



OPEN ACCESS

EDITED BY

Jussi Tohka,
University of Eastern Finland, Finland

REVIEWED BY

Lu Zhao,
University of Southern California,
United States
Juan Miguel Valverde,
University of Eastern Finland, Finland

*CORRESPONDENCE

Fei Shang
bit552sf@bit.edu.cn
Shuai Liu
liushuai@bit.edu.cn

†These authors have contributed
equally to this work and share first
authorship

‡These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Applied Neuroimaging,
a section of the journal
Frontiers in Neurology

RECEIVED 17 August 2022

ACCEPTED 27 October 2022

PUBLISHED 17 November 2022

CITATION

Gong T, Han H, Tan Z, Ning Z, Qiao H,
Yu M, Zhao X, Tang X, Liu G, Shang F
and Liu S (2022) Segmentation and
differentiation of periventricular and
deep white matter hyperintensities in
2D T2-FLAIR MRI based on a cascade
U-net. *Front. Neurol.* 13:1021477.
doi: 10.3389/fneur.2022.1021477

COPYRIGHT

© 2022 Gong, Han, Tan, Ning, Qiao,
Yu, Zhao, Tang, Liu, Shang and Liu. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Segmentation and differentiation of periventricular and deep white matter hyperintensities in 2D T2-FLAIR MRI based on a cascade U-net

Tan Gong^{1†}, Hualu Han^{2†}, Zheng Tan¹, Zihan Ning²,
Huiyu Qiao², Miaoxin Yu³, Xihai Zhao², Xiaoying Tang¹,
Gaifen Liu^{3,4}, Fei Shang^{1*†} and Shuai Liu^{1*†}

¹Department of Biomedical Engineering, Beijing Institute of Technology School of Life Science, Beijing, China, ²Department of Biomedical Engineering, Center for Biomedical Imaging Research, Tsinghua University School of Medicine, Beijing, China, ³Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, ⁴China National Clinical Research Center for Neurological Diseases, Beijing, China

Background: White matter hyperintensities (WMHs) are a subtype of cerebral small vessel disease and can be divided into periventricular WMHs (pvWMHs) and deep WMHs (dWMHs). pvWMHs and dWMHs were proved to be determined by different etiologies. This study aimed to develop a 2D Cascade U-net (Cascade U) for the segmentation and differentiation of pvWMHs and dWMHs on 2D T2-FLAIR images.

Methods: A total of 253 subjects were recruited in the present study. All subjects underwent 2D T2-FLAIR scan on a 3.0 Tesla MR scanner. Both contours of pvWMHs and dWMHs were manually delineated by the observers and considered as the gold standard. Fazekas scale was used to evaluate the burdens of pvWMHs and dWMHs, respectively. Cascade U consisted of a segmentation U-net and a differentiation U-net and was trained with a combined loss function. The performance of Cascade U was compared with two other U-net models (Pipeline U and Separate U). Dice similarity coefficient (DSC), Matthews correlation coefficient (MCC), precision, and recall were used to evaluate the performances of all models. The linear correlations between WMHs volume (WMHV) measured by all models and the gold standard were also conducted.

Results: Compared with other models, Cascade U exhibited a better performance on WMHs segmentation and pvWMHs identification. Cascade U achieved DSC values of 0.605 ± 0.135 , 0.517 ± 0.263 , and 0.510 ± 0.241 and MCC values of 0.617 ± 0.122 , 0.526 ± 0.263 , and 0.522 ± 0.243 on the segmentation of total WMHs, pvWMHs, and dWMHs, respectively. Cascade U exhibited strong correlations with the gold standard on measuring WMHV ($R^2 = 0.954$, $p < 0.001$), pvWMHV ($R^2 = 0.933$, $p < 0.001$), and dWMHV ($R^2 = 0.918$, $p < 0.001$). A significant correlation was found on lesion volume between Cascade U and gold standard ($r > 0.510$, $p < 0.001$).

Conclusion: Cascade U showed competitive results in segmentation and differentiation of pvWMHs and dWMHs on 2D T2-FLAIR images, indicating potential feasibility in precisely evaluating the burdens of WMHs.

KEYWORDS

periventricular white matter hyperintensities, deep white matter hyperintensities, image segmentation, cascade U-net, 2D T2-FLAIR

Introduction

Cerebral small vessel disease (CSVD) is characterized as a group of pathological processes with various etiologies affecting small arteries, arterioles, venules, and capillaries of the brain (1). White matter hyperintensities (WMHs) are commonly observed MRI-based biomarkers for CSVD. Some studies have validated that WMHs are closely related to an increased risk of stroke (2, 3), and higher WMHs load is observed in patients with depression (4), Alzheimer's disease (AD) (5), and migraine (6). WMHs, in view of the location, can be further divided into periventricular WMHs (pvWMHs) which extend from the ventricular wall and deep WMHs (dWMHs) in deep white matter area. pvWMHs and dWMHs show distinct risk factors and clinical implications. pvWMHs are associated with a decline in cerebral blood flow and cognitive function, and dWMHs are of hypoxic/ischemic origin and linked with a higher incidence of migraine and mood disorder (7, 8). pvWMHs-dWMHs dichotomization plays an important role in potential therapeutic intervention. The changes on fluidity and water content of interstitial tissue fluid in patient with WMHs may be reversible if intervention is conducted at early stage (9).

Magnetic resonance imaging T2-weighted fluid-attenuated inversion recovery (T2-FLAIR) has been widely used to evaluate WMHs. On T2-FLAIR images, WMHs appear as hyperintense objects scattered throughout the white matter and cerebrospinal fluid is nullified for enhanced discrimination of ischemic pathology (10). Traditionally, qualitative and quantitative evaluations of WMHs relied on the radiologists' subjective scale or manual delineations, which were time-consuming and laborious (11). Convolutional neural network (CNN) has been validated as an efficient tool for the automatic segmentation of WMHs (12–18). A fine-tuned fully convolutional network (FCN) which combined linearly fine to coarse feature maps of a pretrained Visual Geometry Group was designed for the automatic segmentation of WMHs (12). Wang et al. developed an FCN structure that consisted of three U-shaped networks to segment WMHs using different shapes of patch (13). Wu et al. segmented WMHs on skull-stripped images using a skip connection U-net which could capture more features and speed the

optimization convergence (14). Li et al. proposed an ensemble architecture that generated the segmentation of WMHs from multiple U-net models and the average of which was taken as the prediction result (15). In addition, two-dimensional (2D) U-net had the potential to distinguish WMHs from acute ischemic lesions (16). Recently, precise segmentations of pvWMHs and dWMHs attracted investigators' attention. An ensemble network architecture (TrUE-Net) of three parallel U-net using coronal, sagittal, and horizontal planes as independent input was proposed to segment and differentiate pvWMHs and dWMHs (17). Mojiri et al. presented a U-shaped three-dimensional (3D) Bayesian network to segment WMHs and a secondary 3D U-net to differentiate pvWMHs and dWMHs (18). However, these networks were designed for 3D thin-section scan. 2D T2-FLAIR was recommended in CSVD studies and acquired with high efficiency in community study (19, 20). Previous studies validated that 3D networks were more susceptible to limited axial slices compared with 2D networks (21).

TABLE 1 Clinical characteristics of the study population.

	All subjects (n = 253)	Train set (n = 176)	Test set (n = 77)	P
Age (year)	57.5 ± 12.7	56.5 ± 12.4	59.8 ± 13.2	0.067
Sex (male)	111 (43.9%)	80 (45.5%)	31 (40.3%)	0.444
BMI (kg/m ²)	24.3 ± 3.1	24.4 ± 3.0	24.1 ± 3.4	0.415
Hypertension	80 (31.5%)	59 (33.5%)	21 (27.3%)	0.325
Hyperlipidemia	98 (38.7%)	66 (37.5%)	32 (41.6%)	0.542
Diabetes	30 (11.9%)	22 (12.5%)	8 (10.4%)	0.886
Smoking ^a	31 (17.6%)	23 (18.5%)	8 (15.4%)	0.615
Volume (ml)				
WMHV	1.53 (0.58–3.60)	1.42 (0.57–3.37)	1.90 (0.71–4.00)	0.328
pvWMHV	0.82 (0.31–1.87)	0.82 (0.32–1.75)	0.86 (0.29–2.19)	0.946
dWMHV	0.44 (0.11–1.75)	0.41 (0.10–1.57)	0.73 (0.20–2.05)	0.094

Values are provided as mean ± standard deviation, median (interquartile range), or numbers (%) for each variable. BMI, body mass index; WMHV, white matter hyperintensity; pvWMHV, periventricular white matter hyperintensity volume; dWMHs, deep white matter hyperintensity.

^a Subjects with unknown smoking status (52 in train set and 25 in test set) were excluded.

Previous studies exhibited potential of cascaded CNN in WMHs segmentation on 3D T2-FLAIR images. However, the segmentation of small WMHs lesions on 2D FLAIR images was still a challenge, and further research was needed for accurate segmentation and differentiation of pvWMHs and dWMHs (13, 16). In this study, a cascade CNN (Cascade U) with a combined loss function was developed to segment and differentiate small pvWMHs and dWMHs on 2D T2-FLAIR images.

Materials

Dataset

In the present study, all subjects were recruited from a community study of Cardio- and cerebrovascular Accident Monitoring, Epidemiology, and caRe quALity system (CAMERA) (22). A total of 253 subjects (30–80 years old, 111 male patients) were included from January 2017 to August 2020. The breakdown of subjects in our study is presented in Table 1. Each subject underwent 2D T2-FLAIR scan (TR/TE = 7000 ms/140 ms, flip angle = 90°, FOV = 230×230×133 mm³, voxel size = 0.9×0.9×5.5 mm³, matrix size = 256×256) on a 3.0 Tesla MR scanner (Achieva TX, Philips Healthcare, Best, The Netherlands) with a 32-channel phase array head coil. Repeat scan was conducted if the image quality was poor by visual assessment with the following criteria: (I) contrast-based: unclear or invisible contrast in gray matter and white matter; and (II) artifact-based: severe head motion, signal drop, or geometric distortion.

Gold standard

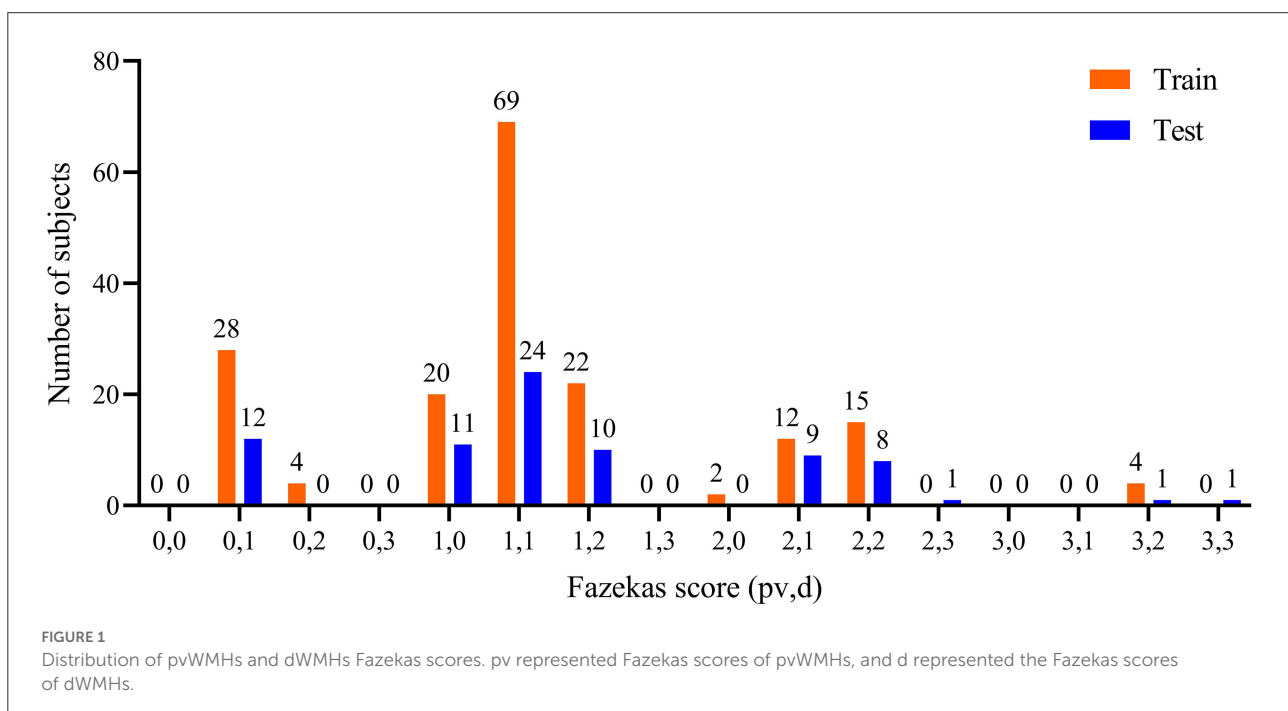
All T2-FLAIR images were reviewed by two radiologists with > 3 years' experience in neuroradiology with consensus. The contour delineated by the observer was considered as the gold standard. Each slice was interpolated from 256×256 to 512×512 for lesion delineation. In addition, Fazekas scales (ranging from 0 to 3 for pvWMHs and dWMHs, respectively) were used for describing the degree of WMHs (23). The distribution of Fazekas scales is shown in Figure 1. Ninety % of the subjects (223/253) were with a total Fazekas score lower than 3 (range 0 to 6).

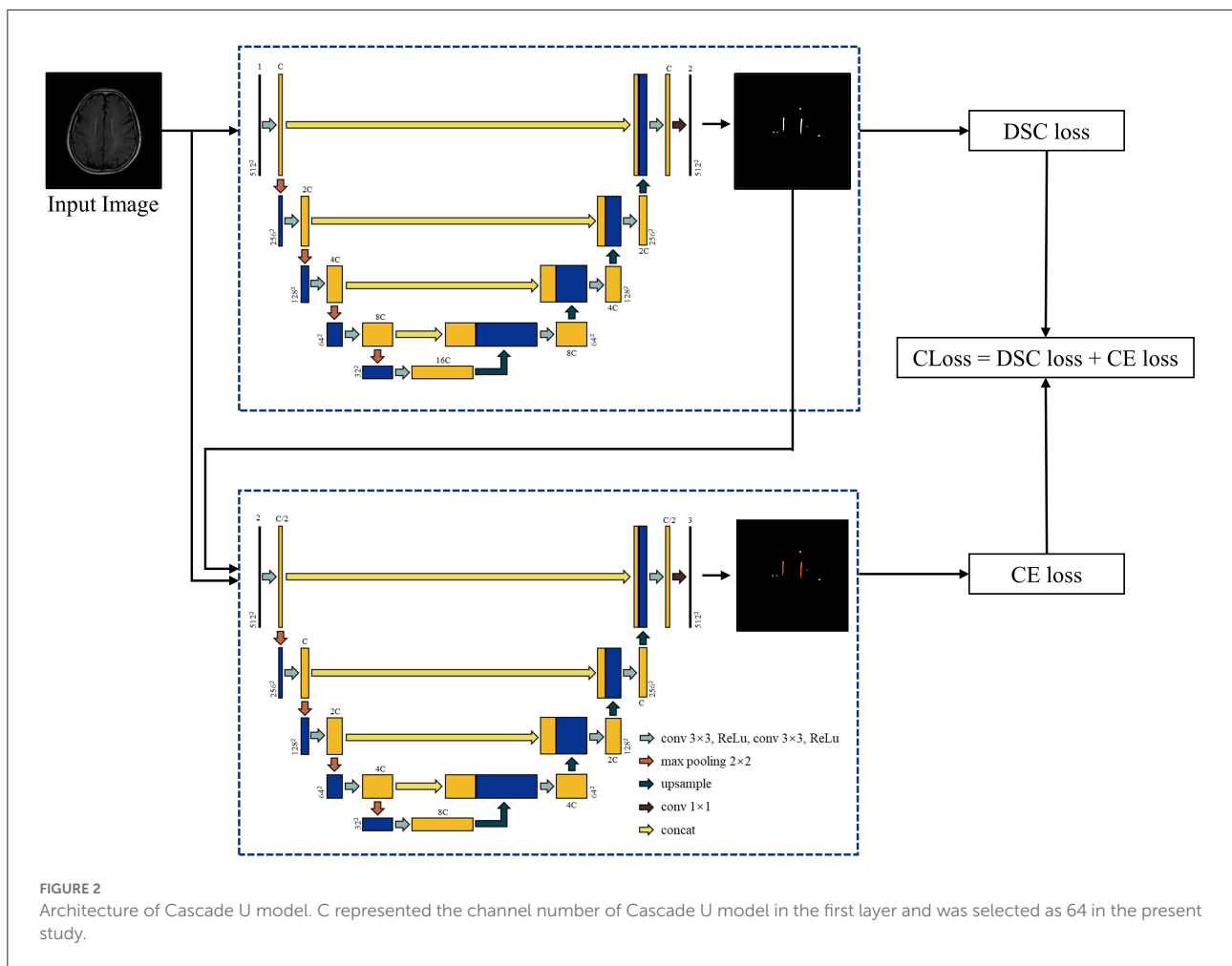
Methods

Network architecture

In the present study, 2D CNNs were applied due to a voxel size of 5.5 mm along the vertical axis and small lesions. The cascade U-net architecture (Cascade U) was constructed based on a U-net architecture proposed by Ronneberger et al. in 2015 (24). Two other models (Pipeline U-net and Separate U-net) were trained as comparison.

a) Cascade U-net: Cascade U was integrated by segmentation stage and differentiation stage, and trained with a combined loss function (CLoss). The architecture is exhibited in Figure 2. In the first stage, the input was each T2-FLAIR slice, and the result was the WMHs segmentation result. In the second stage, pvWMHs and dWMHs were differentiated using the original slice and possibility map of





WMHs segmentation from the first stage. The CLoss was defined by a Dice similarity (DSC) loss from the segmentation stage and cross-entropy (CE) loss from the differentiation stage. The loss functions were defined as follows:

$$\begin{aligned}
 \text{CLoss} &= \text{DSC Loss} + \text{CE Loss} \\
 \text{DSC Loss} &= 1 - \frac{2 \sum_{n=1}^N p_n g_n + \varepsilon}{\sum_{n=1}^N (p_n + g_n) + \varepsilon} \\
 \text{CE Loss} &= -\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{\text{Class}} y_{ni} \log \hat{y}_{ni}
 \end{aligned}$$

where N represented the number of pixels of the input data in both CE loss and DSC loss. In DSC loss, p_n represented the softmax output probability of WMHs at the n th pixel, $g_n = 0$ or 1 represented the ground truth of the n th pixel. The ε term was 10^{-4} to prevent the denominator being zero. In CE loss, Class represented the number of classes, \hat{y}_{ni} represented the softmax output probability of class i at the n th pixel, and y_{ni} represented the given one-hot encoded label for class i of the n th pixel.

b) Pipeline U-net: Pipeline U-net (Pipeline U) consisted of two U-net models. The first model was trained to segment WMHs with the addition of DSC loss and CE loss, while the second model was trained to differentiate pvWMHs and dWMHs with CE loss. Two models were trained independently.

c) Separate U-net: Separate U-net (Separate U) was two U-net models. Two models were trained to segment pvWMHs and dWMHs with the addition of DSC loss and CE loss, respectively.

The segmentation stages in Cascade U and Pipeline U were same, while the differentiation stages in three networks were same. All three models are publicly available at <https://github.com/GGTTGTGT-2020/Cascade-U-Net-white-matter-hyperintensities->.

Model training

A total of 176 subjects were randomly selected as training set (18 subjects for validation), and 77 subjects were testing set.

Considering CNN's ability to segment brain MR images without preprocessing (25), we only conducted a Z-score normalization.

Data augmentation is an effective method to improve the robustness of models and the precision of prediction results. Five transformations were applied on training set for each slice in every epoch: 1) rotating: $[-30^\circ, 30^\circ]$, 2) shifting: $[-26 \text{ pixels}, 26 \text{ pixels}]$, 3) scaling: $[0.9, 1.1]$, 4) horizontal flipping, and 5) changing the intensity of each pixel according to the following formula:

$$\left(\frac{x - \min(x)}{\max(x) - \min(x)}\right)^\gamma \times (\max(x) - \min(x)) + \min(x)$$

where γ was randomly selected from $[0.5, 1]$. The segmentation stage in Cascade U was pretrained for 200

epochs with WMHs gold standard on training set to accelerate convergence speed.

Each model was convergent after being trained for 500 epochs, and the model with the minimal loss on validation set was taken as the prediction model. Batch size was set to 6. Adam optimizer was used with an initial learning rate of 2×10^{-3} (26). The models were trained on an RTX 3090 graphics card with 24G of memory, and all models were constructed based on PyTorch 1.11.0 (27).

Evaluation metrics

DSC, Matthews correlation coefficient (MCC), recall, precision, and the correlation coefficient of lesion volume were used to evaluate the performance of the models.

The DSC is defined as follows:

$$DSC = 2 \times \frac{|G \cap P|}{|G| + |P|}$$

where P represented a binary mask, and G represented the gold standard.

The MCC is an index to measure the quality of a binary classification system when the size of samples in the two classes varies substantially (28). The MCC is defined as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$$

where TP , TN , FP , and FN represented true positive, true negative, false positive, and false negative, respectively. The range of MCC is $[-1, 1]$. A value of 1 means that the prediction is completely consistent with actual result, 0 means not as good as random prediction result, and -1 means completely inconsistent with actual result.

The recall for lesions is defined as follows:

TABLE 2 Performance of three models on total WMHs, pvWMHs, and dWMHs segmentations.

	Total WMHs	pvWMHs	dWMHs
Cascade U			
DSC	0.605 ± 0.135	0.517 ± 0.263	0.510 ± 0.241
MCC	0.617 ± 0.122	0.526 ± 0.263	0.522 ± 0.243
Precision	0.641 ± 0.184	0.551 ± 0.282	0.588 ± 0.293
Recall	0.621 ± 0.131	0.679 ± 0.223	0.626 ± 0.214
Pipeline U			
DSC	0.593 ± 0.135	0.389 ± 0.254	0.405 ± 0.218
MCC	0.607 ± 0.120	0.403 ± 0.254	0.421 ± 0.215
Precision	0.593 ± 0.181	0.476 ± 0.284	0.398 ± 0.242
Recall	0.655 ± 0.137	0.531 ± 0.298	0.629 ± 0.222
Separate U			
DSC	0.540 ± 0.171	0.451 ± 0.247	0.513 ± 0.240
MCC	0.563 ± 0.147	0.469 ± 0.245	0.523 ± 0.239
Precision	0.501 ± 0.219	0.401 ± 0.247	0.515 ± 0.261
Recall	0.687 ± 0.132	0.736 ± 0.192	0.698 ± 0.193

DSC, dice similarity coefficient; MCC, Matthews correlation coefficient.

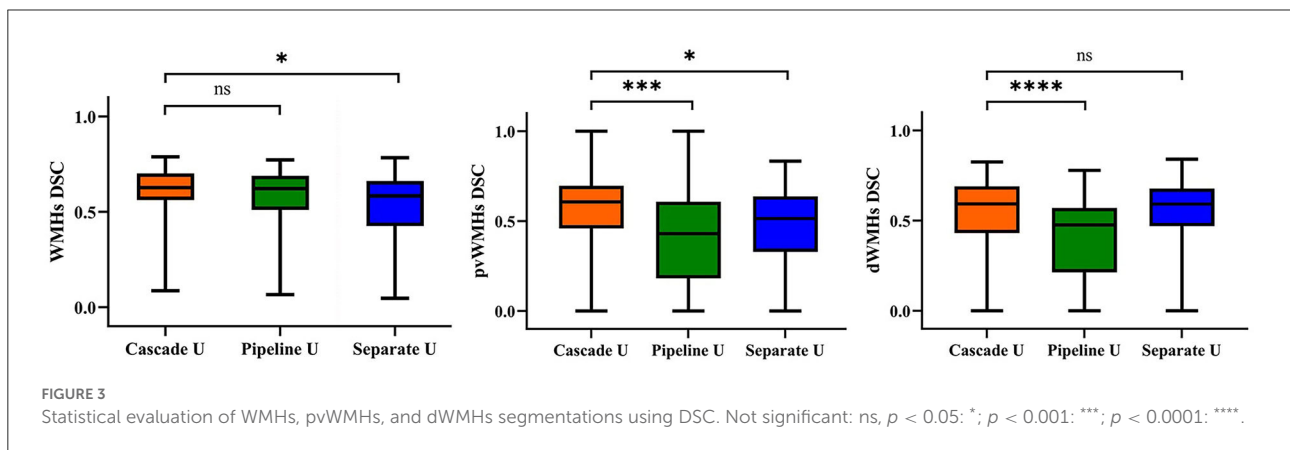


TABLE 3 Performance of three models on subjects' pvWMHs with different Fazekas scores.

Fazekas scores	pvWMHs			
	0 (n = 12)	1 (n = 45)	2 (n = 18)	3 (n = 2)
Cascade U				
DSC	0.083 ± 0.276	0.562 ± 0.174	0.672 ± 0.099	0.729 ± 0.067
MCC	0.083 ± 0.276	0.575 ± 0.169	0.677 ± 0.091	0.740 ± 0.058
Precision	0.083 ± 0.276	0.624 ± 0.194	0.644 ± 0.116	0.865 ± 0.017
Recall	1.000 ± 0.000	0.578 ± 0.207	0.720 ± 0.094	0.640 ± 0.111
Pipeline U				
DSC	0.083 ± 0.276	0.397 ± 0.212	0.556 ± 0.127	0.516 ± 0.185
MCC	0.083 ± 0.276	0.415 ± 0.210	0.564 ± 0.125	0.578 ± 0.153
Precision	0.083 ± 0.276	0.539 ± 0.239	0.550 ± 0.151	0.773 ± 0.004
Recall	1.000 ± 0.000	0.383 ± 0.242	0.594 ± 0.151	0.466 ± 0.230
Separate U				
DSC	0.000 ± 0.000	0.487 ± 0.168	0.635 ± 0.006	0.677 ± 0.094
MCC	0.000 ± 0.000	0.515 ± 0.154	0.643 ± 0.005	0.696 ± 0.075
Precision	0.000 ± 0.000	0.420 ± 0.188	0.573 ± 0.006	0.849 ± 0.045
Recall	1.000 ± 0.000	0.674 ± 0.184	0.730 ± 0.006	0.587 ± 0.154

DSC, dice similarity coefficient; MCC, Matthews correlation coefficient.

TABLE 4 Performance of three models on subjects' dWMHs with different Fazekas scores.

Fazekas scores	dWMHs			
	0 (n = 12)	1 (n = 44)	2 (n = 19)	3 (n = 2)
Cascade U				
DSC	0.000 ± 0.000	0.567 ± 0.142	0.661 ± 0.070	0.583 ± 0.113
MCC	0.000 ± 0.000	0.583 ± 0.136	0.674 ± 0.063	0.584 ± 0.112
Precision	0.000 ± 0.000	0.646 ± 0.194	0.792 ± 0.077	0.584 ± 0.151
Recall	1.000 ± 0.000	0.555 ± 0.178	0.584 ± 0.122	0.587 ± 0.073
Pipeline U				
DSC	0.000 ± 0.000	0.440 ± 0.158	0.563 ± 0.098	0.343 ± 0.140
MCC	0.000 ± 0.000	0.464 ± 0.142	0.572 ± 0.093	0.352 ± 0.144
Precision	0.000 ± 0.000	0.392 ± 0.177	0.635 ± 0.107	0.452 ± 0.186
Recall	1.000 ± 0.000	0.597 ± 0.175	0.527 ± 0.140	0.276 ± 0.112
Separate U				
DSC	0.000 ± 0.000	0.570 ± 0.138	0.663 ± 0.069	0.629 ± 0.058
MCC	0.000 ± 0.000	0.585 ± 0.126	0.669 ± 0.068	0.628 ± 0.057
Precision	0.000 ± 0.000	0.550 ± 0.171	0.718 ± 0.089	0.646 ± 0.044
Recall	1.000 ± 0.000	0.656 ± 0.176	0.633 ± 0.121	0.613 ± 0.070

DSC, dice similarity coefficient; MCC, Matthews correlation coefficient.

$$Recall = \frac{TP}{TP+FN}$$

The precision for lesions is defined as follows:

$$Precision = \frac{TP}{TP+FP}$$

Linear correlation analysis between WMH volumes (WMHVs) measured by three models and the gold standard was also employed to evaluate the performance of all models. The non-parametric Mann-Whitney U-test was conducted to assess the performance between Cascade U and other models. To test the association between clinical characteristics and the segmentation performance of Cascade U, the Mann-Whitney U-test and Spearman's correlation analysis were conducted

TABLE 5 Association between demographic, clinical factors, and the segmentation results of Cascade U on testing set.

BV	Total WMHs		pvWMHs		dWMHs	
	DSC	P	DSC	P	DSC	P
Hypertension (yes = 1)	0.601 ± 0.134	0.575	0.557 ± 0.232	0.211	0.473 ± 0.255	0.010
	0.617 ± 0.146		0.410 ± 0.103		0.608 ± 0.179	
Hyperlipidemia (yes = 1)	0.600 ± 0.130	0.397	0.488 ± 0.263	0.125	0.539 ± 0.226	0.222
	0.612 ± 0.146		0.559 ± 0.267		0.468 ± 0.263	
Diabetes (yes = 1)	0.600 ± 0.141	0.473	0.501 ± 0.275	0.160	0.510 ± 0.247	0.536
	0.648 ± 0.082		0.655 ± 0.078		0.505 ± 0.220	
Smoking ^a (yes = 1)	0.596 ± 0.151	0.389	0.535 ± 0.236	0.576	0.469 ± 0.266	0.909
	0.564 ± 0.131		0.429 ± 0.153		0.490 ± 0.229	
Sex(male)	0.610 ± 0.149	0.355	0.512 ± 0.257	0.399	0.512 ± 0.243	0.763
	0.598 ± 0.116		0.481 ± 0.277		0.507 ± 0.247	
CV						
BMI	r = -0.032	0.783	r = -0.075	0.525	r = 0.032	0.787
Age	r = -0.109	0.346	r = -0.079	0.492	r = 0.150	0.194
Lesion volume	r = 0.689	< 0.001	r = 0.557	< 0.001	r = 0.510	< 0.001

Values are provided as mean ± standard deviation. BV, binary variables; CV, continuous variables; DSC, dice similarity coefficient; BMI, body mass index.

^a Subjects with unknown smoking status (25 in testing set) were excluded.

on binary and continuous variables, respectively. A $p < 0.05$ was considered statistically significant, and all statistical analyses were conducted on SPSS v25.0 (International Business Machines, Inc., New York, USA). Continuous variables were presented as mean ± standard deviation (SD).

Results

The mean values of DSC, MCC, precision, and recall for Cascade U, Pipeline U, and Separate U are summarized in Table 2. Cascade U had the highest DSC and MCC on WMHs segmentation and pvWMHs among the three models. On dWMHs segmentation, DSC and MCC of Cascade U were similar with those of Separate U. Compared with other two models, Cascade U exhibited the highest precision. However, Separate U produced the highest recall. Figure 3 shows the comparison between Cascade U and other models on DSC of WMHs, pvWMHs, and dWMHs segmentation.

The performance of all models on different levels of Fazekas scores is summarized in Tables 3 and 4. On pvWMHs segmentation (Table 3), Cascade U achieved the highest DSC, MCC, and precision on subjects in all levels, but lower recall in subjects with Fazekas scores of 1 and 2 than Separate U. On dWMHs segmentation (Table 4), Separate U showed a similar performance with Cascade U on the cases with a Fazekas score < 3, but the highest DSC, MCC, precision, and recall on the cases with a Fazekas score of 3. The association between demographic, clinical factors, and the segmentation results of Cascade U is summarized in Table 5. A significant difference was found on dWMHs segmentation results between the hypertensive and

non-hypertensive groups. In addition, there was a significant correlation between segmentation results and lesion volume.

The correlations between WMHV measured by radiologists and by three models are summarized in Figure 4. Cascade U model had the highest R^2 in WMHV ($R^2 = 0.954$, $p < 0.001$), pvWMHV ($R^2 = 0.933$, $p < 0.001$), and dWMHV ($R^2 = 0.918$, $p < 0.001$) among the three models.

Discussion

In the present study, a 2D Cascade U model using a combined loss function was proposed to segment and identify pvWMHs and dWMHs simultaneously on T2-FLAIR images. Compared with Pipeline U and Separate U, Cascade U exhibited better performance on WMHs segmentation and pvWMHs identification. In addition, the lesion volume measured by Cascade U had the strongest correlation with the gold standard.

The cascade model exhibited advantages on many tasks, including segmentation of the pancreas (29, 30), brain tumor (31), and bladder cancer (32). Liu et al. developed two sub-models for the segmentation of WMHs and differentiation between focal cerebral ischemia and lacunar infarction (33). The results showed the superiority of the cascade model in the segmentation and differentiation of small brain lesions. In the present study, Cascade U outperformed Pipeline U and Separate U on the segmentation and identification of pvWMHs and dWMHs. Combined loss function made the learning process guided by the segmentation loss function (DSC) and classification loss function (CE) at the same time and improved model's performance.

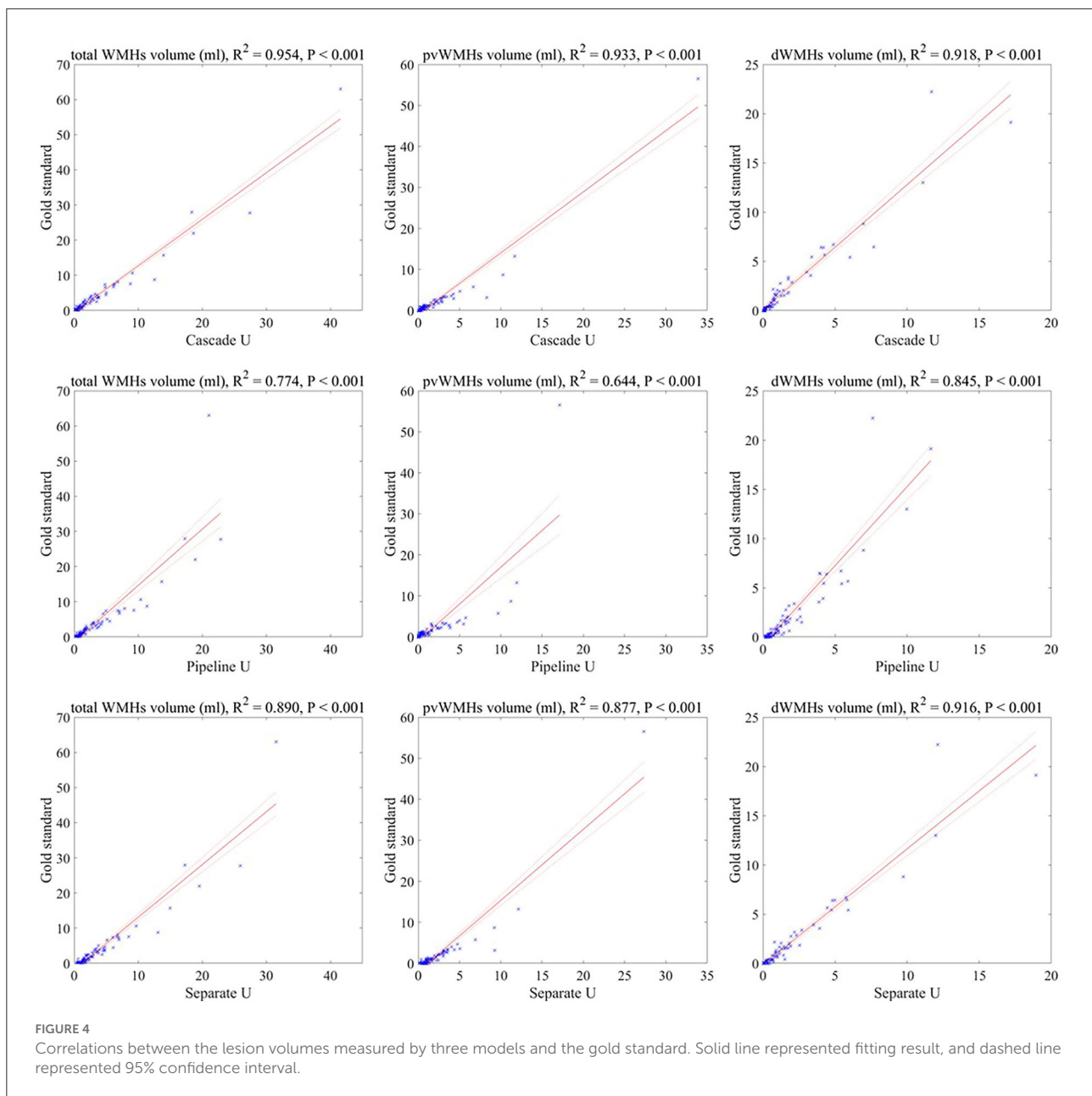


Figure 5 shows the comparisons between the gold standard and three models on the segmentation and differentiation of pvWMHs and dWMHs from a subject with a Fazekas score of 3 (pvWMHs: 2 and dWMHs: 1). Pipeline U underestimated and confused pvWMHs and dWMHs in some cases. In Pipeline U, the identification of pvWMHs and dWMHs relied on the prediction results of segmentation model, leading to the propagation of error. Cascade U overcame this problem *via* training segmentation network together with a differentiation network using a combined loss function. Compared with Cascade U, Separate U exhibited inferior performance on pvWMHs segmentation and similar

performance on dWMHs segmentation. However, overlaps of pvWMHs and dWMHs were found in some cases (blue regions in Figure 6), exhibiting a region identified as pvWMHs and dWMHs simultaneously by two U-net models in Separate U. The example implied that independent training had some limitation in segmentation and differentiation of pvWMHs and dWMHs.

In this study, all subjects were recruited from community-based population aiming to investigate the risk factors of cerebrovascular disease risk. Early characterization of WMHs before cognitive symptoms occur may prevent further deterioration to dementia (34, 35). In the present

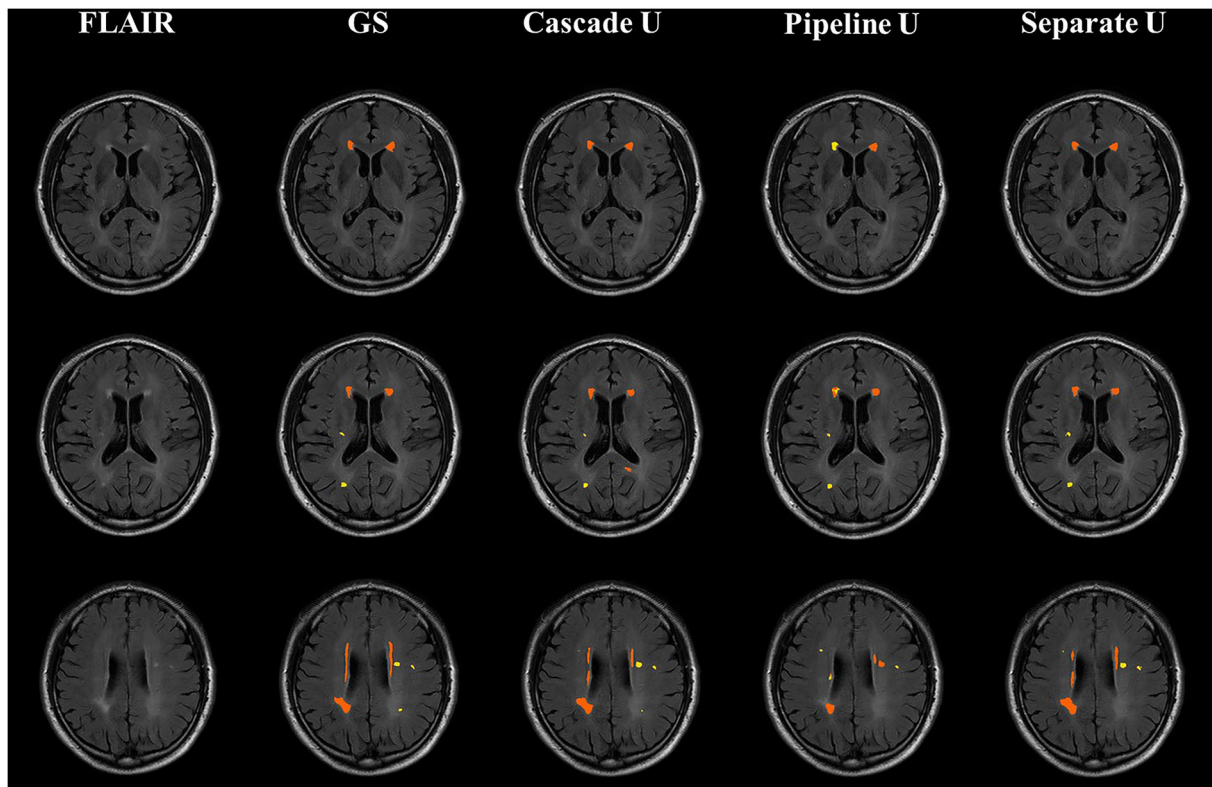


FIGURE 5
Illustration of segmentation results from a subject with Fazekas score of 3. From left to right were T2-FLAIR images, the gold standard, Cascade U prediction results, Pipeline U prediction results, and Separate U prediction results. Orange represented pvWMHs, and yellow represented dWMHs. GS, gold standard.

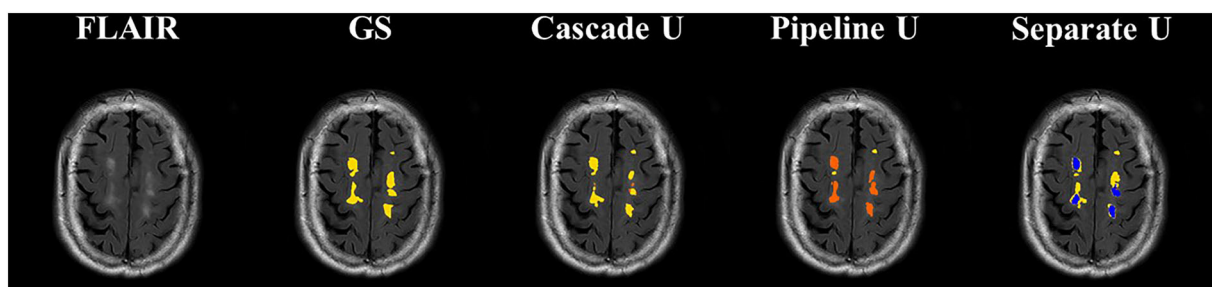
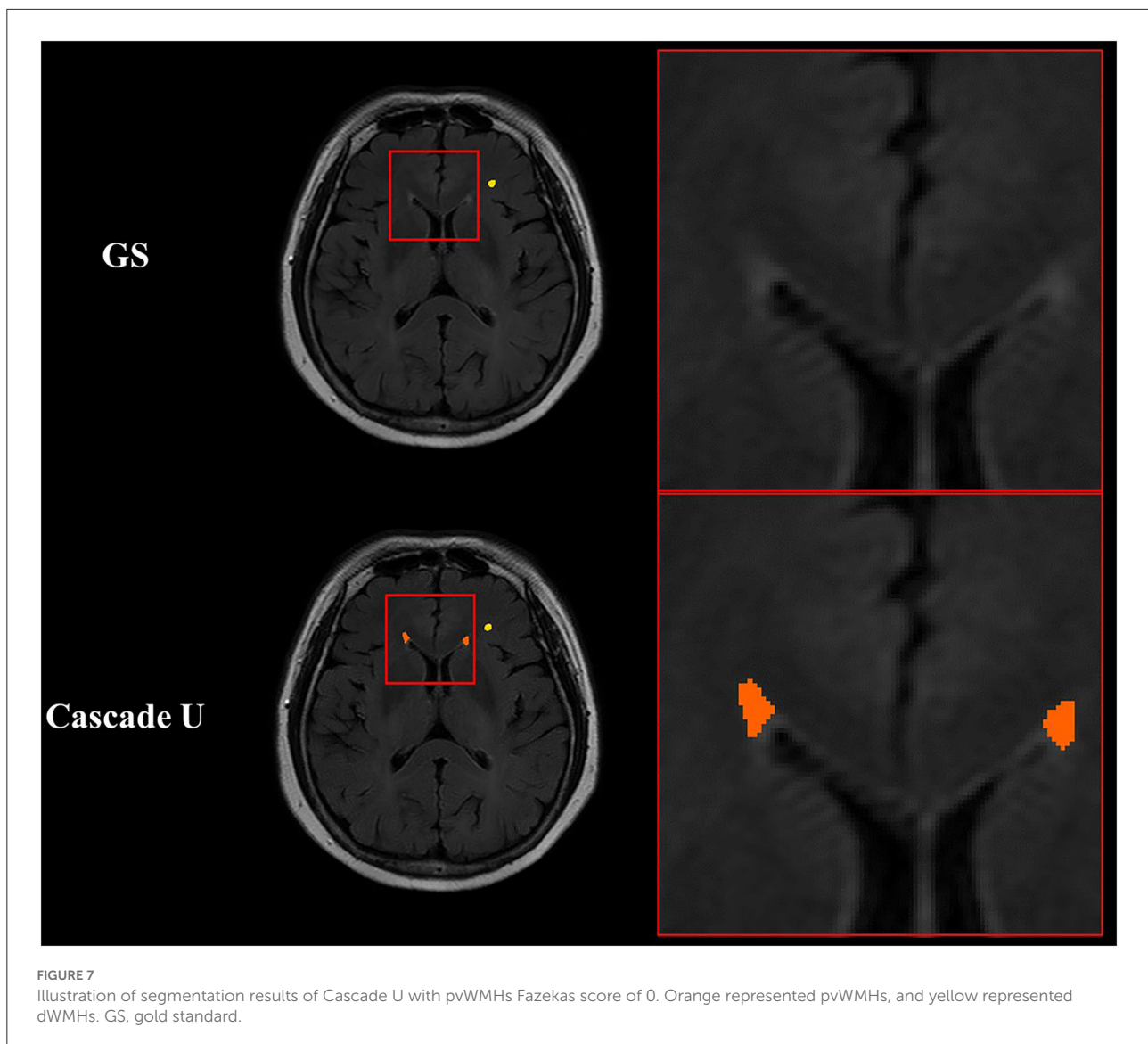


FIGURE 6
Illustration of segmentation results of dWMHs. Orange represented pvWMHs, yellow represented dWMHs, and blue represented the overlaps of the model prediction. GS, gold standard.

study, 90% (223/253) of subjects recruited were at an early stage, with a total Fazekas score lower than 3 (range 0 to 6). The median of pvWMHV and dWMHV on each subject was 0.82 ml and 0.44 ml, respectively. Previous studies had validated that accurate segmentation of small WMHs was more challenging compared with large WMHs (13, 16, 33). Although exhibited potential in WMHs segmentation, CNN-based network was more likely to cause false negative prediction

on the images with smaller lesions and thus led to a lower DSC (36). As exhibited in Table 3, the performance of each model on the segmentation and identification of pvWMHs and dWMHs increased significantly with the burden of WMHs. Our result also exhibited that the segmentation performance was correlated with lesion volume. In the present study, the subjects with hypertension exhibited a severer dWMHs burden compared with those without hypertension



(1.71 ml vs. 0.60 ml), favoring the better segmentation performance on the hypertensive group than that on the non-hypertensive group.

In addition, manual annotation was considered the gold standard, but it may not always reflect the real situation (17). It was more difficult to accurately delineate the contour of small lesions for radiologists due to low contrast between lesion and adjacent tissue or noise inherent in imaging protocol. Since DSC was a metric sensitivity to the size of object, the correlations between WMHs volume measured by three models and the gold standard were also compared in the present study. The strongest correlation between volume measured by Cascade U and the gold standard also demonstrated the advantage of Cascade U on the segmentation and differentiation of pvWMHs and dWMHs. Figure 7 exhibits some lesions detected by Cascade

U on a subject with a pvWMHs score of 0, but missed in manual delineation by radiologists.

There were several limitations in the present study: 1) T1-weighted MRI can provide additional information to the segmentation of WMHs (37). Some subjects did not undergo T1-weighted MR scan, and only 2D T2-FLAIR images were used in the present study. In future, we will attempt to use multi-modality images, such as T1-weighted and diffusion-weighted imaging, as input of the model. 2) Cascade U was designed on a 2D U-net due to a layer thickness of 5.5 mm with 2D T2-FLAIR scan. 3D U-net can provide competitive results *via* leveraging spatial and anatomical information in volumetric organs (31). CNN with a 3D architecture would be investigated on the segmentation and differentiation of pvWMHs and dWMHs on 3D T2-FLAIR data. 3) Finally, our model was trained and tested

on a single-center dataset. The models should be tested on multi-center datasets or on the crowds with different population characteristics in future.

Conclusion

In this study, a cascade 2D U-net (Cascade U) was proposed for the segmentation and differentiation of pvWMHs and dWMHs. Cascade U was composed by a segmentation stage and a differentiation stage and trained with a combined loss function. Cascade U achieved better segmentation and differentiation of pvWMHs and dWMHs on 2D T2-FLAIR images and showed potential feasibility in precisely evaluating the burden of WMHs.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Tsinghua University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

FS and SL designed the study. HH and HQ contributed to the data collection. XZ and GL contributed to the data management. HH, ZN, and MY interpreted the data. TG and ZT developed the network. TG, HH, and ZT performed the statistical analysis. TG, FS, and SL drafted the manuscript. SL, FS, HH, XT, and XZ

revised the manuscript. All authors have read the manuscript and approved the submitted version.

Funding

This study was funded by the grants of the National Natural Science Foundation of China (Nos. 82102136 and 81771825), the Beijing Municipal Commission of Health and Family Planning (No. 2016-1-2041), the Beijing Municipal Science and Technology Commission (Nos. D17110003017002 and D17110003017003), and the Ministry of Science and Technology of China (Nos. 2017YFC1307904, 2017YFC1307702, and 2016YFC0901001).

Acknowledgments

We thank all participants of the CAMERA (Cardio- and cerebrovascular Accident Monitoring, Epidemiology, and caRe quAlity system) study for their contribution.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol.* (2010) 9:689–701. doi: 10.1016/S1474-4422(10)70104-6
- Debette S, Markus H S. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *Br Med J.* (2010) 341:c3666. doi: 10.1136/bmj.c3666
- Wong TY, Klein R, Sharrett AR, Couper DJ, Klein BE, Liao DP et al. Cerebral white matter lesions, retinopathy, and incident clinical stroke. *JAMA.* (2002) 288:67–74. doi: 10.1001/jama.288.1.67
- Devantier TA, Nørgaard BL, Poulsen MK, Garde E, Øvrehus KA, Marwan M et al. White matter lesions, carotid and coronary atherosclerosis in late-onset depression and healthy controls. *Psychosomatics.* (2016) 57:369–77. doi: 10.1016/j.psych.2016.02.005
- Zhuang L, Sachdev PS, Trollor JN, Kochan NA, Reppermund S, Brodaty H, et al. Microstructural white matter changes in cognitively normal individuals at risk of amnesic MCI. *Neurology.* (2012) 79:748–54. doi: 10.1212/WNL.0b013e3182661f4d
- Hamedani AG, Rose KM, Peterlin BL, Mosley TH, Coker LH, Jack CR et al. Migraine and white matter hyperintensities: the ARIC MRI study. *Neurology.* (2013) 81:1308–13. doi: 10.1212/WNL.0b013e3182a8235b
- Kim K W, MacFall J R, Payne M E. Classification of white matter lesions on magnetic resonance imaging in elderly persons. *Biol Psychiat.* (2008) 64:273–80. doi: 10.1016/j.biopsych.2008.03.024
- Kilgore CB, Strain JF, Benzinger TL, Morris JC, Ances BM. Increases in periventricular white matter Hyperintensities associate with Alzheimer Disease biomarkers: Neuroimaging/multi-modal comparisons. *Alzheimers Dement.* (2020) 16:e041286. doi: 10.1002/alz.041286

9. Wardlaw J M, Valdés Hernández M C, Muñoz-Maniega S. What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. *J Am Heart Assoc.* (2015) 4:e001140. doi: 10.1161/JAHA.114.001140
10. Malloy P, Correia S, Stebbins G, Laidlaw DH. Neuroimaging of white matter in aging and dementia. *Clin Neuropsychol.* (2007) 21:73–109. doi: 10.1080/13854040500263583
11. Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage.* (2002) 17:479–89. doi: 10.1006/nimg.2002.1040
12. Xu Y, Géraud T, Puybureau É, Bloch I, Chazalon J. White matter hyperintensities segmentation in a few seconds using fully convolutional network and transfer learning. In: *International MICCAI Brain lesion Workshop*. Quebec City: Springer (2017). p. 501–14.
13. Wang Z, Smith C D, Liu J. Ensemble of multi-sized FCNs to improve white matter hyperintensities segmentation from MRI. *IEEE Access.* (2019) 7:155194–202. doi: 10.1109/ACCESS.2019.2948476
15. Li H, Jiang G, Zhang J, Wang R, Wang Z, Zheng WS et al. Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. *Neuroimage.* (2018) 183:650–65. doi: 10.1016/j.neuroimage.2018.07.005
16. Liu S, Wu X, He S, Song X, Shang F, Zhao X. Identification of white matter lesions in patients with acute ischemic lesions using U-net. *Front Neurol.* (2020) 11:1008. doi: 10.3389/fneur.2020.01008
17. Sundaresan V, Zamboni G, Rothwell PM, Jenkinson M, Griffanti L. Triplanar ensemble U-Net model for white matter hyperintensities segmentation on MR images. *Med Image Anal.* (2021) 73:102184. doi: 10.1016/j.media.2021.102184
18. Mojiri Forooshani P, Biparva M, Ntiri EE, Ramirez J, Boone L, Holmes MF, et al. Deep Bayesian networks for uncertainty estimation and adversarial resistance of white matter hyperintensity segmentation. *Hum Brain Mapp.* (2022) 43:2089–108. doi: 10.1002/hbm.25784
19. Zhu W, Huang H, Zhou Y, Shi F, Shen H, Chen R, et al. Automatic segmentation of white matter hyperintensities in routine clinical brain MRI by 2D VB-Net: A large-scale study. *Front Aging Neurosci.* (2022) 14:915009. doi: 10.3389/fnagi.2022.915009
20. Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* (2013) 12:822–38. doi: 10.1016/S1474-4422(13)70124-8
21. Abulnaga, S M, and Rubin J. Ischemic stroke lesion segmentation in CT perfusion scans using pyramid pooling and focal loss. In: *International MICCAI Brainlesion Workshop*. Granada: Springer (2018). p. 352–63. doi: 10.1007/978-3-030-11723-8_36
22. Han H, Zhang R, Liu G, Qiao H, Chen Z, Liu Y, et al. Reduction of cerebral blood flow in community-based adults with subclinical cerebrovascular atherosclerosis: A 30 T magnetic resonance imaging study. *Neuroimage.* (2019) 188:302–8. doi: 10.1016/j.neuroimage.2018.12.021
23. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 15 T in Alzheimer's dementia and normal aging. *Am J Roentgenol.* (1987) 149:351–6. doi: 10.2214/ajr.149.2.351
24. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer International Publishing. (2015). doi: 10.1007/978-3-662-54345-0_3
25. Valverde JM, Shatillo A, De Feo R, Grohn O, Sierra A, Tohka J, RatLesNetv2: a fully convolutional network for rodent brain lesion segmentation. *Front Neurosci-Switz.* (2020) 14:610239. doi: 10.3389/fnins.2020.610239
26. Kingma D P, Ba J. Adam: Adam: a method for stochastic optimization. In: *International Conference on Learning Representations*. San Diego: Ithaca. (2015). doi: 10.48550/arXiv.1412.6980
27. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G et al. Pytorch: an imperative style, high-performance deep learning library. *Advan Neural Inform Process Sys.* (2019) p. 8024–35. doi: 10.48550/arXiv.1912.01703
28. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics.* (2020) 21:1–13. doi: 10.1186/s12864-019-6413-7
29. Zhao N, Tong N, Ruan D, Sheng K Fully automated pancreas segmentation with two-stage 3D convolutional neural networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen: Springer (2019). p. 201–9. doi: 10.1007/978-3-030-32245-8_23
30. Xue J, He K, Nie D, Adeli E, Shi Z, Lee SW, et al. Cascaded multitask 3-D fully convolutional networks for pancreas segmentation. *IEEE T Cybernetics.* (2019) 51:2153–65. doi: 10.1109/TCYB.2019.2955178
31. Jiang Z, Ding C, Liu M, Tao D. Two-stage cascaded u-net: 1st place solution to BraTS challenge 2019 segmentation task. In: *International MICCAI brain lesion workshop*. Shenzhen: Springer (2019). p. 231–41. doi: 10.1007/978-3-030-46640-4_22
32. Yu J, Cai L, Chen C, Fu X, Wang L, Yuan B et al. Cascade path augmentation Unet for bladder cancer segmentation in MRI. *Med Phys.* (2022) 49:4622–31. doi: 10.1002/mp.15646
33. Liu Y, Li X, Li T, Li B, Wang Z, Gan J et al. A deep semantic segmentation correction network for multi-model tiny lesion areas detection. *Bmc Med Inform Decis.* (2021) 21:1–9. doi: 10.1186/s12911-021-01430-z
34. Chen Y, Wang X, Guan L, Wang Y. Role of white matter hyperintensities and related risk factors in vascular cognitive impairment: a review. *Biomol Ther.* (2021) 11:1102. doi: 10.3390/biom11081102
35. Prins N D, Scheltens P. White matter hyperintensities, cognitive impairment and dementia: an update. *Nat Rev Neurol.* (2015) 11:157–65. doi: 10.1038/nrneuro.2015.10
36. Kuijff HJ, Biesbroek JM, De Bresser J, Heinen R, Andermatt S, Bento M et al. Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge. *IEEE T Med Imaging.* (2019) 38:2556–68. doi: 10.1109/TMI.2019.2905770
37. Dadar M, Maranzano J, Ducharme S, Carmichael OT, Decarli C, Collins DL. Validation of T1w based segmentations of white matter hyperintensity volumes in large scale datasets of aging. *Hum Brain Mapp.* (2018) 39:1093–107. doi: 10.1002/hbm.23894