IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 April 2023, accepted 26 May 2023, date of publication 29 May 2023, date of current version 7 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3281201

RESEARCH ARTICLE

A Nested Attention Guided UNet++ Architecture for White Matter Hyperintensity Segmentation

HAO ZHANG[1], CHENYANG ZHU[2], XUEGAN LIAN[1], AND FEI HUA[1]
[1]The Third Affiliated Hospital of Soochow University, Changzhou 213000, China
[2]School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213000, China

Corresponding authors: Xuegan Lian (13961188116@163.com) and Fei Hua (huafei1970@suda.edu.cn)

This work was supported in part by the Science and Technology Bureau of Changzhou under Grant CE20215044 and Grant CE20215053; and in part by the Key Research and Development Program (Applied Basic Research) of Changzhou, China, under Grant CJ20210123.

ABSTRACT White Matter Hyperintensity (WMH) is a common finding in Magnetic Resonance Imaging (MRI) of patients with cerebral infarction and is associated with poor prognosis. Accurate and rapid segmentation of WMH lesions is critical for clinicians to assess the risk of rebleeding and the long-term prognosis of thrombolytic patients. However, segmentation can be challenging due to the erratic signals of WMH in MRI, leading to imprecise results. Deep learning-based approaches have been proposed, but the dice similarity coefficient remains low. Atlas images are navigation maps that integrate various medical information expressions. In this study, we propose a nested attention-guided UNet++ framework that employs attention mechanisms to capture local and global features of WMH lesions using atlas images for segmentation. The framework consists of two modules, the atlas attention module, and the nested attention-guided nested U-Net module. The atlas attention module generates the atlas attention map, which is used as the input for the nested attention-guided nested U-Net module that generates the segmentation map of the FLAIR image. Experimental results demonstrate that the proposed NAUNet++ framework converges faster than conventional UNet and UNet++ approaches. Moreover, the nested architecture enhances recall and f1 scores of the segmentation results compared to the attention-guided approach.

INDEX TERMS MRI segmentation, UNet, attention mechanism.

I. INTRODUCTION

White matter hyperplasia (WMH) is a pathological condition characterized by abnormal white matter volume increase in the brain. In Magnetic Resonance Imaging (MRI), WMH appears as punctate, patchy, or confluent high signal areas in bilateral periventricular or subcortical white matter on T2-weighted imaging or T2 fluid-attenuated inversion recovery, with isomorphic or slightly low signal on T1-weighted imaging. WMH is associated with cognitive deterioration, dementia, an increased risk of stroke, and is considered a sign of cerebrovascular disease. Furthermore, it is a predictor of poor outcomes in patients with Alzheimer's disease [1]. Hence, rapid and accurate segmentation of WMH lesions is critical to assess the severity of WMH accurately and aid clinicians in evaluating the risk of rebleeding and long-term prognosis of thrombolytic patients. However, human interpretation of WMH is subject to biases and varying interpretations, leading to subjectivity and human error in the results [2]. Therefore, automated segmentation methods based on deep learning algorithms are needed to provide accurate and objective results.

WMH segmentation serves to identify and isolate areas of elevated white matter volume in MRI images. However, the uneven distribution and varying sizes of WMH often pose challenges in accurately demarcating small lesion areas. Furthermore, WMH detection in MRI is susceptible to signal fluctuations and noise, which can lead to imprecise segmentation outcomes. Thus, the process of WMH detection remains intricate. Evaluating the automatic segmentation results is also complicated due to the diverse data sets and evaluation criteria involved. The Dice Similarity Coefficient

The associate editor coordinating the review of this manuscript and approving it for publication was Yongjie Li.

(DSC) is a commonly used metric to measure the quality of segmentation, with a DSC coefficient of over 0.7 deemed satisfactory [3]. Currently, manual and deep learning-based methods are available for WMH segmentation. Manual segmentation entails a human observer delineating the areas of increased white matter volume on MRI scans, which is both time-consuming and susceptible to inter-observer variability, particularly since brain MRI imaging often comprises multiple sections.

Deep learning approaches have recently become prevalent in automated WMH segmentation. Deep learning has yielded remarkable results in numerous fields, including image classification, object detection, and image segmentation. This method involves feature extraction from input images, utilizing the back-propagation algorithm to train networks, and ultimately producing segmentation outcomes [4]. Consequently, deep learning has emerged as a prominent technique for image segmentation. Within medical image segmentation, deep learning has been widely employed in brain tumor segmentation [5], brain lesion segmentation [6], brain tissue segmentation [7], and WMH segmentation [8], [9], [10], [11], [12]. However, the segmentation accuracy of WMH is still insufficient. The performance metrics of segmentation, such as DSC, AUC, recall, and f1 scores, remain limited. Medical imaging, particularly in MRI, requires accurate segmentation of anatomical structures. However, inter-image variability across individuals can make this task challenging. Atlas-based segmentation is a prevalent approach that employs a pre-existing image or set of images as a reference or template for segmentation. This technique involves developing a model for the image population that learns parameters from the training dataset. Atlas-based segmentation reduces the amount of manual annotation and labeling required, while improving accuracy and consistency of segmentation results, especially for complex or highly variable anatomies. Recently, the incorporation of atlas-based segmentation into Convolutional Neural Network (CNN) architectures has shown promise in improving the segmentation performance of white matter hyperintensities (WMH) in MRI scans. Xu and Niethammer proposed to use the atlas image as prior knowledge to jointly train the CNN for WMH segmentation [13]. Wickramasinghe et al. used rough atlases under a CNN architecture to improve the segmentation results for structures with complex shapes and deformations [14]. However, traditional CNNs may not be as effective in preserving spatial information, which is crucial for accurate segmentation in MRI. In contrast, the UNet architecture allows for the extraction of features at different scales and resolutions, with skip connections to maintain spatial information, making it a highly effective approach for segmentation tasks [15].

This work employs the atlas image, which provides a reference image with a standardized coordinate system, to aid in the interpretation and analysis of WMH segmentation. We mainly make two contributions in this work.

- We designed an attention-guided module to incorporate the information of the atlas image for WMH segmentation. This attention-guided module incorporates two attention blocks, one dealing with the upsampling path and the other with the atlas path.
- We designed a Nested Attention guided U-Net++ (NAUNet++) framework for WMH segmentation, which comprises two key modules: the atlas attention module and the attention-guided nested U-Net module. The former generates the atlas attention map, which serves as input for the latter. Both modules utilize the nested UNet architecture. During the training phase, the NAUNet++ framework was utilized to process the FLAIR image and the atlas image. The atlas image was directed fed to the atlas attention module, while the attention-guided nested U-Net module processed the FLAIR image.

The experimental results demonstrate that the proposed NAUNet++ converges more quickly than typical UNet and UNet++ approaches. Furthermore, compared to the attention-guided approach, the nested architecture enhances the recall and f1 scores of the segmentation results.

The following sections are structured as follows. Section II outlines related work on white matter hyperintensity segmentation. Section III provides an overview of the network architecture used in this paper, including U-Net, Nested U-Net, and Attention U-Net. In Section IV, we present our main framework, which utilizes an atlas image as a guide for WMH segmentation. We provide a more detailed explanation of the atlas attention module and the attention-guided nested U-Net module in this section, along with illustrations of the network structure and attention mechanism. Section V presents the validation of our approach using the 2017 MICCAI WMH segmentation challenge dataset. We compare our proposed NAUNet++ framework with three other baselines, namely UNet, UNet++, and BAGAU-Net, and demonstrate that NAUNet++ performs better in recall and f1 scores while converging faster than these three baselines. Finally, in Section VI, we conclude our work and describe future research directions.

II. RELATED WORK

Numerous studies have identified white matter hyperintensities (WMH) as a marker of cerebrovascular disease that is closely associated with cognitive impairment and dementia [16], [17], [18]. One study found that, in addition to neurodegenerative changes, patients with Alzheimer's disease exhibited WMH [19]. In a longitudinal study [20], a cognitively normal population had a larger WMH volume and more significant cognitive decline, which was independent of traditional risk factors for Alzheimer's disease and MRI-related imaging markers.

Research also suggests that WMH is associated with the intrinsic vulnerability of brain tissue to ischemic injury [21]. Moderate and severe WMH can reduce cerebral ischemia

tolerance, possibly due to insufficient microvascular brain reserve. As a result, in patients with moderate and severe WMH, rapid development of irreversible cerebral infarction after vascular obstruction can cause futile recanalization and increase the risk of bleeding. The studies by Charidimou and Shoamanesh have shown that patients with acute cerebral infarction and moderate to severe WMH were significantly more common than those without WMH, indicating that moderate and severe WMH can increase the risk of cerebral infarction [22].

Ghafoorian et al. employed CNN for white matter hyperintensity (WMH) segmentation and proposed several deep CNN architectures based on positional features [8]. Akkus et al. conducted quantitative analysis of brain MRI and categorized the network models into three types: CNN architecture trained by blocks, CNN architecture introduced by semantics, and cascaded CNN architecture. Rachmadi et al. proposed a method to incorporate spatial information into CNN networks [23]. Their proposed approach generated four images containing spatial position information by processing the MRI images, and then combined the four images and the original MRI as input. The experimental results showed that introducing global spatial information can address localization problems, although the Dice similarity coefficient (DSC) matrix computed was only around 0.5. Thus, there is still significant room for improvement in WMH segmentation with deep learning approaches.

Zhang et al. proposed a U-net-based post-processing technique for averaging and thresholding the U-net outputs in different random initializations [9]. This approach is independent of the model used and can be applied to other model structures, leading to improved accuracy of WMH segmentation. Wu et al. proposed a novel jump link U-net, which introduces a graph-based approach in the pre-processing stage to remove non-brain tissue and adds jump connections to the U-net model to improve segmentation accuracy [10]. Jeong et al. used a U-net model with expansion convolution to learn more context information on MRI slices through expansion convolution, enhancing the probability of the network identifying bulk WMH [11]. However, the segmentation accuracy achieved was only 0.56, indicating that the network structure based on deep learning requires further development. Zhang et al. proposed an attention U-net model named BAGAU-Net to improve the segmentation accuracy of WMH [12]. The proposed model leverages the attention mechanism to learn the essential features of the input image. The experimental results demonstrate that the proposed model can improve the segmentation accuracy of WMH. In this work, we further improve the results of BAGAU-Net by introducing the nested U-Net architecture and the nested attention mechanism into the task of WMH segmentation.

III. PRELIMINARIES

A. U-Net

Deep Convolutional Neural Networks (DCNN) have found widespread applications in various image processing tasks, such as image classification, visual recognition, and image segmentation [24]. However, training a deep CNN requires a significant amount of data, which poses a challenge in the case of medical images that are often limited in quantity. To address this issue, Ronneberger et al. proposed the U-Net architecture that employs data augmentation to use the limited annotated medical image samples more efficiently [15]. The U-Net is a fully convolutional network (FCN) that utilizes skip connections to preserve spatial information, and it has gained widespread adoption in medical image segmentation. Specifically, the U-Net comprises an encoder and a decoder, where the encoder constitutes a downsampling path, and the decoder constitutes an upsampling path that consists of a series of convolutional layers and max pooling layers. The skip connections connect the encoder and decoder to retain information from earlier layers, thereby preserving spatial information from the input image. Upon processing with the U-Net architecture, the output of the network is a pixel-wise classification map.

B. NESTED U-Net

Nested U-Net is a variant of the U-Net segmentation architecture that utilizes nested and dense skip connections to bridge the semantic gap [25]. In the original U-Net architecture, loose skip connections are used to connect the encoder and decoder, resulting in the fusion of semantically distinct features. In contrast, the Nested U-Net employs dense skip connections to enable the encoder and decoder to communicate more directly, thereby enhancing the integration of features at different scales.

The Nested U-Net architecture is composed of an encoder and decoder that are connected through skip pathways, which consist of a dense convolution block with several convolution layers where a concatenation layer precedes each convolution layer. The concatenation layer concatenates the output of the previous convolution layer with the input of the current convolution layer. Let $x^{i,j}$ denote the output of the layer where i is the index of the down-sampling layer and j is the index of the convolution layer of the dense convolution block. Given that $\mathcal{D}$ is the convolution block with the activation function, $\mathcal{P}$ is the max pooling operation, and $\mathcal{T}$ is the upsampling function, then $x^{i,j}$ can be calculated as shown in Equation (1).

$$x^{i,j} = \begin{cases} \mathcal{D}(\mathcal{P}\left(x^{i-1,j}\right)), & j = 0 \\ \mathcal{D}(\mathcal{P}\left(\left[\left[x^{i,k}\right]_{k=0}^{j-1}, \mathcal{T}\left(x^{i+1,j-1}\right)\right]\right)), & j > 0 \end{cases} \tag{1}$$

As shown in the equation, layer 0 receives only the input from the previous encode layer, and the layer $j > 0$ receives the input from the previous j layers in the same skip pathway and the up-sampled output $\mathcal{T}\left(x^{i+1,j-1}\right)$ from the lower skip pathway.

The Nested U-Net architecture offers several advantages over the original U-Net, including better preservation of fine details and improved segmentation accuracy. The dense skip

connections in the Nested U-Net enable the network to capture more contextual information and enhance the feature integration across different scales. Additionally, the nested skip connections in the Nested U-Net provide a hierarchical way of integrating features and can better handle objects of different scales.

C. ATTENTION U-Net

The attention gate model is an innovative neural network architecture that integrates the attention mechanism to focus on important regions of an input image, leading to improved localization accuracy of deep learning models by emphasizing salient features [26]. The attention gate model can be incorporated into both standard U-Net or Nested U-Net structures. However, U-Net and Nested U-Net architectures face challenges in reducing false-positive predictions for small objects. In contrast, the attention gate model can suppress feature responses in irrelevant background regions without cropping regions of interest between networks. The attention gate model scales the input features x^l of layer l with the attention coefficients α^l while the gating signal g is collected from a coarser scale. The attention coefficients are determined using Equation (2), where σ_1 represents the Relu activation function, σ_2 represents the sigmoid activation function, W_x, W_g, and ψ are weight matrices that perform linear transformations, and b_g and b_ψ are bias vectors. Standard back-propagation updates can be used to train the attention gate model [26].

$$\alpha^l = \sigma_2\left(\psi^T\left(\sigma_1\left(W_x^T x^l + W_g^T g + b_g\right)\right) + b_\psi\right) \quad (2)$$

Here the weight matrices are used to learn the importance of different features in the input image. The weight matrix W_x is applied to the input features x^l of layer l, while W_g is applied to the gating signal g collected from a coarser scale. The resulting transformed feature maps are then added together and passed through a Relu activation function represented by σ_1. Finally, the resulting attention coefficients α^l are obtained by applying another linear transformation ψ to the output, followed by a sigmoid activation function represented by σ_2. These attention coefficients are then used to scale the input features to selectively emphasize important regions of the input image. Moreover, the bias vectors are used to introduce an additional degree of freedom to the learned linear transformations performed by the weight matrices. The bias term b_g is added to the gating signal g collected from a coarser scale, while b_ψ is added to the output of the linear transformation ψ applied to the sigmoid output of the transformed features. The incorporation of bias terms into the model facilitates an adjustment in the output of the activation function, thus providing a means to enhance model performance via improved alignment with the training data.

IV. NESTED ATTENTION GUIDED UNet++

A. ATLAS IMAGE REGISTRATION

Segmentation of MRI images can be time-consuming and challenging. To overcome these difficulties, atlas-based image segmentation is commonly used in MRI, which involves using an atlas, a pre-labeled image dataset, to label structures of interest in a new image automatically. This method can improve efficiency and accuracy in image analysis, particularly for complex anatomical structures or large datasets. An atlas image is a pre-existing image or set of images that serve as a reference or template for image segmentation in medical imaging, particularly in MRI. Due to the distinct shapes and sizes of human organs, medical images can vary significantly across individuals, making it essential to account for inter-image variability. One prevalent approach for representing medical images is using an atlas image, which entails developing a specific model for the image population that learns parameters from the training dataset. Atlas-based segmentation methods can reduce the manual annotation and labeling required in the segmentation process. Moreover, using an atlas can improve the accuracy and consistency of the segmentation results, particularly in cases where the target anatomy has a high degree of inter-subject variability or complexity. Also, atlas images can facilitate comparing results across different subjects or studies, enabling more robust analysis and interpretation of MRI data.

In atlas-based MRI segmentation, a pre-existing atlas image is first registered to the target image. The atlas labels are then propagated to the target image based on the registration, resulting in a segmentation of the target image. Our work, following the method described in [12], utilized the MNI152 database [27] as the standard reference space and applied Elastix [28] with a standard gradient descent optimizer and B-spline interpolator to register the MNI152 atlas to the training samples of MRI images.

B. FRAMEWORK

In this study, we propose a novel framework, named Nested Attention guided UNet++ (NAUNet++), which utilizes an atlas image to facilitate the segmentation of the FLAIR image. By incorporating an atlas image into the segmentation process, NAUNet++ aims to provide a more robust and accurate means of segmenting the FLAIR image. As depicted in Figure 1, the framework comprises two key modules: the atlas attention module and the attention-guided nested U-Net module. The former generates the atlas attention map, which serves as input for the latter. The attention-guided nested U-Net module subsequently generates the segmentation map of the FLAIR image. During the training process, both the FLAIR image and the atlas image are fed into the NAUNet++ framework, with the atlas image being directed to the atlas attention module and the FLAIR image being processed by the attention-guided nested U-Net module.

The atlas attention module in the proposed NAUNet++ framework is an adaptation of the UNet architecture, comprising an encoder and a decoder. The encoder comprises four down-sampling layers, with each layer containing two convolution layers followed by a batch normalization layer and a ReLU activation function. The down-sampling

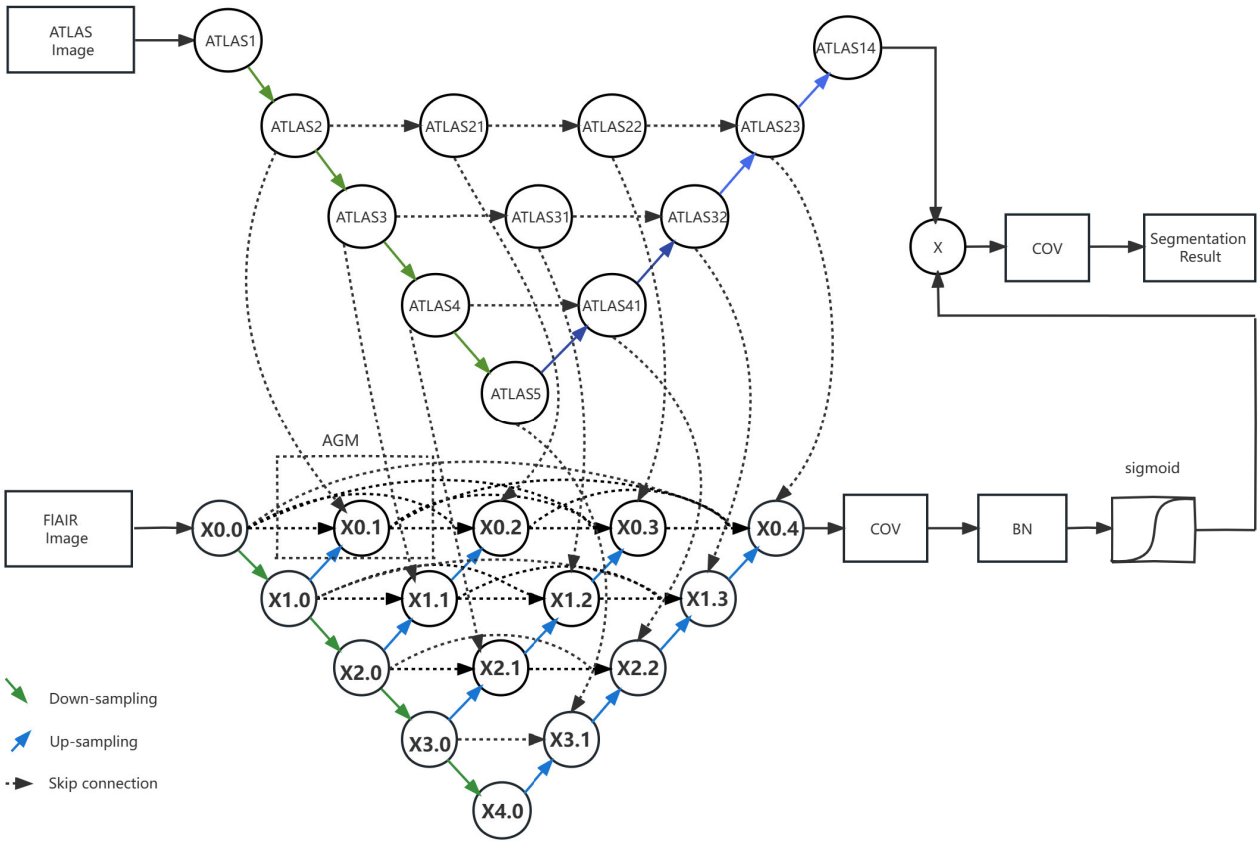


FIGURE 1. The overall framework for nested attention guided UNet++ (NAUNet++).

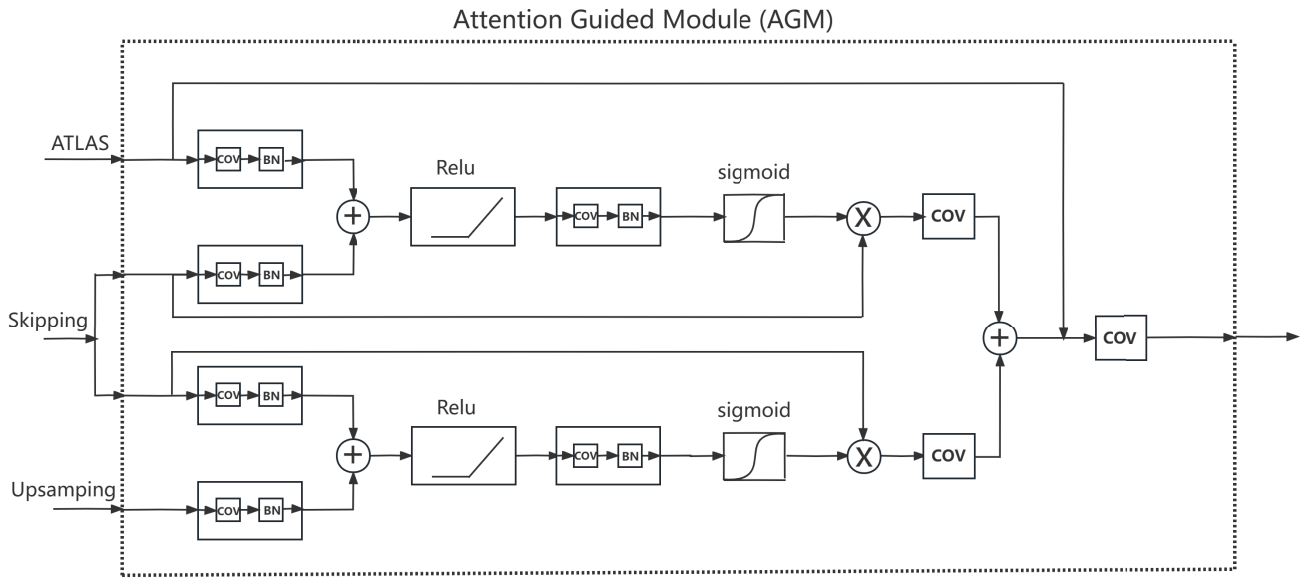


FIGURE 2. The attention guided module of the nested attention-guided nested U-Net module.

layers are also connected to max pooling layers. The decoder, on the other hand, comprises four up-sampling layers, each containing two convolution layers followed by a batch normalization layer and a ReLU activation function.

The encoder and decoder are connected via convolutional layers that serve as skip connections. For instance, Atlas 2 is obtained by down-sampling Atlas 1, while Atlas 21 is processed with a double convolutional layer following Atlas 2. Each feature map is used as supplementary features during

TABLE 1. The input and output channels of downsampling and upsampling paths.

downsampling	[input,output]	upsampling	[input,output]
atlas1->atlas2	[64,96]	atlas5->atlas4	[512,256]
atlas2->atlas3	[96,128]	atlas4->atlas3	[256,128]
atlas3->atlas4	[128,256]	atlas3->atlas2	[128,96]
atlas4->atlas5	[256,512]	atlas2->atlas1	[96,64]

the upsampling process. The input and output channels of the downsampling and upsampling paths are summarized in Table 1.

The attention-guided nested U-Net module in the proposed NAUNet++ framework adapts the nested UNet architecture, which includes four downsampling and four upsampling layers. Each downsampling layer comprises two convolution layers followed by a batch normalization layer, ReLU activation function, and a max pooling layer. Similarly, each upsampling layer comprises two convolution layers followed by a batch normalization layer and ReLU activation function.

The attention-guided nested U-Net module also uses loose skip connections to connect the downsampling and upsampling layers with the feature map generated from the atlas attention module.We develop an attention-guided module for each skip connection, which takes the upsampling path, skipping path, and atlas path as inputs to generate segmentation feature maps. Let $x^{i,j}$ denote the output of the layer, where i represents the index of the downsampling layer, and j represents the index of the output of the attention guided module. The downsampling and upsampling operations are represented by $\mathcal{E}$ and $\mathcal{F}$, respectively. Let the attention guided module denoted as $\mathcal{G}(up, skip, atlas)$ that takes three input, namely the upsampling path up, skipping path $skip$, and atlas path $atlas$. Then $x^{i,j}$ in the NAUNet++ framework can be calculated as Equation (3), which shows that the output feature maps for each layer can be calculated based on the feature maps from the previous layers and the atlas image. In the first case, when $j = 0$, the output feature map $x^{i,j}$ is obtained by applying the down-sampling operation $\mathcal{E}$ to the output feature map from the previous down-sampling layer, $x^{i-1,j}$. In the second case, when $j > 0$, the output feature map $x^{i,j}$ is obtained by applying the attention-guided module $\mathcal{G}$ to a concatenation of three input feature maps: the up-sampled feature map from the next up-sampling layer $\mathcal{F}(x^{i+1,j-1})$, the concatenation of all the previous output feature maps from the current down-sampling layer $[x^{i,k}]_{k=0}^{j-1}$, and the atlas image from two levels down $atlas^{i+2,j-1}$. Table 2 shows our design for the number of input channels, output channels, attention channels, encoding channels for the attention-guided module in different layers.

$$x^{i,j} = \begin{cases} \mathcal{E}\left(x^{i-1,j}\right), & j = 0 \\ \mathcal{G}\left(\left[\mathcal{F}(x^{i+1,j-1}), \left[x^{i,k}\right]_{k=0}^{j-1}, atlas^{i+2,j-1}\right]\right), & j > 0 \end{cases} \tag{3}$$

In the end, we pass the output of the attention-guided nested U-Net module through a convolution layer followed by a batch normalization layer and a ReLu activation function, which is then multiplied with the output of the atlas attention module using element-wise multiplication. The final output is obtained by passing the result through a convolution layer.

C. ATTENTION GUIDED MODULE

In ou proposed model, the attention-guided nested U-Net module utilizes an Attention Guided Module (AGM) to compute the target features through the upsampling path, skipping path, and atlas path. Our AGM design, as illustrated in Figure 2, incorporates two attention blocks from [26], with one block dealing with the upsampling path and the other with the atlas path. The first attention block uses the upsampling path as the gating signal and the skipping path as the input features, while the second attention block uses the atlas path as the gating signal and the skipping path as the input features. We denote the attention coefficients for the upsampling path and atlas path as the gating signal, respectively, at layer l with α^l_{up} and α^l_{atlas}, which can be calculated based on Equation (2). Using x^l to represent the skipping path and $\mathcal{V}$ to represent the convolutional operation, the output of AGM $\mathcal{F}^l$ can be expressed as Equation (4).

$$\mathcal{F}^l = \mathcal{V}(\mathcal{V}(x^l \times \alpha^l_{up}) + \mathcal{V}(x^l \times \alpha^l_{atlas})) \tag{4}$$

As can be seen from the equation, AGM first multiplies the input feature x^l by the attention coefficient α^l_{up} for the upsampling path, and then applies the convolutional operation $\mathcal{V}$ on the result. Similarly, the AGM multiplies the input feature x^l by the attention coefficient α^l_{atlas} for the atlas path, and then applies the convolutional operation $\mathcal{V}$ on the result. The two resulting feature maps are added together using the element-wise addition operation, and the final output of the AGM at layer l is obtained by applying the convolutional operation $\mathcal{V}$ on the sum of the two feature maps.

To summarize, we incorporate the attention machanism and nested UNet architecture for WMH segmentation, which could have the following advantages over typical UNet architecture:

- By incorporating an attention module with atlas images, our model can selectively focus on specific regions or features during segmentation. Additionally, the nested UNet architecture includes multiple levels of nested subnetworks, allowing for more refined feature extraction and segmentation. This can lead to improved accuracy in WMH segmentation.
- The nested UNet architecture is adept at detecting small objects that may be overlooked by standard UNet models. This feature is particularly useful for WMH segmentation as small areas may be present.
- The nested UNet architecture is robust to variations in input data, such as differences in size of the structure of interest. This robustness is achieved through the incorporation of skip connections and residual connections,

TABLE 2. The number of input channels, output channels, attention channels, encoding channels for the attention-guided module in different layers, each cell is presented as [input, output, attention, encoding].

	1	2	3	4
layer 0	[160,64,32,96]	[224,128,32,96]	[448,256,32,192]	[896,512,32,384]
layer 1	[224,96,64,128]	[320,192,64,128]	[640,384,64,256]	-
layer 2	[384,128,96,256]	[512,256,96,256]	-	-
layer 3	[768,256,128,512]	-	-	-

which help to reduce overfitting and improve the generalization of the model.

V. EXPERIMENTS

A. EXPERIMENT SETUP

In this section, we present our methodology for validating the effectiveness of our approach. Initially, we introduce the dataset and the evaluation metrics. Subsequently, we delineate the experimental settings, which include the ablation study and segmentation result analysis. The 2017 MICCAI WMH segmentation challenge dataset [29] was employed in this study to assess the efficiency of our proposed method. A detailed description of the dataset is presented in Table 3. Specifically, the dataset encompasses 60 sets of brain MRI images from various scanners, each consisting of 20 samples. Each sample contains different layers of the MRI image, all together 2528 images. Figure 3 showcases two distinct layers of the same MRI image, with the first column portraying the bias field-corrected FLAIR image, while the second column demonstrates the bias field-corrected T1 image, aligned with the corresponding FLAIR images. We implemented our models to the images to perform WMH segmentation. We then implemented common data augmentation techniques, namely scaling, shearing, and rotation, to expand the dataset size. This process resulted in the creation of a dataset comprising 10112 images.

In this experiment, the dataset was split into three parts for model development: 80% was used as the training set, 10% as the testing set, and 10% as the validation set. To improve the accuracy of segmentation models, Tversky loss was utilized as a common loss function in image segmentation tasks. As shown in previous studies [30], Tversky loss measures the dissimilarity between the predicted segmentation mask and the ground truth mask based on the number of true positives (TP), false positives (FP), and false negatives (FN), instead of solely relying on the overlap between the two masks, which is the approach used in DSC loss. The formal definition of Tversky loss, denoted as $\mathcal{L}_t$, is presented in Equation (5).

$$\mathcal{L}_t = \frac{TP}{TP + \zeta FP + (1-\zeta)FN} \tag{5}$$

Here the hyperparameter ζ controls the relative importance of false positives and false negatives, respectively. We take $\zeta = 0.7$ in the experiment.

We conducted a comparative analysis of our proposed NAUNet++ approach with four other frameworks, namely the UNet architecture, UNet++ architecture, and BAGAU-Net [12]. Here the UNet architecture and UNet++ architecture does not incorporate the attention mechanism with atlas images. The BAGAU-Net architecture incorporates attention mechanism into the UNet architecture. We also implemented the concatenation of feature maps under UNet++ architecture, denoted as CUNet++ to validate whether the nested architecture and attention mechanism can enhance the segmentation outcome for WMH. We implemented the model using PyTorch and trained the model on 4*Nvidia 1080Ti GPU with 11GB memory. The hyperparameter settings adopted for the experiment are presented in Table 4. The Adam optimizer [31] was employed with a batch size of 32. The model was trained for 20 epochs.

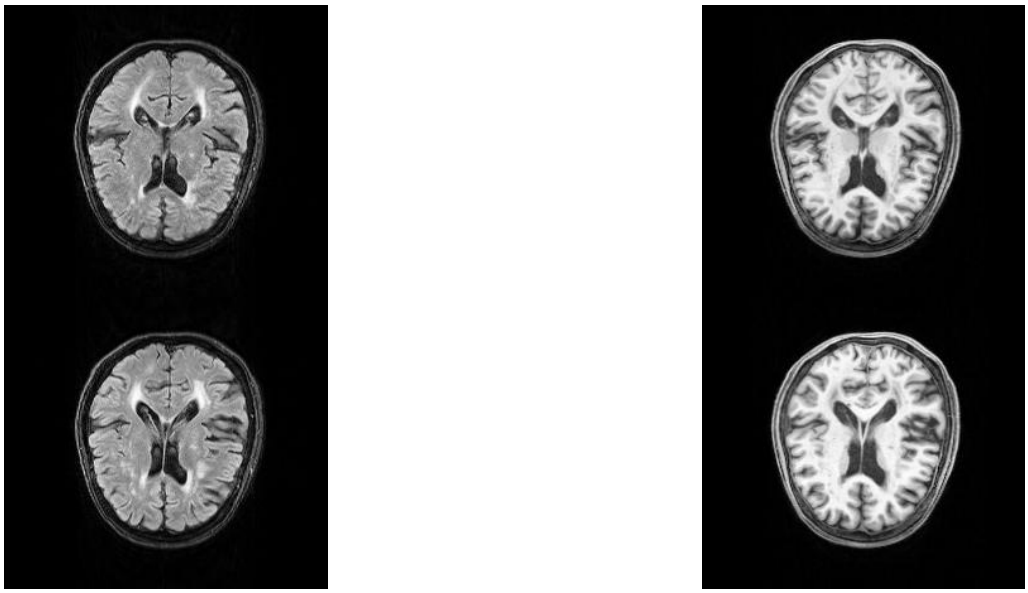
TABLE 3. The dataset name and number of training samples in each dataset for the experiment.

Dataset	Training Samples
UTrecht	20
Singapore	20
Amsterdam GE3T	20

This work mainly uses four evaluation metrics to evaluate the segmentation result, including DSC, Area Under Curve (AUC), recall and f1 scores. The DSC served as the primary statistical validation metric to evaluate the spatial overlap accuracy between the segmentation result and the ground truth [32]. The DSC was computed as the ratio of the number of true positive regions correctly classified as belonging to the structure of interest to the total number of regions in both the predicted and ground-truth segmentations. Using A and B to denote the segmentation result and ground truth, respectively, the DSC is expressed in Equation (6). Here the symbols $|A|$ and $|B|$ represent the total number of pixels in the predicted and ground truth segmentations, respectively, while $|A \cap B|$ represents the number of pixels that are correctly classified as belonging to the structure of interest in both the predicted and ground truth segmentations. A higher DSC value indicated superior segmentation performance, and it ranged from 0 to 1.

$$DSC = \frac{2\|A \cap B\|}{\|A\| + \|B\|} \tag{6}$$

The AUC is used to evaluate the segmentation result in terms of the probability of the segmentation result, commonly used in medical imaging to evaluate the accuracy of image segmentation and classification tasks. The recall and f1 scores are used to evaluate the segmentation result regarding the segmentation accuracy. Given that the number of detected WMH as N_{WMH}, the number of ground truth WMH as N_{GT},



(a) (b)

FIGURE 3. The two different layers of the same MRI image sample. The first column (a) shows the bias field corrected FLAIR images, and the second column (b) shows the bias field corrected T1 image, aligned with the corresponding FLAIR images.

TABLE 4. The hyperparameter settings for the experiment.

parameter	explanation	value
epoch	number of epochs	20
batch size	batch size used for the experiment	32
learning rate	learning rate	1e-3,6e-4,2e-4
gradient accumulation steps	Number of updates steps to accumulate before performing a backward/update pass	2
ζ	ζ value for various loss functions	0.7

and the number of all predictions as N_{all}, the recall is defined as Equation (7) and the f1 score is defined as Equation (8). The f1 scores are usually used as a measure of the balance betwen precision and recall.

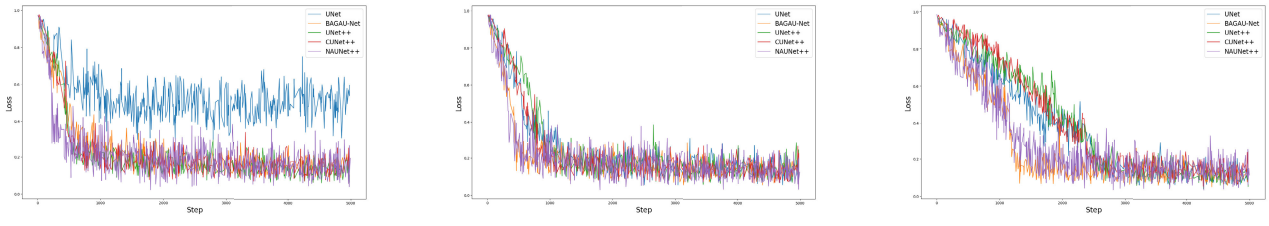
$$\begin{cases} recall = \dfrac{N_{WMH}}{N_{GT}} \\ precision = \dfrac{N_{WMH}}{N_{all}} \end{cases} \tag{7}$$

$$f1 = \frac{2 \times recall \times precision}{recall + precision} \tag{8}$$

B. ABLATION STUDY

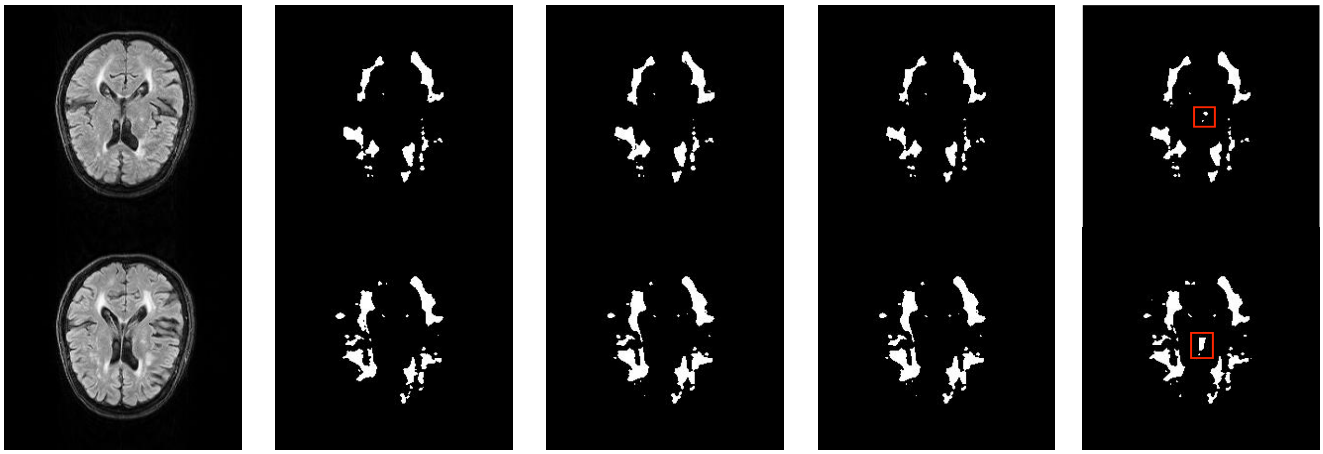
In this section, an ablation study was conducted to evaluate the effectiveness of the proposed NAUNet++ architecture in comparison to three other UNet-based frameworks, namely UNet, BAGAU-Net, UNet++, and CUNet++. To compare the convergence rate of the four architectures during the training process under different learning rates, Figure 4 was plotted. As shown in Figure 4(a), the UNet model converged to sub-optimal solutions under the learning rate 1e-3. The convergence rate indicates the speed at which the model's segmentation performance improves during training. From different learning rate settings, it was observed that NAUNet++ and BAGAU-Net converged faster compared to the other three models, suggesting that the attention mechanism could help the model to converge faster.

Table 5 presents the AUC, DSC, F1, and Recall scores for the four frameworks evaluated in this study. The results indicate that NAUNet++ outperforms the other three frameworks, with higher AUC, F1, and Recall scores. Specifically, the AUC score for NAUNet++ is 0.97, the highest among all models, indicating its superior ability to distinguish between different classes of voxels. Additionally, NAUNet++ has the highest F1 score of 0.86, indicating a good balance of true positive and false positive predictions. Moreover, NAUNet++ also has the highest recall score of 0.94, which is 6% higher than the best score achieved by the other three approaches. This result suggests that the NAUNet++ model has the highest true positive rate among all models. Based on these results, we conclude that NAUNet++ is capable of achieving higher accuracy in detecting WMH and improving the segmentation accuracy.



(a) learning rate = 1e-3 (b) learning rate = 6e-4 (c) learning rate = 2e-4

FIGURE 4. Comparison of convergence rate of four frameworks, namely UNet, BAGAU-Net, UNet++, CUnet++ and NAUNet++ under different learning rate.



(a) (b) (c) (d) (e)

FIGURE 5. Segmentation result of Figure 3. The first column (a) shows the bias field corrected FLAIR images, the second column (b) shows the segmentation result of the FLAIR images with UNet model, the third column (c) shows the segmentation result of the FLAIR images with BAGAU-net model, the forth column (d) shows the segmentation result of the FLAIR images with UNet++ model, the fifth column (e) shows the segmentation result of the FLAIR images with NAUNet++ model.

TABLE 5. The result of AUC, DSC, F1, and Recall matrix for the five frameworks, namely UNet, UNet++, CUNet++, BAGAU-Net and NAUNet++.

	AUC	DSC	F1	RECALL
UNet	0.94	**0.89**	0.75	0.88
BAGAU-Net	0.93	0.86	0.83	0.88
UNet++	0.95	0.88	0.80	0.76
CUNet++	0.95	0.88	0.81	0.82
NAUNet++	**0.97**	0.88	**0.86**	**0.94**

In terms of DSC, all models perform similarly, with UNet, UNet++, and NAUNet++ having the highest DSC of 0.88 and 0.89, respectively. Among the models, UNet++ has a slightly lower recall, indicating a lower true positive rate. BAGAU-Net exhibits similar AUC and recall performance as UNet, while UNet has a lower f1 score. A similar trend is observed when comparing UNet++ and NAUNet++, where NAUNet++ outperforms in terms of f1 and recall scores. Introducing the attention gate model to the UNet structure can achieve a better balance between true positive and false positive predictions. Additionally, the results of BAGAU-Net and NAUNet++ demonstrate that the nested structure of UNet can improve the performance of model.

UNet++ and CUNet++ have similar performances across AUC and DSC, but CUNet++ has higher scores of f1 and recall, suggesting that incorperating atlas image could help to achieve a better balance between true positive and false positive predictions. It is also observed that the difference in DSC score between UNet and NAUNet++ is relatively small, but the BAGAU-Net has lower f1 and recall scores compared with NAUNet++, which shows that BAGAU-Net is less effective at correctly identifying the segmentation areas.

In Figure 5, the segmentation results from the four different models are presented for the example image shown in Figure 3. The comparison of the predicted segmentation with the original image reveals that the NAUNet++ model can capture more WMH information than the other three models.

VI. CONCLUSION

In this study, we introduce a novel approach for White Matter Hyperintensity (WMH) segmentation, called Nested Attention guided U-Net++ (NAUNet++). The proposed

NAUNet++ comprises two modules: the atlas attention module and the attention-guided nested U-Net module. The atlas attention module generates the atlas attention map, which serves as input to the attention-guided nested U-Net module, which generates the segmentation map of the FLAIR image. The learning curve of different evaluation metrics shows that the NAUNet++ can learn faster than the other three frameworks. The results of the AUC, DSC, F1, and Recall metrics indicate that the NAUNet++ can achieve higher accuracy in detecting WMH and segmentation accuracy. Therefore, the proposed NAUNet++ can be employed for the detection of WMH in brain MRI images with improved performance.

As future work, we plan to extend the proposed NAUNet++ to brain MRI images with different modalities, including T1, T2, and FLAIR. Additionally, we aim to explore the transferability of the model to brain MRI images acquired from different scanners or annotated by different experts.

REFERENCES

[1] G. Sibolt, S. Curtze, S. Melkas, T. Pohjasvaara, M. Kaste, P. J. Karhunen, N. K. J. Oksala, and T. Erkinjuntti, "Severe cerebral white matter lesions in ischemic stroke patients are associated with less time spent at home and early institutionalization," *Int. J. Stroke*, vol. 10, no. 8, pp. 1192–1196, Dec. 2015.

[2] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017.

[3] M. E. Caligiuri, P. Perrotta, A. Augimeri, F. Rocca, A. Quattrone, and A. Cherubini, "Automatic detection of white matter hyperintensities in healthy aging and pathology using magnetic resonance imaging: A review," *Neuroinformatics*, vol. 13, no. 3, pp. 261–276, Jul. 2015.

[4] C. Zhu, Z. Zhu, Y. Xie, W. Jiang, and G. Zhang, "Evaluation of machine learning approaches for Android energy bugs detection with revision commits," *IEEE Access*, vol. 7, pp. 85241–85252, 2019.

[5] R. Ranjbarzadeh, A. B. Kasgari, S. J. Ghoushchi, S. Anari, M. Naseri, and M. Bendechache, "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," *Sci. Rep.*, vol. 11, no. 1, pp. 1–17, May 2021.

[6] R. McKinley, R. Wepfer, F. Aschwanden, L. Grunder, R. Muri, C. Rummel, R. Verma, C. Weisstanner, M. Reyes, A. Salmen, A. Chan, F. Wagner, and R. Wiest, "Simultaneous lesion and brain segmentation in multiple sclerosis using deep neural networks," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, Jan. 2021.

[7] S. Roy and S. K. Bandyopadhyay, "A new method of brain tissues segmentation from MRI with accuracy estimation," *Proc. Comput. Sci.*, vol. 85, pp. 362–369, Jan. 2016.

[8] M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. M. van Uden, C. I. Sanchez, G. Litjens, F.-E. de Leeuw, B. van Ginneken, E. Marchiori, and B. Platel, "Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities," *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, Jul. 2017.

[9] Y. Zhang, W. Chen, Y. Chen, and X. Tang, "A post-processing method to improve the white matter hyperintensity segmentation accuracy for randomly-initialized U-Net," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.

[10] J. Wu, Y. Zhang, K. Wang, and X. Tang, "Skip connection U-Net for white matter hyperintensities segmentation from MRI," *IEEE Access*, vol. 7, pp. 155194–155202, 2019.

[11] Y. Jeong, M. F. Rachmadi, M. D. C. Valdés-Hernández, and T. Komura, "Dilated saliency U-Net for white matter hyperintensities segmentation using irregularity age map," *Frontiers Aging Neurosci.*, vol. 11, p. 150, Jun. 2019.

[12] Z. Zhang, K. Powell, C. Yin, S. Cao, D. Gonzalez, Y. Hannawi, and P. Zhang, "Brain atlas guided attention U-Net for white matter hyperintensity segmentation," *AMIA Summits Transl. Sci.*, vol. 2021, p. 663, May 2021.

[13] Z. Xu and M. Niethammer, "DeepAtlas: Joint semi-supervised learning of image registration and segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*. Shenzhen, China: Springer, Oct. 2019, pp. 420–429.

[14] U. Wickramasinghe, G. Knott, and P. Fua, "Probabilistic atlases to enforce topological constraints," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*. Shenzhen, China: Springer, Oct. 2019, pp. 218–226.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* New York, NY, USA: Springer, 2015, pp. 234–241.

[16] E. van den Berg, M. I. Geerlings, G. J. Biessels, P. J. Nederkoorn, and R. P. Kloppenborg, "White matter hyperintensities and cognition in mild cognitive impairment and Alzheimer's disease: A domain-specific meta-analysis," *J. Alzheimer's Disease*, vol. 63, no. 2, pp. 515–527, Apr. 2018.

[17] G. Tosto, M. E. Zimmerman, J. L. Hamilton, O. T. Carmichael, and A. M. Brickman, "The effect of white matter hyperintensities on neurodegeneration in mild cognitive impairment," *Alzheimer's Dementia*, vol. 11, no. 12, pp. 1510–1519, Dec. 2015.

[18] H.-Y. Hu, Y.-N. Ou, X.-N. Shen, Y. Qu, Y.-H. Ma, Z.-T. Wang, Q. Dong, L. Tan, and J.-T. Yu, "White matter hyperintensities and risks of cognitive impairment and dementia: A systematic review and meta-analysis of 36 prospective studies," *Neurosci. Biobehav. Rev.*, vol. 120, pp. 16–27, Jan. 2021.

[19] A. Kapasi, C. DeCarli, and J. A. Schneider, "Impact of multiple pathologies on the threshold for clinically overt dementia," *Acta Neuropathol.*, vol. 134, no. 2, pp. 171–186, Aug. 2017.

[20] C. Puzo, C. Labriola, M. A. Sugarman, Y. Tripodis, B. Martin, J. N. Palmisano, E. G. Steinberg, T. D. Stein, N. W. Kowall, A. C. McKee, J. Mez, R. J. Killiany, R. A. Stern, and M. L. Alosco, "Independent effects of white matter hyperintensities on cognitive, neuropsychiatric, and functional decline: A longitudinal investigation using the national Alzheimer's coordinating center uniform data set," *Alzheimer's Res. Therapy*, vol. 11, no. 1, pp. 1–13, Dec. 2019.

[21] N. Gilberti, M. Gamba, E. Premi, A. Costa, V. Vergani, I. Delrio, R. Spezi, M. Dikran, M. Frigerio, R. Gasparotti, A. Pezzini, A. Padovani, and M. Magoni, "Leukoaraiosis is a predictor of futile recanalization in acute ischemic stroke," *J. Neurol.*, vol. 264, no. 3, pp. 448–452, Mar. 2017.

[22] A. Charidimou and A. Shoamanesh, "Clinical relevance of microbleeds in acute stroke thrombolysis: Comprehensive meta-analysis," *Neurology*, vol. 87, no. 15, pp. 1534–1541, Oct. 2016.

[23] M. F. Rachmadi, M. D. C. Valdés-Hernández, M. L. F. Agan, C. Di Perri, and 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. Imag. Graph.*, vol. 66, pp. 28–43, Jun. 2018.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[25] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Berlin, Germany: Springer, 2018, pp. 3–11.

[26] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[27] B. McConnell, B. Pike, C. Holmes, L. Aparicio, A. C. Evans, and D. L. Collins. (2009). *ICBM152NLin2009: A Neuroanatomically Constrained Nonlinearly Registered Template Based on the 2009 Version of the ICBM Average Brain*. [Online]. Available: http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009

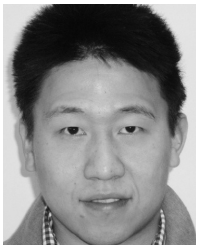
[28] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 196–205, Jan. 2010.

[29] H. J. Kuijf et al., "Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 11, pp. 2556–2568, Nov. 2019.

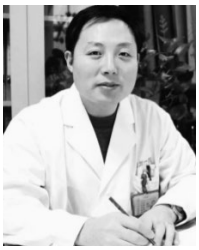
[30] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *Machine Learning in Medical Imaging*. Quebec City, QC, Canada: Springer, Sep. 2017, pp. 379–387.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[32] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. C. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index1," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, Feb. 2004.

HAO ZHANG received the bachelor's degree in medicine and the master's degree in neurology from the School of Medicine, Shanghai Jiao Tong University, Shanghai, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree in endocrinology and metabolism with Soochow University, China. Since 2016, he has been a Clinician with the Department of Neurology, The Third Affiliated Hospital of Soochow University.



CHENYANG ZHU received the B.S. degree in telecommunication engineering from the Huazhong University of Science and Technology, Hubei, China, in 2012, the M.S. degree in embedded systems from the University of Pennsylvania, Philadelphia, PA, USA, in 2014, and the Ph.D. degree in computer science from the University of Southampton, Southampton, U.K., in 2020. Since 2020, he has been an Associate Professor with the School of Computer Science and Artificial Intelligence, Changzhou University.



XUEGAN LIAN received the M.D. degree in neurology from the Jinling Hospital, Nanjing, China, in 2011. Since 2019, he has been an Associate Professor with The Third Affiliated Hospital of Soochow University.

FEI HUA received the M.D. degree in endocrinology and metabolism from Soochow University, Soochow, China, in 2016. Since 2018, he has been a Professor with The Third Affiliated Hospital of Soochow University.