Original Article

Ensemble learning accurately predicts the potential benefits of thrombolytic therapy in acute ischemic stroke

Zhihong Chen[1#], Qingqing Li[2#], Renyuan Li[3,4#], Han Zhao[1], Zhaoqing Li[4,5], Ying Zhou[2], Renxiu Bian[3], Xinyu Jin[1], Min Lou[2], Ruiliang Bai[3,4,5]

[1]Institute of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China; [2]Department of Neurology, The Second Affiliated Hospital of Zhejiang University, School of Medicine, Hangzhou, China; [3]Department of Physical Medicine and Rehabilitation, The Affiliated Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China; [4]Interdisciplinary Institute of Neuroscience and Technology, School of Medicine, Zhejiang University, Hangzhou, China; [5]Key Laboratory of Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

Contributions: (I) Conception and design: R Bai, M Lou, X Jin, Z Chen; (II) Administrative support: R Bai, M Lou, X Jin; (III) Provision of study materials or patients: R Bai, M Lou; (IV) Collection and assembly of data: Q Li, R Li; (V) Data analysis and interpretation: Z Chen, H Zhao, Y Zhou, R Bian; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Dr. Ruiliang Bai. 268 Kaixuan Road, South Central Building, Room 708, Hangzhou 310027, China. Email: ruiliangbai@zju.edu.cn; Dr. Min Lou. Zhejiang University, Hangzhou 310027, China. Email: lm99@zju.edu.cn; Dr. Xinyu Jin. Zhejiang University, Hangzhou 310027, China. Email: jinxy@zju.edu.cn.

Background: Finding methods to accurately predict the final infarct volumes for acute ischemic stroke patients with full or no recanalization would significantly help to evaluate the potential benefits of thrombolytic therapy. We proposed such a method by constructing a model of ensemble deep learning and machine learning using diffusion-weighted imaging (DWI) only.

Methods: The proposed prediction model (named AUNet) combines an adaptive linear ensemble model (ALEM) of machine learning and a deep U-Net network with an accelerated non-local module (U-NL-Net) to learn voxel-wise and spatial features, respectively. Of 40 patients with acute ischemic stroke who received thrombolytic therapy, 17 were fully recanalized, 14 were not recanalized, and nine were partially recanalized. The AUNet was separately trained for full recanalization conditions (AUNetR) and no recanalization (AUNetN) as the best and worst outcomes of thrombolysis, respectively.

Results: AUNet performed significantly better in predicting the final infarct volumes in both the recanalization and non-recanalization conditions [area under the receiver operating characteristic curve (AUC) =0.898±0.022, recanalization; AUC =0.875±0.036, non-recanalization: Matthew's correlation coefficient (MCC) =0.863±0.033, recanalization; MCC =0.851±0.025, non-recanalization] than the fixed-thresholding method (AUC =0.776±0.021, $P<0.0001$, recanalization; AUC =0.692±0.023, $P<0.0001$, non-recanalization: MCC =0.742±0.035, recanalization; MCC =0.671±0.024, non-recanalization), the logistic regression method (AUC =0.797±0.023, $P<0.003$, recanalization; AUC =0.751±0.030, $P<0.003$, non-recanalization: MCC =0.762±0.035, recanalization; MCC =0.730±0.031, non-recanalization), and a recently developed convolutional neural network (AUC =0.814±0.013, $P<0.003$, recanalization; AUC =0.781±0.027, $P<0.003$, non-recanalization: MCC =792±0.022, recanalization; MCC =0.758±0.016, non-recanalization). The potential benefit of thrombolysis calculated from AUNetR and AUNetN showed large individual differences (from 12.81% to 239.73%)

Conclusions: AUNet improved predictive accuracy over current state-of-the-art methods. More importantly, the accurate prediction of infarct volumes under different recanalization conditions may provide benefitial information for physicians in selecting suitable patients for thrombolytic therapy.

Keywords: Infarct volume prediction; thrombolytic therapy; recanalization; ensemble learning; convolutional neural networks; computer-aided diagnosis

Submitted Jan 08, 2021. Accepted for publication Apr 16, 2021.
doi: 10.21037/qims-21-33
View this article at: http://dx.doi.org/10.21037/qims-21-33

Introduction

Stroke is one of the leading causes of death worldwide, and almost half of stroke deaths result from ischemic stroke (1). Thrombolytic therapy, a procedure used for ischemic stroke, aims to recanalize the occluded vessel and reperfuse the ischemic tissue (2). However, as noted by Dr. Lyden, the decision to use thrombolytic therapy is "*among the most difficult treatment decisions in medicine, given the risks involved and the compressed time frame available*" (3). Complications related to thrombolysis include symptomatic intracranial hemorrhage, major systemic hemorrhage, and angioedema (4,5). Any decision to use thrombolytic therapy needs to take into account the risk of these complications and the potential benefits of thrombolytic therapy, including the volume of salvageable tissue. One way to quantitatively evaluate treatment benefit is to predict and compare the final infarct volume after full recanalization (best outcome) with that of no recanalization (worst outcome).

Several quantitative methods have been developed to predict final infarct volumes from baseline neuroimaging data; however, prediction methods for different recanalization levels are still lacking. The most popular method involves uniform thresholding on the baseline apparent diffusion coefficient (ADC) map, which is calculated from diffusion-weighted imaging (DWI) to determine the ischemic core (6,7). By using different thresholds to obtain receiver operating characteristic (ROC) curves classified by pixels in data from 14 stroke patients, it was previously found that the ADC of the final infarcted area was lower than that of the non-infarcted area and that the best segmentation threshold was 620×10^{-6} mm^2/s (7). A similar approach was used with magnetic resonance imaging (MRI)/CT perfusion-weighted images (PWIs) to obtain the optimized T_{max} (time point for the maximum of the residue function) threshold (6). The DWI/PWI mismatch is considered to represent salvageable tissue (8-11). However, tissue heterogeneity exists, and fixed thresholds might oversimplify the problem and limit the prediction accuracy. Recently, threshold-free predictive algorithms with the capability of including multimodal images have been developed, including the logistic regression (LR) model (12) and the general linear model (GLM) (13). These advanced machine-learning methods have substantially improved predictive accuracy. However, these methods are still performed pixel-by-pixel without full consideration of spatial information, limiting their accuracy.

The convolutional neural network (CNN) is an emerging technique that provides lesion segmentation applications and other imaging texture-recognition applications (14-16). CNN has the advantage of including spatial information, which makes CNN a suitable candidate for stroke progression assessment. However, studies using CNN for final infarct prediction of acute ischemic stroke (17), especially under different treatment conditions, are still limited. Nielsen *et al.* (17) implemented several CNN algorithms for predicting tissue outcome and assessing treatment effects in acute ischemic stroke; they found that deep CNN performed better than other CNNs, fixed-threshold methods (6,7), or GLMs (8). However, this study used patients receiving thrombolytic therapy to train the CNN models but did not assess the degree of recanalization, which is problematic as the degree of recanalization significantly affects the final tissue outcomes. Additionally, due to the excessively deep network structure, this method ignores the image's low-level information details, resulting in blurring the edges of the prediction area and overestimating the final infarct volume.

This study aimed to develop prediction models for the best (full recanalization) and worst (no recanalization) outcomes of thrombolytic therapy in acute ischemic stroke, using baseline DWI only. For this reason, we used two groups of stroke patients, one consisting of patients who had undergone full recanalization (recanalization group) and the other consisting of patients without recanalization (non-recanalization group), to train the prediction models for the best and worst outcomes, respectively. We further aimed to improve deep learning algorithms by incorporating different spatial scale information (e.g., fine structural information and long-distance information) and combining the U-Net network (17) with an accelerated non-local (NL) module (U-NL-Net) (18). Additionally, we combined machine learning, which has advantages in voxel-level sorting, with deep learning, which provides spatial information integration. This ensemble learning was used to make full use of local tissue information (voxel-wise) and spatial distribution information to improve prediction accuracy. Further, we compared the prediction performance of our proposed method (AUNet) with other state-of-the-art methods, including a fixed-thresholding method (6), GLM (12), and a CNN-based method (16) in patients with ischemic stroke who had full, partial, or no recanalization after thrombolytic therapy.

Methods

Patients and image acquisition

In this retrospective study, 40 patients with symptoms consistent with acute ischemic stroke and treated with thrombolytic therapy within 4.5 hours after stroke onset were selected. Baseline DWI images were collected before treatment, and T2-weighted-fluid-attenuated inversion recovery (T2-FLAIR) images or CT images were obtained 7 days after thrombolytic therapy. All MRIs were performed on a GE 3.0T scanner at the admitting hospital. Echo-planar DWIs were obtained at a diffusion weighting (*b*) of 0 s/mm^2 and 1,000 s/mm^2. The b =1,000 s/mm^2 images were acquired at 3 to 12 directions, depending on the scanner/vendor type at the admitting hospital. The MR image acquisition parameters for DWI were repetition time =4,000 ms, echo time =69.3 ms, b-value =1,000 s/mm^2, three slabs, slice thickness 5.0 mm, interslice gap 1.0 mm, spatial resolution of 0.94 mm/pixel, field of view (FOV) =24 cm, and matrix size =160×160. The two DWI images were used to obtain the ADC images. Also, digital subtraction angiography (DSA)/computed tomography perfusion (CTP)/magnetic resonance angiography (MRA) images were collected after thrombolytic therapy to evaluate the recanalization status. Because of the sensitive nature of the data and compliance regulations on general data protection, requests to access the dataset from qualified researchers trained in human subject confidentiality protocols may be sent to Zhejiang University. The human ethics committee approved this study protocol at our center. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Written informed consent was obtained from all patients or their designated proxy.

Patient groups and final infarct volume identification

After receiving thrombolytic therapy, 17 patients were fully recanalized [arterial occlusive lesion (AOL) score =3, recanalization group], 14 patients were not recanalized (AOL =0, non-recanalization group), and 9 patients were partly recanalized (AOL =2, partial-recanalization group). The final infarct volumes were marked by a neurologist on the 7-day T2-FLAIR or CT images, which were then used as labels for training and evaluation of the prediction models.

Image preprocessing

Because the grayscale values of DWI and ADC images differ greatly in magnitude, data normalization operations were performed on all images. Specifically, to make the model a better fit for the dataset's overall distribution, the mean and variance used in our normalization were calculated through the whole dataset. For the normalization of the DWI images, the mean value (DWI_{mean}) and standard deviation (SD) (DWI_{sd}) were computed by taking the mean and the SD of the brain region. The DWI signal intensity was normalized as the ratio of DWI signal intensity – DWI_{mean} to DWI_{sd}. For the normalization of the ADC images, the mean value (ADC_{mean}) was set to 600 (10^{-6} mm^2/s), which is a balance value of the mean ADC in the infarction core (400×10^{-6}–499×10^{-6} mm^2/s) and that in the non-core brain tissue (700×10^{-6}–799×10^{-6} mm^2/s) as reported by Jonsdottir *et al.* (19). The SD (ADC_{sd}) was computed by taking the SD of the brain region. The ADC signal intensity was normalized as the ratio of ADC signal intensity – ADC_{mean} to ADC_{sd}. DWI, ADC, and lesion mask images were resampled to a 256×256 resolution for all slices. All images were further co-registered onto the 7-day MRI T2-FLAIR or CT images using a fully automated and affine algorithm with FMRIB's Linear Image Registration Tool (FLIRT) (20).

The proposed prediction model: AUNet

Figure 1 shows the architecture of the proposed prediction model, AUNet. It is an ensemble learning method combining two prediction models: the adaptive linear ensemble model (ALEM), which is a shallow machine-learning model, and a

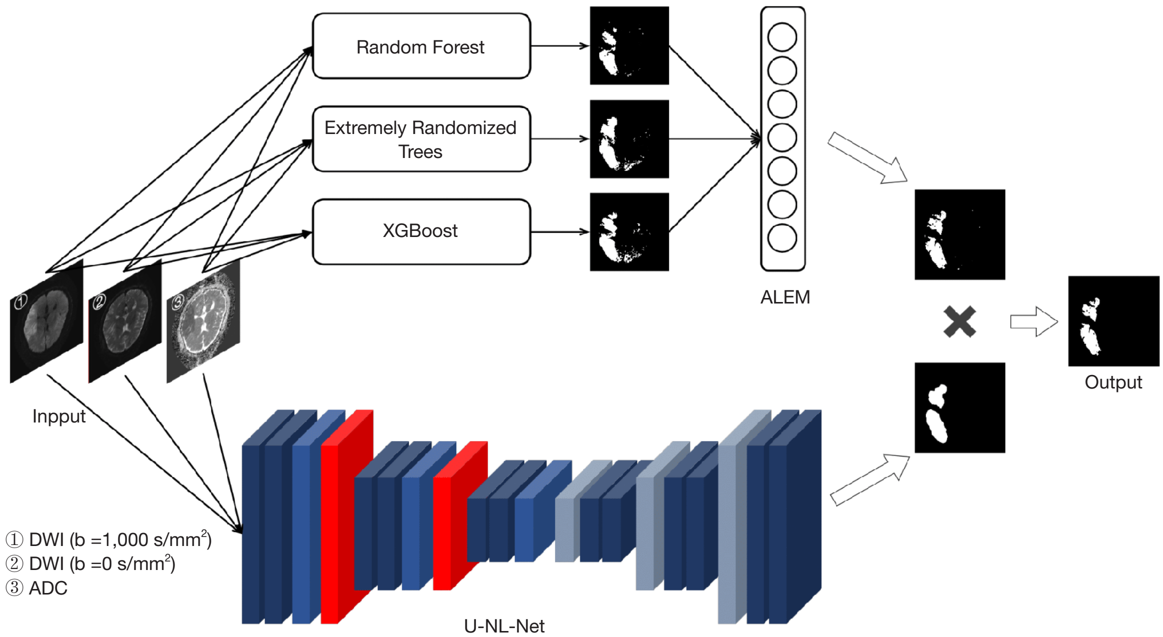

Figure 1 The architecture of the proposed AUNet method. The inputs of AUNet are baseline DWI and ADC images and the outputs of AUNet are the predicted infarct volume. The upper panel represents the adaptive linear ensemble model (ALEM), in which three machine-learning-based classifiers are combined. The lower panel illustrates the designed U-Net network with the accelerated non-local module U-NL-Net. The red module represents the non-local module (best viewed in color). The final prediction results were obtained by logic computation of the vetting results from the two channels. DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient.

deep U-Net network with an accelerated non-local module (U-NL-Net); these models learn the voxel-wise and spatial features, respectively. The final prediction results were obtained by combining the results from these two methods to obtain the AUNet model.

ALEM for voxel-wise selection

Due to the limitations of a deep model's learning ability to incorporate low-level information and data distribution diversity, it is not easy to find one model that can make accurate predictions for inputs with various distributions. Thus, we developed the ALEM, which combines three base classifiers for the voxel-wise prediction of the final infarct tissue: Random forest (21), extremely randomized trees (22), and XGBoost (23). ALEM was designed to combine these three different types of base classifiers with improving this model's robustness for cases in which one base classifier could not provide correct prediction values due to the sparsity of the data. For the three base classifiers, each forest contained 100 subtrees, and the maximum split depth of each tree was 10. ALEM used the mean decrease impurity (MDI) as the evaluation index of feature importance, which adaptively uses feature importance to assign weights to different models for fitting diverse data distributions. The details of implementation are provided in Supplement 1. The model implementation was based on the Scikit-Learn Algorithm Library.

U-NL-Net prediction model for spatial features learning

By observing the brain MRI image slices, it can be seen that a large proportion of patients exhibit a unilateral distribution or clustered distribution of the final infarct tissue. Therefore, the long-distance information, including the difference between the infarct side and the mirror side, or infarct region and normal-appearing region distant from the infarct region, is an important feature in determining the infarct area of the stroke. However, in the conventional convolutional operation used in the baseline network (17), the convolution kernel can only process local areas such as 3×3 or 5×5 pixels when extracting features. Generally, it is impossible to consider the low-level information and establish a connection between opposite sides' long-distance pixels. To include the long-distance information, we improved the conventional U-Net network (18) by adding a NL module (24,25) so that the new network, U-NL-Net, could efficiently model long-distance features while capturing low-level features that reflect the detailed information of the MRI images. More details of U-NL-Net are provided in the Supplementary information. We used two different pooling scales (26) to improve the NL module structure to create an accelerated NL module, which optimized the computational complexity of the original NL module for the current applications. The U-NL-Net model architecture was based on Dong's implementation in PyTorch and was performed on a workstation computer [CentOS Linux 7, Intel Xeon (R) Processor E5-2667, 32 CPU cores with 3.2 GHz, 254 GB RAM, 8 GB graphics processing unit (Tesla P4)]. The following parameters were chosen for the U-NL-Net model: the number of epochs =100, batch size =10, and Adam (27) optimizer with learning rate =0.01. The final loss function was the weighted summation of Dice Loss (28) and Focal Loss (29). To prevent the model from overfitting, deep neural network training often requires a large amount of data. Considering that there were fewer infarct volumes than normal-appearing volumes in our dataset, a series of image augmentation algorithms were used to enhance the data representation, such as random horizontal flip, random vertical flip, random affine change, random elastic transformation, random contrast adjustment, random add Gaussian noise and random position cropping.

Ensemble predictive results—AUNet

The U-NL-Net prediction network can obtain richer features through stacked convolutional layers, which can be regarded as a "deep" prediction model containing high-level information. The prediction results were obtained simultaneously through one calculation. ALEM belongs to a "shallow" model containing low-level information, and the results were obtained by voxel-by-voxel prediction. Therefore, the two models may provide complementary characteristics for data. The final prediction result was obtained by the product of each position's probability value in the two results provided by ALEM and U-NL-Net.

Predictions for best and worst outcomes and the evaluation of treatment benefits

The AUNet models were trained on the recanalization group and non-recanalization group data separately to predict the final infarct volume with full recanalization (best outcome, $AUNet_R$) and no recanalization (worst outcome, $AUNet_N$), respectively. Subjects in each group were randomly assigned to either the training set or the external testing set with a training/testing ratio of approximately 70:30. The difference between the predicted infarct volumes of $AUNet_R$ ($Volume_R$) and those of $AUNet_N$ ($Volume_N$) was defined as the potential benefit of thrombolytic therapy, as follows:

$$Benefit = \frac{Volume_N - Volume_R}{Volume_R} \times 100\% \qquad [1]$$

Statistical analysis

The infarct volume prediction accuracy was evaluated using the area under the ROC curve (AUC) and Matthew's correlation coefficient (MCC). The AUC was calculated according to the method used by Jonsdottir *et al.* (19), which has the advantage of being threshold-independent and tests whether positives (correct predictions) are ranked higher than negatives (incorrect predictions). The MCC is a robust method to calculate metric success in unbalanced datasets. The calculated AUC and MCC were pairwise compared using paired *t*-tests. The AUC and MCC are presented as mean ± SD.

Results

The characteristics of all included patients are summarized in *Table 1*. There was no significant difference among patients concerning these characteristics, which included age (P=0.919), time of onset to MRI (P=0.973), admission score on the National Institutes of Health Stroke Scale (NIHSS) (P=0.808), DWI volume (ADC <600 $\mu m^2/s$, P=0.5647), and PWI volume (T_{max} ≥6 s, P=0.1714).

Model comparisons

Figure 2 shows the final infarct volumes (in red) defined as the 7-day MRI/CT (fourth column, manually drawn by one neurologist) and predicted from different methods, including fixed-thresholding on ADC, the LR method, conventional CNN, ALEM, U-NL-Net, and AUNet (ALEM + U-NL-Net). In the final fusion result, the prediction error is lower. Visually, fixed-thresholding-based predictions led to a scattered distribution of infarct

Table 1 Patient characteristics

Characteristics	Non-recanalization median (min–max)	Recanalization median (min–max)	Partial-recanalization median (min–max)
Age, years	67.2 (55–82)	67.6 (47–81)	71 (53–81)
Time, onset to MRI, minutes	247 (121–350)	248 (118–390)	163 (70–309)
Admission NIHSS	15.3 (2.0–21.0)	15.0 (3.0–23.0)	12.4 (1.0–22.0)
DWI volume, mL	45.6 (1.0–124.8)	37.0 (1.0–144.9)	34.69 (1.0–109.3)
PWI volume, mL	128.4 (37.8–356.4)	93.4 (30.2–226.3)	130.79 (14.7–238.7)

NIHSS, National Institutes of Health Stroke Scale; DWI, diffusion-weighted imaging; PWI, perfusion-weighted imaging.

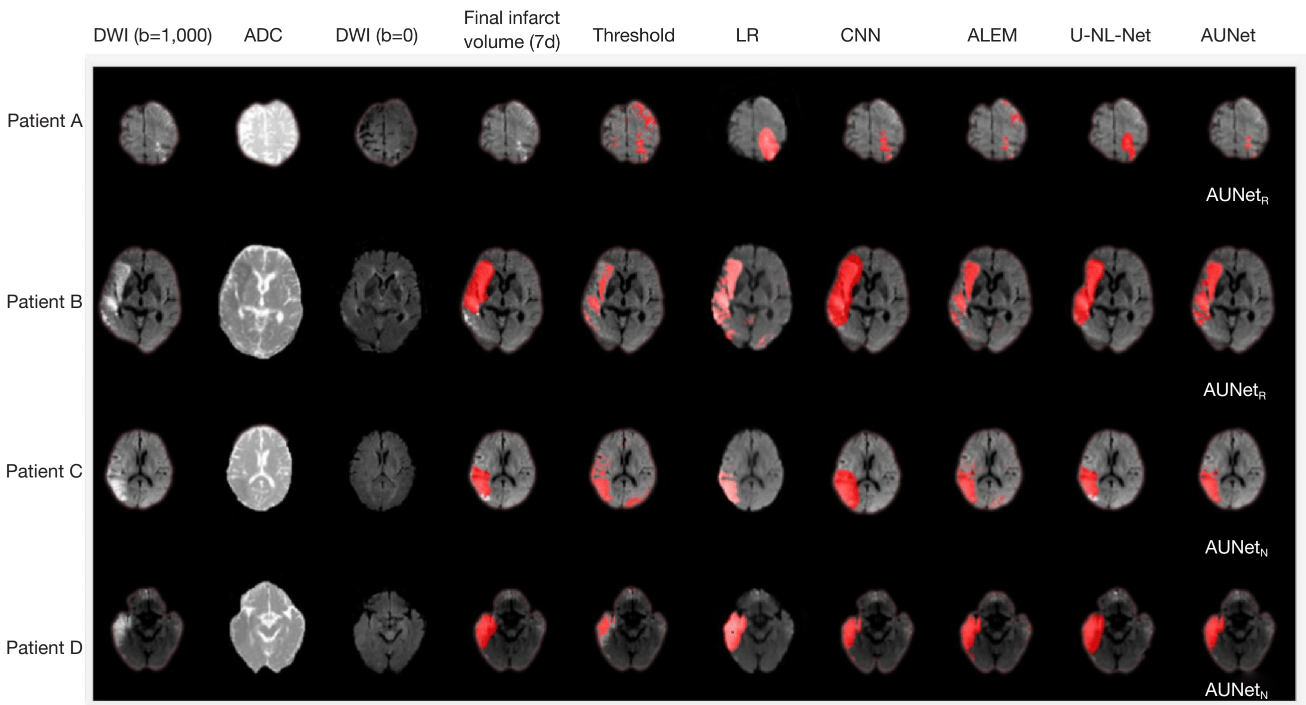

Figure 2 The final infarct volumes predicted by the different models. Patients A and B are from the recanalization group and Patients C and D are from the non-recanalization group. The final infarct volumes from 7-day data and predicted from different methods are labeled in red. For the fixed-thresholding method, the same ADC threshold was used for all patients. For other learning-based methods, the results of the recanalization model are illustrated for Patients A and B and those of the non-recanalization model are displayed for Patients C and D. Patient A is a 72-year-old woman (NIHSS =23), scanned 273 minutes after symptom onset. AUNetR shows the least overestimation of the final infarct volumes (none on this slice). Patient B is a 55-year-old man (NIHSS =13) scanned after 244 minutes. Patient C is an 80-year-old woman (NIHSS =14) scanned after 193 minutes. Patient D is a 71-year-old woman (NIHSS =4) scanned after 281 minutes. LR, logistic regression method; Threshold, fixed-thresholding method; NIHSS, National Institutes of Health Stroke Scale; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; CNN, convolutional neural network; ALEM, adaptive linear ensemble model.

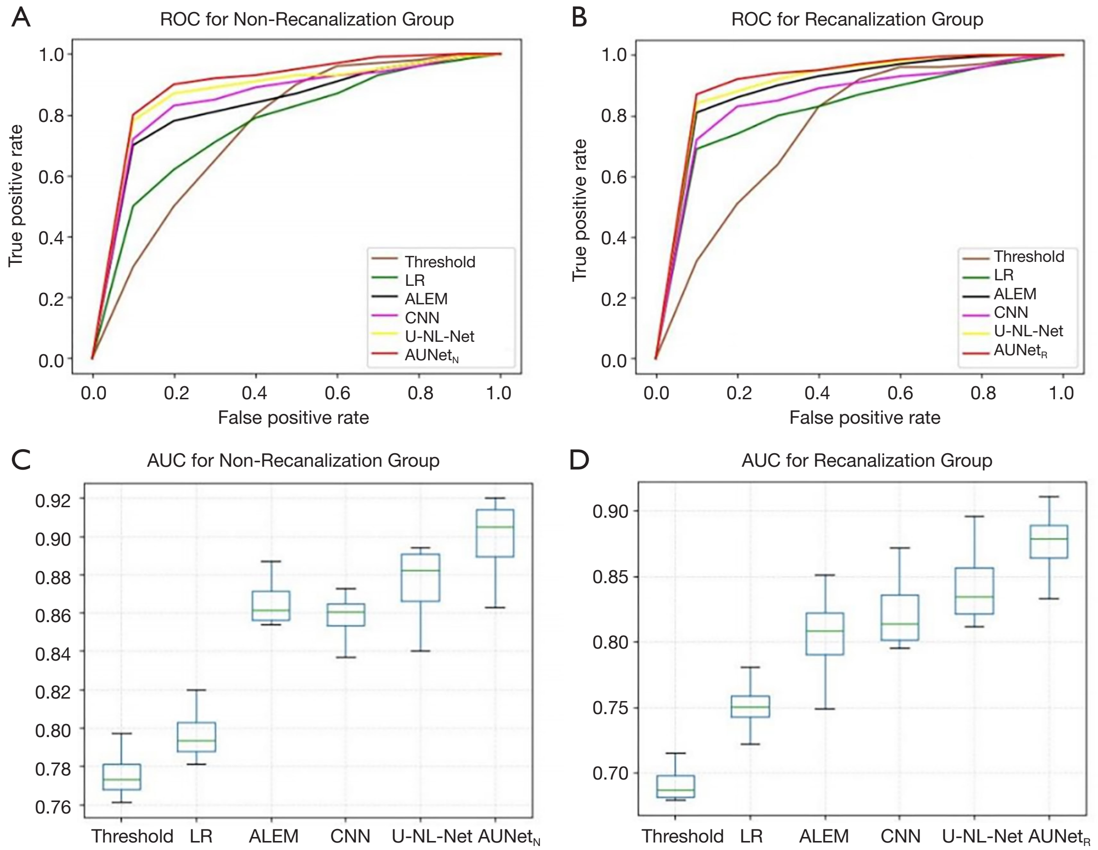

Figure 3 Receiver operating characteristic (ROC) curves (A,B) and box plots of the areas under the curve (AUCs) for the recanalization and non-recanalization datasets (C,D). LR, logistic regression method; Threshold, fixed-thresholding method; CNN, convolutional neural network; ALEM, adaptive linear ensemble model.

pixels, and the use of the machine-learning-based ALEM significantly improved the voxel-by-voxel sorting. The conventional CNN methods showed an overestimation of the final infarct volume, while the U-NL-Net showed a better performance than conventional CNN. AUNet, which uses ALEM to narrow the area of interest and U-NL-Net in negating discrete voxels, showed the best performance among all methods.

To further quantitatively compare the prediction performances of these methods, ROC and AUC results from the test subgroup in the recanalization group and non-recanalization group are separately shown in *Figure 3*. The highest AUC was obtained by AUNet (AUC =0.898±0.022, recanalization; AUC =0.875±0.036, non-recanalization: MCC =0.863±0.033, recanalization; MCC =0.851±0.025, non-recanalization), followed by U-NL-Net (AUC =0.875±0.019, P=0.02, recanalization; AUC =0.844±0.052, P=0.037, non-recanalization: MCC =0.853±0.017, recanalization; MCC =0.827±0.025, non-recanalization), ALEM (AUC =0.866±0.021, P<0.003, recanalization; AUC =0.804±0.047, P=0.004, non-recanalization: MCC =0.841±0.008, recanalization; MCC =765±0.021, non-recanalization), the CNN method (AUC =0.814±0.013, P<0.003, recanalization; AUC =0.781±0.027, P<0.003, non-recanalization: MCC =792±0.022, recanalization; MCC =0.758±0.016, non-recanalization), the LR method (AUC =0.797±0.023, P<0.003, recanalization; AUC =0.751±0.030, P<0.003, non-recanalization: MCC =0.762±0.035, recanalization; MCC =0.730±0.031, non-recanalization), the fixed-thresholding method (AUC =0.776±0.021, P<0.0001, recanalization; AUC =0.692±0.023, P<0.0001, non-recanalization: MCC =0.742±0.035, recanalization; MCC =0.671±0.024, non-recanalization), and the U-Net (AUC =0.872±0.031, P<0.003, recanalization; AUC =0.812±0.024, P<0.003, non-recanalization: MCC =0.849±0.013, recanalization; MCC =786±0.018, non-

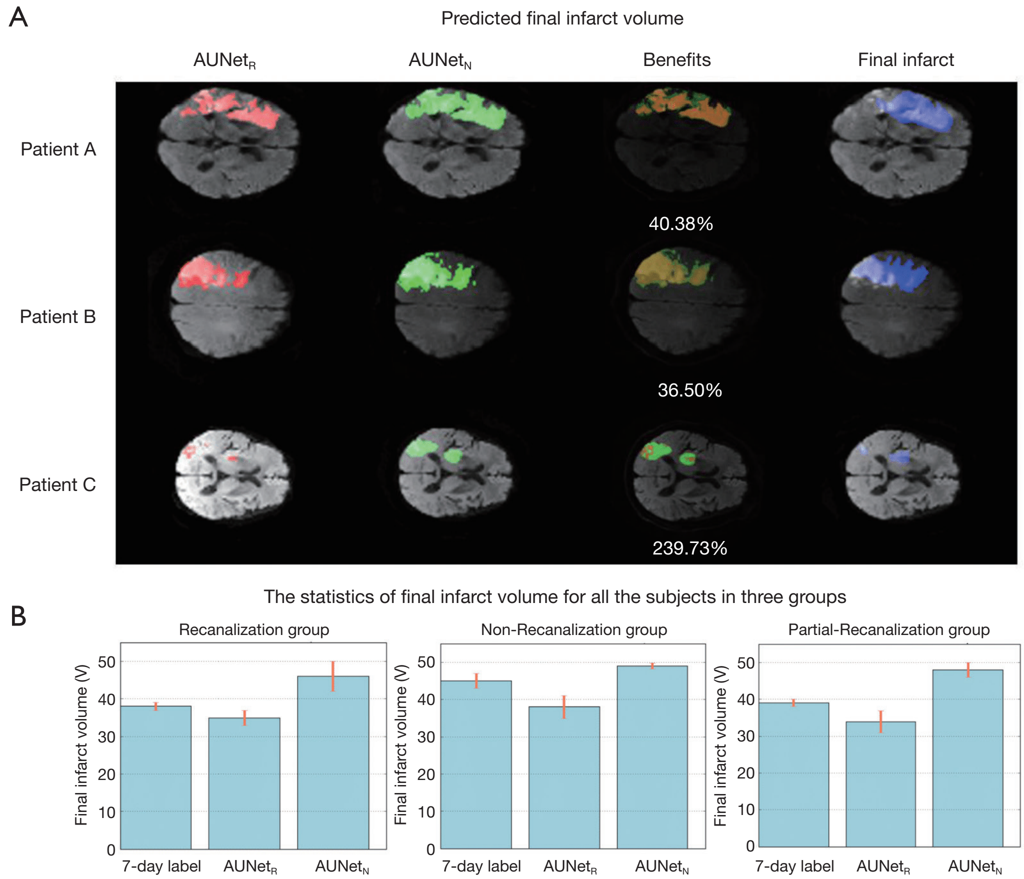

Figure 4 The analysis of predicted benefits of recanalization. (A) Examples of the final infarct volume obtained from the AUNetR and AUNetN and the benefits calculated according to Eq. [1] on three patients. The difference of the predicted infarct volumes by the two models (shown as green, overlaps are shown in yellow) are displayed in the third column. The real infarct volumes are shown in the fourth column. Patients A, B, and C are from the recanalization, non-recanalization, and partial-recanalization groups, respectively. For recanalized Patient A, AUNetN shows a clear overestimation of the infarct volume and AUNetR shows a better performance. For non-recanalized Patient B, AUNetR shows underestimation of the final infarct volume in some regions. For partially-recanalized Patient C, the final infarct volume is larger than that predicted by AUNetR but smaller than that of AUNetN. (B) The statistics of the final infarct volumes obtained from the different prediction models and the ground truth for all the subjects in the three patient groups; the bar height represents the average volume of all subjects in each group.

recanalization). The proposed voxel-based ALEM method also showed significantly higher AUCs than those obtained by other voxel-based methods (the LR method and the fixed-thresholding method). Additionally, the proposed U-NL-Net demonstrated better performance than the conventional CNN method, suggesting the NL module could significantly improve the prediction model. The further combination of ALEM and U-NL-Net, i.e., AUNet, also showed significant improvement over either U-NL-Net or ALEM alone, suggesting that ensemble learning was responsible for the better performance of AUNet.

The predicted benefits of recanalization

Figure 4A shows the predicted final infarct volumes using the recanalization model ($AUNet_R$) and non-recanalization model ($AUNet_N$) on three representative patients from each group: the recanalization group (Patient A), the non-recanalization group (Patient B), and the partial-recanalization group (Patient C), along with the recanalization benefits and the final infarct volume defined on the 7-day data. For Patient A, $AUNet_R$ showed the best prediction of the final infarct volume; the predicted therapy benefit was 40.38%. In other words, the final infarct volume will increase by 40.38% if no thrombolytic therapy is received or if there is a failure in the recanalization after thrombolytic therapy. For Patient B, $AUNet_N$ showed the best final infarct volume prediction due to recanalization failure, and the predicted benefit was 36.50%. For Patient C, the final infarct volume was smaller than that predicted by $AUNet_N$ but larger than that predicted by $AUNet_R$. Patient C would show a predicted benefit of 239.7% due to the relatively small infarct volume if full recanalization were achieved, as predicted by $AUNet_R$.

Overall, as shown in *Figure 4B*, for all subjects in the recanalization group, the final infarct volume (V =36.9 mL) was close to the predicted volume of $AUNet_R$ (V =35.8 mL), but significantly smaller than that of $AUNet_N$ (V =44.9 mL) (P<0.0001). For the non-recanalization group, the final infarct volume (V =46.2 mL) was closer to that predicted by $AUNet_N$ (V =47.3 mL) but significantly larger than that predicted by $AUNet_R$ (V =38.5 mL) (P<0.0001). The partial-recanalization group's final infarct volume (V =38.9 mL) lay between the results predicted by $AUNet_R$ and $AUNet_N$. The predicted therapy benefits for the recanalization, partial-recanalization, and non-recanalization groups were 41.94%, 39.87%, and 40.63%, respectively; there were no significant differences between groups. However, the predicted benefits showed large individual variances in all groups, from the smallest at 12.81% to the largest at 239.73%.

DWI maps contain additional information compared with ADC maps for final infarct volume prediction

While most existing methods (12,17) use only ADC maps to predict the infarct volume, we included DWI maps (including the b =0 and 1,000 s/mm^2) in our model training. To prove that the fusion into DWI maps can provide additional information for the final infarct volume prediction; we further trained the AUNet model on ADC maps only (ADC-only AUNet). *Figure 5A* shows representative examples for the AUNet and ADC-only AUNet. For most images, we found that the ADC-only AUNet overestimated the final infarct volume in both the recanalization and non-recanalization models. Based on the AUCs of the ROC (*Figure 5B,C*), the AUNet (AUC =0.898±0.022, recanalization; AUC =0.875±0.036, non-recanalization) showed better performance than the ADC-only AUNet (AUC =0.852±0.023, P<0.003, recanalization; AUC =0.830±0.037, P<0.003, non-recanalization). These results suggest that DWI images can provide fine-grained information to improve the performance of the model.

Discussion

Current guidelines for ischemic stroke patients limit the time to initiate intravenous thrombolytic therapy within 4.5 hours after stroke onset. A meta-analysis (30) showed that only about one-third of patients within the 4.5-hour time limit benefited from intravenous thrombolysis. It suggested that more detailed information may be needed to screen for more suitable thrombolytic candidates. Convenient methods for the accurate prediction of infarct volume and salvageable brain tissue are required. In this study, we demonstrated that the AUNet model, which is an ensemble learning method combining machine learning and deep learning to improve prediction accuracy, can predict the final infarct volume with full or no recanalization from baseline DWIs only. To the best of our knowledge, these are the first predictive models of the final infarct volume for different recanalization levels. The proposed AUNet algorithm, once trained, can be readily appended to automated post-processing software in a computationally inexpensive manner and can provide a convenient way to select suitable patients for thrombolytic therapy with only DWIs needed.

The advantageous performance of the proposed AUNet model is due to several factors. First, the inclusion of long-distance spatial information by adding the NL module to the U-NL-net improves the model over conventional CNN. This NL module may benefit the network by learning the infarct tissue features by comparison with normal long-distance tissue. Second, we adaptively used feature importance as weighting to combine three machine learning methods [random forest (21), extremely randomized trees (22), and XGBoost (23)], which makes the voxel-wise sorting of infarct tissue more accurate than an LR or fixed-thresholding method. Third, the final AUNet ensemble machine-learning-based voxel sorting and deep-learning-based image segmentation combine both shallow (voxel) and deep (spatial distribution) information from baseline DWIs. Fourth, we found that including the DWI raw images could significantly improve the prediction performance, suggesting that DWI raw images contain additional information not found in ADC images. Additionally, DWI raw images may have relatively higher signal-to-noise ratios than ADC images, explaining the additional fine-grained features learned from the DWI raw images.

This study has several limitations. First, we only trained

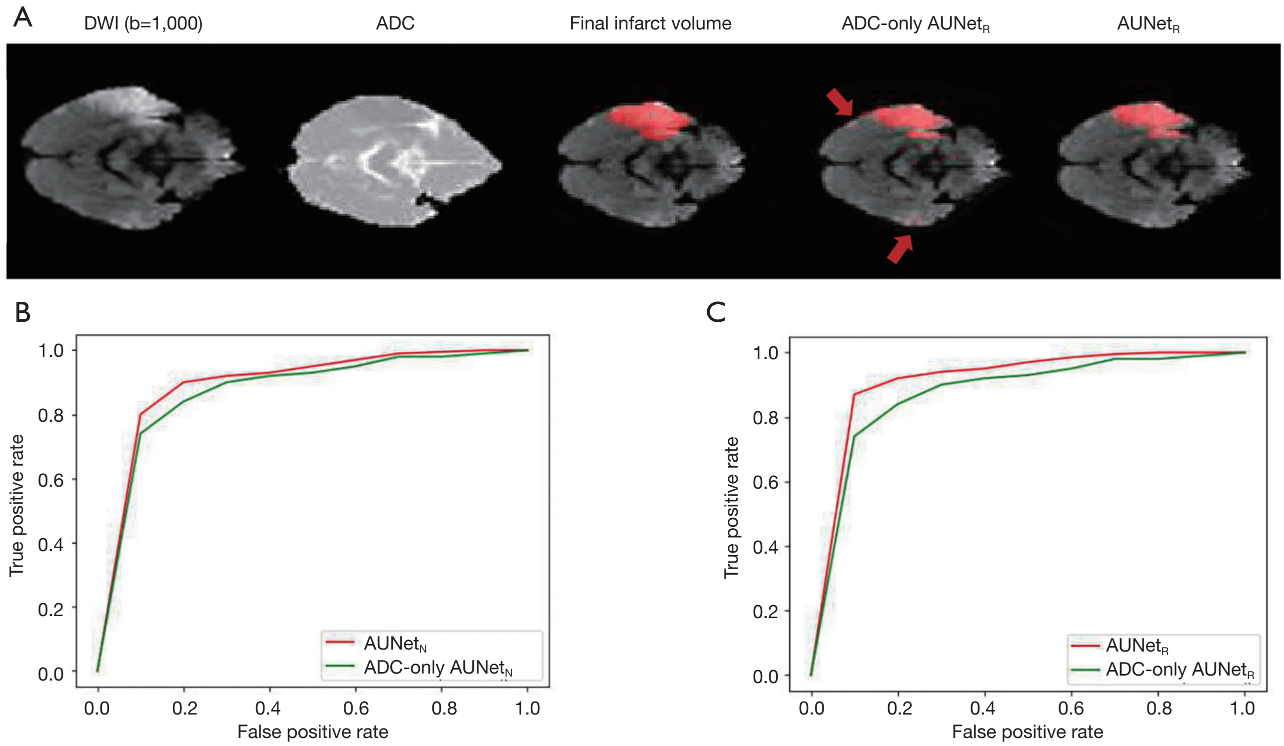

Figure 5 Verify the gain of DWI maps compared with ADC maps. (A) The predicted infarct volumes of the ADC-only AUNet and AUNet. The patient is a 74-year-old woman (NIHSS =7) from the recanalization group, scanned 164 minutes after symptom onset. The ADC-only AUNet always overestimates the infarct volume, as shown by the red arrows. (B,C) The receiver operating characteristic (ROC) curves of the non-recanalization model (B) and recanalization model (C) using the ADC-only AUNet and AUNet. NIHSS, National Institutes of Health Stroke Scale; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient.

and tested the models on a dataset from a limited number of patients, though proper data enhancement was used. However, the current AUNet model can be readily trained and tested with additional data. In the future, it would be interesting to adapt our current methods for patients beyond 4.5 hours, for whom CT perfusion or perfusion-diffusion MRI is required to receive thrombolytic therapy. Second, we only used DWI data (and ADC data derived from DWI) in the model. Whether other imaging modalities could improve the final infarct volume prediction should be further investigated. However, including more image modalities would significantly increase the scan time, which is not preferred for most patients. Also, contrast agents in perfusion MRI/CT are restricted in some specific cases, e.g., an MRI contrast agent is not allowed for patients with kidney problems. Other patient information, such as age, stroke onset time, or position of the occluded vessels, may also be included in the model to improve prediction accuracy (31). Third, CT, especially CTP, is also frequently used for acute stroke evaluation.

Further testing of the proposed methods on CT images is warranted. Fourth, the U-NL-Net network portion of the proposed method is based on a two-dimensional convolution; therefore, the slice direction's spatial information is not fully utilized. Using a 3D convolution for feature extraction operations may be considered. And lastly, to achieve personalized medicine, an interesting but more challenging direction would be to explore prediction models in partial-recanalization conditions further to allow predictions of all possible outcomes, thus creating the ability to predict each recanalization outcome for each patient.

Conclusions

The proposed AUNet model demonstrates significant advantages over currently available methods for predicting final infarct volumes. Additionally, we have established methods to predict the final infarct volumes under different recanalization conditions to evaluate thrombolytic therapy's potential benefits quantitatively. The proposed model needs to be validated in patients beyond 4.5 hours and may evolve into a quantitative framework for ischemic core diagnosis, therapeutic decision-making, and prognostic evaluation of therapeutic efficacy at an individual level.

Acknowledgments

Funding: The National Natural Science Foundation of China (NSFC) (grant no. 81873894), the Zhejiang Provincial Natural Science Foundation of China (no. LR20H180001), the Science Technology Department of Zhejiang Province (2018C04011), the National Natural Science Foundation of China (81971101), and the National Key Research and Development Program of China (2016YFC1301503).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/qims-21-33). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The human ethics committee at our center approved this study protocol. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Written informed consent was obtained from all patients or their designated proxy.

References

1. Wang Y, Li Z, Wang Y, Zhao X, Liu L, Yang X, Wang C, Gu H, Zhang F, Wang C, Xian Y, Wang DZ, Dong Q, Xu A, Zhao J. Chinese Stroke Center Alliance: a national effort to improve healthcare quality for acute stroke and transient ischaemic attack: rationale, design and preliminary findings. Stroke Vasc Neurol 2018;3:256-62.
2. Laredo C, Renú A, Tudela R, Lopez-Rueda A, Urra X, Llull L, Macías NG, Rudilosso S, Obach V, Amaro S, Chamorro Á. The accuracy of ischemic core perfusion thresholds varies according to time to recanalization in stroke patients treated with mechanical thrombectomy: A comprehensive whole-brain computed tomography perfusion study. J Cereb Blood Flow Metab 2020;40:966-77.
3. Lyden PD. Thrombolytic Therapy for Acute Ischemic Stroke: A Very Great Honor. Stroke 2019;50:2597-603.
4. Yu AY, Hill MD, Coutts SB. Should minor stroke patients be thrombolyzed? A focused review and future directions. Int J Stroke 2015;10:292-7.
5. Miller DJ, Simpson JR, Silver B. Safety of thrombolysis in acute ischemic stroke: a review of complications, risk factors, and newer technologies. Neurohospitalist 2011;1:138-47.
6. Purushotham A, Campbell BC, Straka M, Mlynash M, Olivot JM, Bammer R, Kemp SM, Albers GW, Lansberg MG. Apparent diffusion coefficient threshold for delineation of ischemic core. Int J Stroke 2015;10:348-53.
7. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. Biom J 2008;50:419-30.
8. Straka M, Albers GW, Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke. J Magn Reson Imaging 2010;32:1024-37.
9. Hacke W, Albers G, Al-Rawi Y, Bogousslavsky J, Davalos A, Eliasziw M, Fischer M, Furlan A, Kaste M, Lees KR, Soehngen M, Warach S; DIAS Study Group. The Desmoteplase in Acute Ischemic Stroke Trial (DIAS): a phase II MRI-based 9-hour window acute stroke thrombolysis trial with intravenous desmoteplase. Stroke 2005;36:66-73.
10. Davis SM, Donnan GA, Parsons MW, Levi C, Butcher KS, Peeters A, Barber PA, Bladin C, De Silva DA, Byrnes G. Effects of alteplase beyond 3 h after stroke in the Echoplanar Imaging Thrombolytic Evaluation Trial (EPITHET): a placebo-controlled randomised trial. Lancet Neurol 2008;7:299-309.
11. Albers GW, Thijs VN, Wechsler L, Kemp S, Schlaug G, Skalabrin E, Bammer R, Kakuda W, Lansberg MG, Shuaib A. Magnetic resonance imaging profiles predict clinical

response to early reperfusion: The diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE) study. Ann Neurol 2006;60:508-17.
12. Flottmann F, Broocks G, Faizy TD, Ernst M, Forkert ND, Grosser M, Thomalla G, Siemonsen S, Fiehler J, Kemmling A. CT-perfusion stroke imaging: a threshold free probabilistic approach to predict infarct volume compared to traditional ischemic thresholds. Sci Rep 2017;7:6679.
13. Lee J, Lee D, Choi JY, Shin D, Shin HG, Lee J. Artificial neural network for myelin water imaging. Magn Reson Med 2020;83:1875-83.
14. Chen Z, Chen C, Cheng Z, Jiang B, Fang K, Jin X. Selective Transfer with Reinforced Transfer Network for Partial Domain Adaptation. arXiv:1905.10756 [cs, stat]. 2020 [cited 2020 May 16]. Available online: http://arxiv.org/abs/1905.10756
15. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM 2017;60:84-90.
16. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs]. 2015 [cited 2020 May 16]. Available online: http://arxiv.org/abs/1409.1556
17. Nielsen A, Hansen MB, Tietze A, Mouridsen K. Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning. Stroke 2018;49:1394-401.
18. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF. editors. Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015. Cham: Springer International Publishing, 2015:234-41.
19. Jonsdottir KY, Østergaard L, Mouridsen K. Predicting tissue outcome from acute stroke magnetic resonance imaging: improving model performance by optimal sampling of training data. Stroke 2009;40:3006-11.
20. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal 2001;5:143-56.
21. Loh WY. Classification and regression trees. WIREs Data Mining Knowl Discov 2011;1:14-23.
22. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006;63:3-42.
23. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016:785-94.
24. Wang X, Girshick R, Gupta A, He K. Non-local Neural Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:7794-803.
25. Buades A, Coll B, Morel JM. A Non-Local Algorithm for Image Denoising. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA: IEEE, 2005:60-5.
26. Lazebnik S, Schmid C, Ponce J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06). New York, NY, USA: IEEE, 2006:2169-78.
27. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs]. 2017 [cited 2020 May 16]. Available online: http://arxiv.org/abs/1412.6980
28. Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology 1945;26:297-302.
29. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. IEEE Trans Pattern Anal Mach Intell 2020;42:318-27.
30. Emberson J, Lees KR, Lyden P, Blackwell L, Albers G, Bluhmki E, Brott T, Cohen G, Davis S, Donnan G, Grotta J, Howard G, Kaste M, Koga M, von Kummer R, Lansberg M, Lindley RI, Murray G, Olivot JM, Parsons M, Tilley B, Toni D, Toyoda K, Wahlgren N, Wardlaw J, Whiteley W, del Zoppo GJ, Baigent C, Sandercock P, Hacke W; Stroke Thrombolysis Trialists' Collaborative Group. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. Lancet 2014;384:1929-35.
31. Fabritius MP, Reidler P, Froelich MF, Rotkopf LT, Liebig T, Kellert L, Feil K, Tiedt S, Kazmierczak PM, Thierfelder KM, Puhr-Westerheide D, Kunz WG. Incremental Value of Computed Tomography Perfusion for Final Infarct Prediction in Acute Ischemic Cerebellar Stroke. J Am Heart Assoc 2019;8:e013069.

Cite this article as: Chen Z, Li Q, Li R, Zhao H, Li Z, Zhou Y, Bian R, Jin X, Lou M, Bai R. Ensemble learning accurately predicts the potential benefits of thrombolytic therapy in acute ischemic stroke. Quant Imaging Med Surg 2021;11(9):3978-3989. doi: 10.21037/qims-21-33

Supplement 1

Mean decrease impurity (MDI) refers to the average Gini coefficient attenuation of this feature through all participating nodes. Assuming that node q is split by feature s, the impurity changes before and after the split:

$$\Delta i(s,q) = i(q) - p_L i(q_L) - p_R i(q_R) \qquad [1]$$

Where i(q) represents the impurity of node q. P_L and P_R represent the probability that the sample is divided into the left and right subtrees, respectively. Based on the impurity change at a single node, the average impurity change can be calculated for the feature X_m:

$$Imp(X_m) = \frac{1}{N_T} T \sum_{q \in T} p(q) \Delta i(s_q, q) \qquad [2]$$

Where p(q) represents the proportion of samples that are classified into node q in all samples, and s_q represents the feature segmentation points used.

Calculate the MDI for all features by Eq. [2], and then sort the results from large to small. The first K features are taken as the most important K features of the model.

The ensemble model is based on the weighting of each base model, and the weighted coefficients are determined by the most important K features of each model:

$$w_i(x) = \sum_{j=1}^{K} v_{ij} f_j(x) \qquad [3]$$

$$E(x) = \sum_{i=1}^{L} w_i b_i(x) \qquad [4]$$

$w_i(x)$ is the weighting coefficient of i-th base classifier, $f_j(x)$ represents the j-th important eigenvalue of the sample x under the i-th base classifier. v_{ij} is a set of parameters to learn, which represents the output of the i-th base classifier, and E(x) is the result after ensemble. According to Eq. [3] and Eq. [4], we can solve variables by minimizing Eq. [5]:

$$\min_v \sum_{x \in X'} \sum_{i,j} (v_{ij} f_j(x) b_i(x) - y_{gt}(x))^2 \qquad [5]$$

X' is a subset of the training set used to train the ensemble model. Directly solving Eq. [5] may lead to the problem of overfitting. Therefore, an additional penalty L2 term for free variables v_{ij} is used to avoid the model being too complicated, which can be regarded as the following ridge regression problem:

$$\min_v \sum_{x \in X'} \sum_{i,j} (v_{ij} f_j(x) b_i(x) - y_{gt}(x))^2 + \lambda \sum_{ij} v_{ij}^2 \qquad [6]$$

λ is the penalty term coefficient. Solving the Eq. [6] and get:

$$v_{opt} = (A^T A + \lambda I)^{-1} y_{gt} \qquad [7]$$

Where $A_{ij} = f_j(x) g_i(x)$.

 http://dx.doi.org/10.21037/qims-21-33

Supplement 2 (Table S1)

Table S1 U-Net (DWI + ADC) architecture used as the baseline

Level	Layer name	Channel	Kernel	Step	Output dimension	Parameters (M)	Computation (GFlops)
Input		–	–	–	224^2×1	–	–
Encode1	Convolution	64	3×3	1	224^2×64	0.0006	0.0321
	BN	64	–	–	224^2×64	0.0001	0.0128
	Dropout	–	–	–	224^2×64	–	–
	Convolution	64	3×3	1	224^2×64	0.0369	1.8529
	BN	64	–	–	224^2×64	0.0001	0.0128
	Dropout	–	–	–	224^2×64	–	–
	MaxPool	–	2×2	2	112^2×64	–	–
Encode2	Convolution	128	3×3	1	112^2×128	0.0739	0.9264
	BN	128	–	–	112^2×128	0.0003	0.0064
	Dropout	–	–	–	112^2×128	–	–
	Convolution	128	3×3	1	112^2×128	0.1476	1.8513
	BN	128	–	–	112^2×128	0.0003	0.0064
	Dropout	–	–	–	112^2×128	–	–
	MaxPool	–	2×2	2	56^2×128	–	–
Encode3	Convolution	256	3×3	1	56^2×256	0.2952	0.9256
	BN	256	–	–	56^2×256	0.0005	0.0032
	Dropout	–	–	–	56^2×256	–	–
	Convolution	256	3×3	1	56^2×256	0.5901	1.8505
	BN	256	–	–	56^2×256	0.0005	0.0032
	Dropout	–	–	–	56^2×256	–	–
	MaxPool	–	2×2	2	28^2×256	–	–
Encode4	Convolution	256	3×3	1	28^2×256	0.5901	0.4626
	BN	256	–	–	28^2×256	0.0005	0.0008
	Dropout	–	–	–	28^2×256	–	–
	Convolution	256	3×3	1	28^2×256	0.5901	0.4626
	BN	256	–	–	28^2×256	0.0005	0.0008
	Dropout	–	–	–	28^2×256	–	–
Docode1	Up&Concat	–	–	–	56^2×512	–	–
	Convolution	256	3×3	1	56^2×256	1.1799	3.7002
	BN	256	–	–	56^2×256	0.0005	0.0032
	Dropout	–	–	–	56^2×256	–	–
	Convolution	256	3×3	1	56^2×256	0.5901	1.8505

Table S1 (*continued*)

 http://dx.doi.org/10.21037/qims-21-33

Table S1 (*continued*)

Level	Layer name	Channel	Kernel	Step	Output dimension	Parameters (M)	Computation (GFlops)
	BN	256	–	–	56^2×256	0.0005	0.0032
	Dropout	–	–	–	56^2×256	–	–
Decode2	Up&Concat	–	–	–	112^2×384	–	–
	Convolution	128	3×3	1	112^2×128	0.4425	5.5507
	BN	128	–	–	112^2×128	0.0003	0.0064
	Dropout	–	–	–	112^2×128	–	–
	Convolution	128	3×3	1	112^2×128	0.1476	1.8513
	BN	128	–	–	112^2×128	0.0003	0.0064
	Dropout	–	–	–	112^2×128	–	–
Decode3	Up&Concat	–	–	–	224^2×192	–	–
	Convolution	64	3×3	1	224^2×64	0.1107	5.5523
	BN	64	–	–	224^2×64	0.0001	0.0128
	Dropout	–	–	–	224^2×64	–	–
	Convolution	64	3×3	1	224^2×64	0.0369	1.8529
	BN	64	–	–	224^2×64	0.0001	0.0128
	Dropout	–	–	–	224^2×64	–	–
Output	Convolution	1	3×3	1	224^2×1	0.0006	0.0290
Sum						4.84	28.84

DWI indicates diffusion weighted imaging; ADC, apparent diffusion coefficient.

http://dx.doi.org/10.21037/qims-21-33

Supplement 3 (Figure S1)

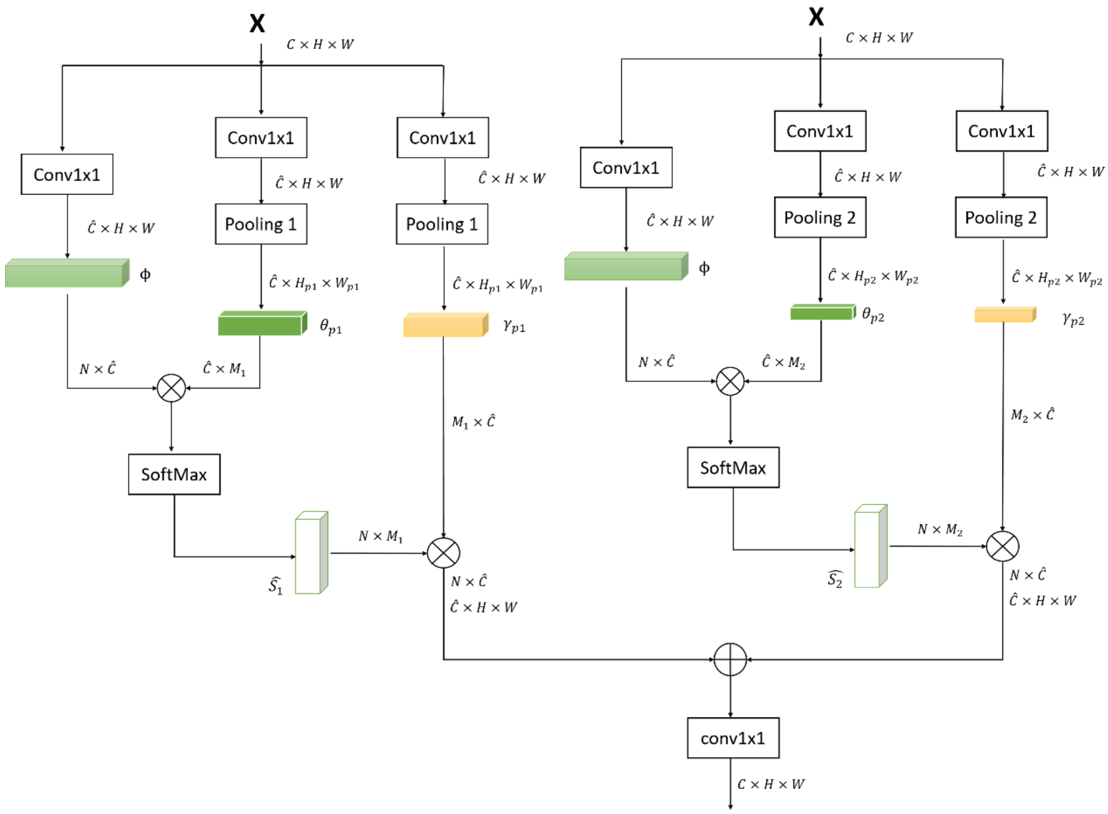

Figure S1 Developed non-local module architecture.

 http://dx.doi.org/10.21037/qims-21-33