S.T. Kim, H. Park, DEWS (DEep white matter hyperintensity segmentation framework): a fully automated pipeline for detecting small deep white matter hyperintensities in migraineurs, NeuroImage Clin. (2018), doi:10.1016/j.nicl.2018.02.033.

[13] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, doi:10.1109/ICCV.2015.178.

[14] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C.I. Sánchez, R. Mann, A. den Heeten, N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, Med. Image Anal. (2017), doi:10.1016/j.media.2016.07.007.

[15] J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A. Guevara Lopez, Representation learning for mammography mass lesion classification with convolutional neural networks, Comput. Methods Programs Biomed. (2016), doi:10.1016/j.cmpb.2015.12.014.

[16] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning Dense Volumetric Segmentation From Sparse Annotation, Medical Image Computing and Computer-Assisted Intervention, 2016, doi:10.1007/978-3-319-46723-8_49.

[17] F. Milletari, N. Navab, S.A. Ahmadi, V-Net: fully convolutional neural networks for volumetric medical image segmentation, in: Proceedings of the Fourth International Conference on 3D Vision, 3DV, 2016, doi:10.1109/3DV.2016.79.

[18] M.F. Rachmadi, M. del, C. Valdés-Hernández, M.L.F. Agan, C. Di Perri, T. Komura, 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 Graph. (2018), doi:10.1016/j.compmedimag.2018.02.002.

[19] H. Li, G. Jiang, J. Zhang, R. Wang, Z. Wang, W.S. Zheng, B. Menze, Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images, Neuroimage (2018), doi:10.1016/j.neuroimage.2018.07.005.

[20] Headache Classification Committee of the International Headache Society (IHS), The international classification of headache disorders, 3rd edition (beta version), Cephalalgia (2013), doi:10.1177/0333102413485658.

[21] J.M. Wardlaw, E.E. Smith, G.J. Biessels, C. Cordonnier, F. Fazekas, R. Frayne, R.I. Lindley, J.T. O'Brien, F. Barkhof, O.R. Benavente, S.E. Black, C. Brayne, M. Breteler, H. Chabriat, C. Decarli, F.-.E. de Leeuw, F. Doubal, M. Duering, N.C. Fox, S. Greenberg, V. Hachinski, I. Kilimann, V. Mok, R. van Oostenbrugge, L. Pantoni, O. Speck, B.C.M. Stephan, S. Teipel, A. Viswanathan, D. Werring, C. Chen, C. Smith, M. van Buchem, B. Norrving, P.B. Gorelick, M. Dichgans, STandards for reporting vascular changes on nEuroimaging (STRIVE v1), neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration, Lancet. Neurol. (2013), doi:10.1016/S1474-4422(13)70124-8.

[22] R.W. Cox, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages, Comput. Biomed. Res. (1996), doi:10.1006/cbmr.1996.0014.

[23] M. Jenkinson, C.F. Beckmann, T.E.J. Behrens, M.W. Woolrich, S.M. Smith, FSL - review, Neuroimage (2012), doi:10.1016/j.neuroimage.2011.09.015.

[24] D. Chyzhyk, R. Dacosta-Aguayo, M. Mataró, M. Graña, An active learning approach for stroke lesion segmentation on multimodal MRI data, Neurocomputing (2015), doi:10.1016/j.neucom.2014.01.077.

[25] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, A. Biller, Deep MRI brain extraction: a 3D convolutional neural network for skull stripping, Neuroimage (2016), doi:10.1016/j.neuroimage.2016.01.024.

[26] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, Medical Image Computing and Computer-Assisted Intervention, 2015, doi:10.1007/978-3-319-24574-4_28.

[27] M.D. Zeiler, ADADELTA: an adaptive learning rate method, *arXiv e-print* 2012, arXiv: 1212.5701.

[28] N. Shiee, P.L. Bazin, A. Ozturk, D.S. Reich, P.A. Calabresi, D.L. Pham, A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions, Neuroimage (2010), doi:10.1016/j.neuroimage.2009.09.005.

[29] L. Griffanti, G. Zamboni, A. Khan, L. Li, G. Bonifacio, V. Sundaresan, U.G. Schulz, W. Kuker, M. Battaglini, P.M. Rothwell, M. Jenkinson, BIANCA (Brain intensity abnormality classification algorithm): a new tool for automated segmentation of white matter hyperintensities, Neuroimage (2016), doi:10.1016/j.neuroimage.2016.07.018.

[30] J. Nickolls, I. Buck, M. Garland, K. Skadron, Scalable parallel programming with CUDA, Queue (2008), doi:10.1145/1365490.1365500.

[31] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, G. Brain, I. Osdi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, TensorFlow : a system for large-scale machine learning, Software available from tensorflow.org., 2015.

[32] T. MathWorks, MATLAB (R2017b), MathWorks Inc., 2017, doi:10.1007/s10766-008-0082-5.

[33] V. Iglovikov, A. Shvets, TernausNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation, *arXiv e-print* 2018, arXiv: 1801.05746.

[34] X. Han, MR-based synthetic CT generation using a deep convolutional neural network method, Med. Phys. (2017), doi:10.1002/mp.12155.

[35] B. Norman, V. Pedoia, S. Majumdar, Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee mr imaging data to determine relaxometry and morphometry, Radiology (2018), doi:10.1148/radiol.2018172322.

[36] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. (2017), doi:10.1109/TPAMI.2016.2572683.

[37] X. Li, H. Chen, X. Qi, Q. Dou, C.W. Fu, P.A. Heng, H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from ct volumes, IEEE Trans. Med. Imaging (2018), doi:10.1109/TMI.2018.2845918.

[38] K. Kamnitsas, C. Ledig, V.F.J. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. (2017), doi:10.1016/j.media.2016.10.004.

[39] J.M. Wardlaw, M.C. Valdés Hernández, S. Muñoz-Maniega, What are white matter hyperintensities made of? Relevance to vascular cognitive impairment, J. Am. Heart Assoc. (2015), doi:10.1161/JAHA.114.001140.