NeuroImage: Clinical 31 (2021) 102744

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

An automatic machine learning approach for ischemic stroke onset time identification based on DWI and FLAIR imaging

Haichen Zhu [a,1], Liang Jiang [b,1], Hong Zhang [c], Limin Luo [a], Yang Chen [a,d,*], Yuchen Chen [b,*]

[a] Lab of Image Science and Technology, Key Laboratory of Computer Network and Information Integration (Ministry of Education), School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[b] Department of Radiology, Nanjing First Hospital, Nanjing Medical University, Nanjing 210006, China
[c] Department of Radiology, Affiliated Jiangning Hospital of Nanjing Medical University, Nanjing 210000, China
[d] School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China

ARTICLE INFO

Keywords:
Machine learning
Segmentation and classification
Time since stroke
Magnetic resonance imaging

ABSTRACT

Current thrombolysis for acute ischemic stroke (AIS) treatment strictly relies on the time since stroke (TSS) less than 4.5 h. However, some patients are excluded from thrombolytic treatment because of the unknown TSS. The diffusion-weighted imaging (DWI) and fluid-attenuated inversion recovery (FLAIR) mismatch can simply identify TSS since lesion intensities are not identical at different onset time. In this paper, we propose an automatic machine learning method to classify the TSS less than or more than 4.5 h. First, we develop a cross-modal convolutional neural network to accurately segment the stroke lesions from DWI and FLAIR images. Second, the features are extracted from DWI and FLAIR according to the segmentation regions of interest (ROI). Finally, the features are fed to machine learning models to identify TSS. In DWI and FLAIR ROI segmentation, the networks obtain high Dice coefficients with 0.803 and 0.647. The classification test results show that our model achieves an accuracy of 0.805, with a sensitivity of 0.769 and a specificity of 0.840. Our approach outperforms human reading DWI-FLAIR mismatch model, illustrating the potential for automatic and fast TSS identification.

1. Introduction

Stroke ranks as one of the leading causes of death in the world, with acute ischemic stroke (AIS) being the most common subtype (Wang, 2020; Zhu et al., 2020). According to the AIS treatment guidelines, recombinant tissue plasminogen activator (rtPA) thrombolysis is strictly limited to time since stroke (TSS) less than 4.5 h (Furie and Jayaraman, 2018). Thrombolysis on TSS more than 4.5 h will increase the risk of hemorrhage significantly (Campbell et al., 2019). Hence, some patients with unknown TSS are excluded from thrombolysis although they are probably within the rtPA time window (Furie and Jayaraman, 2018; Jauch et al., 2013).

A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association (AHA/ASA) recommends that DWI-positive FLAIR-negative lesions (DWI-FLAIR mismatch) can be useful for selecting patients with unknown TSS who can benefit from rtPA administration within 4.5 h of stroke symptom recognition (Powers et al., 2019). The DWI-FLAIR mismatch model is constructed based on the fact that high intensity appears immediately in DWI at AIS onset whereas it takes 3-4 h for AIS to emerge in FLAIR (Ebinger et al., 2010). Fig. 1 illustrates two typical DWI-FLAIR mismatch examples of different stroke onset time. Some researchers have investigated the DWI-FLAIR mismatch in TSS identification. However, human readings of DWI-FLAIR mismatch showed the low-to-moderate accuracy in these studies (Thomalla et al., 2009; Aoki et al., 2010; Kim et al., 2014; Lee et al., 2020). These evaluation studies reported the mismatch model could achieve a specificity no more than 0.80 and a sensitivity no more than 0.6. Some patients who could benefit from thrombolysis may be misclassified since the mismatch model is too stringent. Previous work calculated the lesion water uptake from CT and set the water uptake threshold to classify patients within or beyond 4.5 h (Broocks et al., 2020), yet the research of TSS classification on CT is still limited. In

* Corresponding authors at: Lab of Image Science and Technology, Key Laboratory of Computer Network and Information Integration (Ministry of Education), School of Computer Science and Engineering, Southeast University, Nanjing 210096, China (Yang Chen); Department of Radiology, Nanjing First Hospital, Nanjing Medical University, Nanjing 210006, China (Yuchen Chen).
E-mail addresses: chenyang.list@seu.edu.cn (Y. Chen), chenyuchen1989@126.com (Y. Chen).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.nicl.2021.102744
Received 3 April 2021; Received in revised form 22 June 2021; Accepted 23 June 2021
Available online 3 July 2021

order to mine more latent features, machine learning (ML) approaches have been applied for stroke imaging analysis (e.g. brain tumor segmentation and classification (Tiwari et al., 2020), stroke tissue outcome prediction (Nielsen et al., 2018)). Lee et al. (Lee et al., 2020) extracted features (e.g. intensity, gradient, gray level co-occurrence matrices) from apparent diffusion coefficient (ADC) map and FLAIR, and constructed three ML classifiers to distinguish the patients within or after 4.5 h of onset. Ho et al. (King Chung et al., 2019) used four-modal imaging to generate features, i.e. DWI, ADC, FLAIR, and perfusion weighted imaging (PWI). Besides the area, volume, sphericity and some statistical features, 384 deep features were also generated by deep autoencoder. Then, the features were fed into ML models for TSS classification. Although these approaches have outperformed human-reading mismatch, they still need much effort for ROI generation and feature extraction. In this paper, we developed an automatic algorithm for TSS classification and evaluated this method on a large dataset which is collected from two stroke centers. The main contributions of our work can be summarized as follows:

(1) We proposed an effective segmentation network to segment infarct lesions in DWI and FLAIR. In FLAIR segmentation, the cross-modal network employed DWI prior information to obtain accurate results. The automatic segmentation network can keep radiologists from labeling regions of interest (ROIs) with extra time.
(2) Only two-modal MR (DWI and FLAIR) images were used for TSS classification and achieved comparable results. Our algorithm may be more generalizable than the previous study since PWI was not required.
(3) We constructed five classifiers (i.e., logistic regression (LR), support vector machine (SVM), random forest (RF), gradient boosted decision tree (GBDT), extra trees (ET)). The final result was decided by the voting of five classifiers for robust classification.

2. Materials

2.1. Dataset

A dataset of 268 anonymized AIS patients from January 2016 to December 2020 was collected from two stroke centers in China for retrospective evaluation. 180 cases are from the Nanjing First Hospital, Nanjing Medical University and 88 cases are from the Affiliated Jiangning Hospital of Nanjing Medical University. Each AIS patient has recorded time of stroke symptom onset and time of MR imaging. These patients were categorized into two classes according to TSS: negative ($\leqslant 4.5\ h$) and positive (>4.5 h). The characteristics of the patient cohort are shown in Table 1. All scans were performed using 3-Tesla echo planar scanner of Philips. The pixel dimension of DWI varies from $0.893 \times 0.893 \times 6.6$ to $1.198 \times 1.198 \times 7.3\ mm^3$. The pixel dimension of FLAIR varies from $0.411 \times 0.411 \times 7$ to $0.599 \times 0.599 \times 7.3\ mm^3$. DWI and FLAIR images of each patient have 18 slices to cover the cerebrum from top to bottom.

Table 1
AIS patient characteristics.

	Patients (n = 268)
Female	90
Male	178
Age	67.7 ± 13.2
Time since stroke (hour)	10.26 ± 18.11
NIHSS on admission	9.65 ± 7.01
Stroke location: left	123
Stroke location: right	145
Classification label (cases)	$\leqslant 4.5\ h$ (173); >4.5 h (95)

2.2. Image preprocessing

For each patient, the DWI images were first registered to FLAIR images using Elastix rigid registration tool in 3D Slicer. Secondly, the FLAIR and registered DWI images were respaced to $1 \times 1 \times N\ mm^3$. The images were not respaced in the z-dimension. Afterwards, we cropped each DWI and FLAIR volume to the size of $224 \times 224 \times 18$. Finally, the stroke lesions were drawn manually by one radiologist with 6-year experience and then checked by another radiologist with 15-year experience in the sagittal-sectional slices. We divided the 268-patient dataset into training, validation, test sets by a ratio of 3:1:1.

3. Methods

For DWI-FLAIR mismatch analysis, stroke lesion ROIs should be first delineated. However, accurate labeling will take a lot of manpower. Considering the deep convolutional neural networks have been widely utilized in MR image segmentation (Chen et al., 2017; Zhang et al., 2018; Pinto et al., 2018), we propose an EfficientNet-B0 based U-Net to segment ROIs from DWI and FLAIR. A multi-scale atrous convolution (MSAC) block was first introduced for fine segmentation. Then, ROI features of DWI and FLAIR were generated using PyRadiomics. Finally, we constructed five ML models and identified TSS based on five-model voting. The whole process of our method is shown in Fig. 2. All the experiments were carried out on our PC (Intel Core i5 CPU, 1080Ti 11 GB GPU, 16 GB RAM).

3.1. Stroke Lesion Segmentation

U-Net has shown great potential in medical image segmentation (Ronneberger et al., 2015). U-Net and its variants have achieved remarkable results in some medical challenges, such as Kidney Tumor Segmentation (KiTS) Challenge (Isensee and Maier-Hein, 2019), Brain Tumor Segmentation (BraTS) Challenge (Zeineldin et al., 2020), and so on (Yin et al., 2019; Ge et al., 2019). The encoder of vanilla U-Net has the limitation that the depth of the consecutive 3×3 convolutional layers should be increased to capture comprehensive features. This will bring in too many parameters which need large memory. To balance the segmentation accuracy and memory use, we proposed to replace the 3×3 convolutional encoder with EfficientNet-B0 (Tan and Le, 2019). EfficientNet-B0 is mainly composed of 7 mobile inverted bottleneck

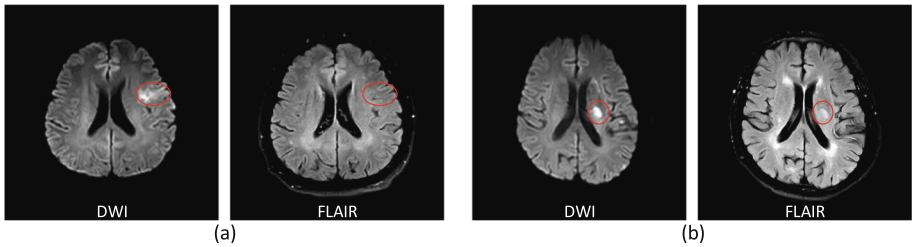

Fig. 1. Two examples of DWI-FLAIR mismatch (TSS = 2 h) and no DWI-FLAIR mismatch (TSS = 6 h).

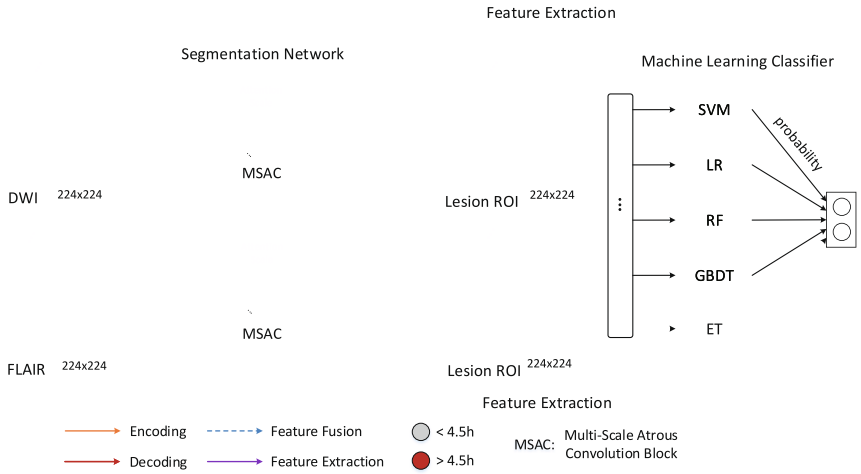

Fig. 2. The workflow of automatic TSS identification method. SVM: support vector machine, LR: logistic regression, RF: random forest, GBDT: gradient boosted decision tree, ET: extra trees.

(MBConv) blocks and 2 convolutional layers. The parameter amount of EfficientNet-B0 is only about one third of that of the vanilla U-Net's encoder (5.3 million vs. 14.1 million). Compared with DWI, lesion ROIs in FLAIR have lower contrast and are more easily interfered by other hyperintensities. These barriers will impede the pace of accurate FLAIR ROI segmentation. We constructed a cross-modal segmentation network for FLAIR segmentation, which is shown in Fig. 3. The encoding features of DWI were fused with the features of FLAIR level by level. DWI can provide the lesion location prior information which supervises FLAIR segmentation. To further concentrate on discriminative regions, a multi-scale atrous convolution block combined with attention mechanism (MSAC) was first proposed (Fig. 4). The MSAC block had four parallel convolutional paths to generate multi-scale features. Receptive fields of multiple sizes were realized by setting the dilation rates of three paths as 2,4,8. The attention scale was computed from the previous layer and then applied to multi-scale convolutional operations.

The segmentation models were trained by minimizing the proposed loss function using Adam optimization (Kingma and Ba, 2014). The loss function consisted of weighted cross entropy (WCE) and Dice loss, which is illustrated in Eq. (1). x_{ci} is the i-th pixel prediction probability of c-th class, y_{ci} is the ground truth label value, w_c is the weight for c-th class. $\varphi(\theta)$ computes the L2-norm loss of the model parameters to prevent

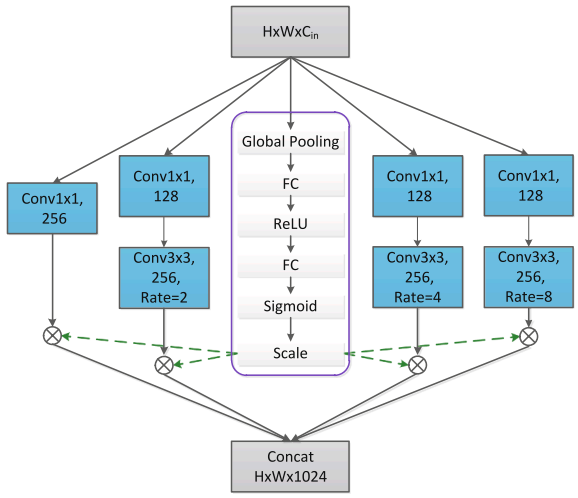

Fig. 4. The proposed multi-scale atrous convolution (MSAC) block. Attention mechanism is contained in the purple box.

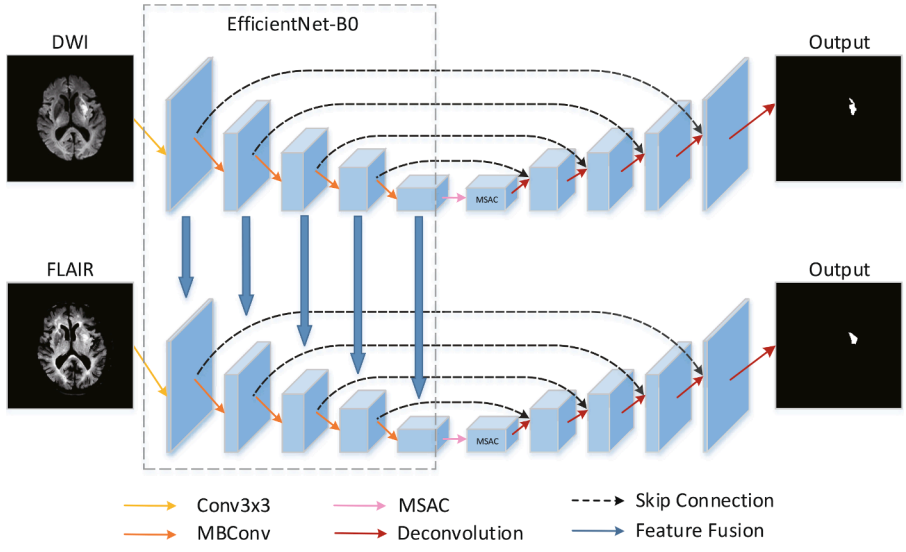

Fig. 3. The architecture of cross-modal segmentation network. The input and output size are 224 × 224.

overfitting. θ denotes the trainable parameters of model. α is L2 loss weight and was set as 1e-6. X is predicting map, Y is ground truth map. We utilized data augmentation, i.e., random flip, rotation, scaling, to prevent overfitting during the training period. The initial learning rate was set to 1e-4 and then decayed by multiplying 0.98 in each epoch.

$$\begin{aligned} Loss &= L_{WCE} + L_{Dice} \\ L_{WCE} &= -\sum_{i=1}^{n} w_c[y_{ci}\cdot\log x_{ci} + (1-y_{ci})\cdot\log(1-x_{ci})] + \alpha\cdot\varphi(\theta) \\ L_{Dice} &= 1 - \frac{2|X\cap Y|+1}{|X|+|Y|+1} \end{aligned} \tag{1}$$

3.2. Feature extraction

Radiomics is an important method in medical image analysis, which digitally describes the images containing lesion information with high-dimensional morphological features (Shi et al., 2019). Radiomics has been widely employed in tumor feature generalization (Aerts et al., 2014; Coroller et al., 2015; Lambin et al., 2017). In this work, we proposed to extract the stroke infarct lesion features automatically with PyRadiomics (Van Griethuysen et al., 2017). PyRadiomics is an open-source radiomics package in Python, which can be applied on both two and three-dimensional images. We extracted 107 features from DWI ROIs and 107 features from FLAIR ROIs. The 107 features can be divided into seven categories, i.e. first-order statistics features, shape-based features, gray level co-occurrence matrix features, gray level run length matrix features, gray level size zone matrix features, gray level dependence matrix features, and neighbouring gray tone difference matrix features. For training data and test data, we generalized features using radiologist's labeling and segmentation ROI respectively. The features were then normalized to zero mean and unit variance. Features of DWI and FLAIR were concatenated as 214-dimension for TSS classification.

3.3. TSS classification

Five machine learning classifiers were constructed for TSS classification (TSS $\leqslant 4.5h$ or TSS $>$ 4.5 h). The five models were logistic regression (LR), support vector machine (SVM), random forest (RF), gradient boosted decision tree (GBDT), extra trees (ET). LR is a statistical model that uses a logistic function to model a binary dependent variable (Bewick et al., 2005). SVM can learn a hyperplane from training examples for classification (Cortes and Vapnik, 1995). RF is an ensemble learning method which constructs a multitude of decision trees and averages the probabilities of different trees in classification tasks (Breiman, 2001). GBDT also constructs different trees but optimizes the model by using derivatives (Friedman, 2001). ET is similar to RF but has a stronger generalization ability than RF (Geurts et al., 2006). Considering the small amount of data, the deep neural network was not developed in this study. To improve the classification robustness, we integrated the five models by voting to obtain the final prediction.

4. Results

4.1. DWI segmentation

The vanilla U-Net was utilized as a baseline model. We trained the proposed EfficientNet-B0 based U-Net combined with MSAC block, abbreviated as B0-UNet w/ MSAC, to segment the infarct lesions on DWI. The hyper-parameters were chosen according to the segmentation results on DWI validation set. Finally, the pre-trained model was evaluated on DWI test set. We also used High-Resolution Network (HRNet) as a comparison. HRNet is one of the state-of-the-art methods on Cityspaces, COCO, ImageNet datasets, and achieved impressive results in segmentation, classification, detection tasks (Wang et al., 2020). Table 2 presents the evaluation metrics of different models. The proposed method obtained the best sensitivity (0.791), specificity (0.999), and Dice coefficient (0.803). B0-Unet acquired the best specificity (0.999) and precision (0.862). The B0-Unet made use of much less parameters than U-Net (12.3 million vs. 28.9 million) but achieved close dice to U-Net (0.785 vs. 0.795). HRNet outperformed U-Net in terms of sensitivity and specificity. The specificities were all greater than 0.995 since most of the brain images were non-infarct pixels.

Table 2
Evaluation results of DWI segmentation. Best results are marked in bold.

	Sensitivity	Specificity	Precision	Dice
HRNet	0.783	0.998	0.792	0.788
U-Net	0.772	0.995	0.815	0.795
B0-UNet	0.721	0.999	**0.862**	0.785
B0-UNet w/ MSAC	**0.791**	**0.999**	0.817	**0.803**

Fig. 5 demonstrates several sample results that visually compare our method with the other three methods. True positive pixels were denoted in red, false positive pixels were given in green, and false negative pixels were in blue. The results of B0-UNet w/ MSAC show less missing detection and false detection pixels. Compared with the three models, the segmentation precision of infarct boundary and isolated little lesions can be significantly improved by employing MSAC, as shown in the last column.

4.2. FLAIR segmentation

Considering the difficulties in FLAIR infarct lesion segmentation, we trained a cross-modal network based on B0-UNet w/ MSAC, abbreviated as CM_B0-UNet w/ MSAC. The corresponding DWI images were input for training synchronously as supplementary information. The hyper-parameters were also chosen according to the results on FLAIR validation set. Afterwards, the pre-trained model was evaluated on FLAIR test set. We also built a cross-modal HRNet (CM_HRNet) and a cross-modal U-Net (CM_U-Net) for comparison. Table 3 gives the evaluation results on FLAIR images. The one-modal networks (U-Net and B0-UNet) using FLAIR only achieved low metrics, with Dice both less than 0.3. The cross-modal B0-UNet w/ MSAC achieved the best sensitivity (0.561), specificity (0.999), precision (0.763) and Dice coefficient (0.647). By introducing the MSAC block to CM_B0-UNet, the Dice coefficient and other metrics can be improved around by 0.1. Some cases of FLAIR segmentation using cross-modal networks are exhibited in Fig. 6 for visual inspection. CM_B0-UNet w/ MSAC obtained comparatively accurate segmentation results in most cases. CM_B0-UNet did not detect the lesion in case 3 and CM_HRNet did not detect the lesion in case 4. In case 3, the isolated little lesion was also undetected by CM_U-Net.

4.3. TSS classification

After segmenting the infarct lesions, we extracted the features from DWI and FLAIR with PyRadiomics. Afterwards, we trained the LR, SVM, RF, GBDT, ET using the features of the training set. The hyper-parameters of each method, e.g. iterations and tree depth, were optimized on the validation set. At last, the methods were evaluated on test set. The feature extraction ROI of training cases were delineated by radiologists and the ROIs of validation and test cases were model segmentation results. The receiver operating characteristic (ROC) curves of five classifiers are depicted in Fig. 7.

The ET model showed the highest area under the curve (AUC) while LR showed the lowest AUC (0.824 vs. 0.765). The radiologist performance using DWI-FLAIR mismatch model is also marked in Fig. 7. All ROCs of the five classifiers can cover the red cross, indicating the classifiers had higher sensitivities than human-derived DWI-FLAIR mismatch when having the same specificities. The sensitivity, specificity and accuracy of human-derived DWI-FLAIR mismatch were 0.817,

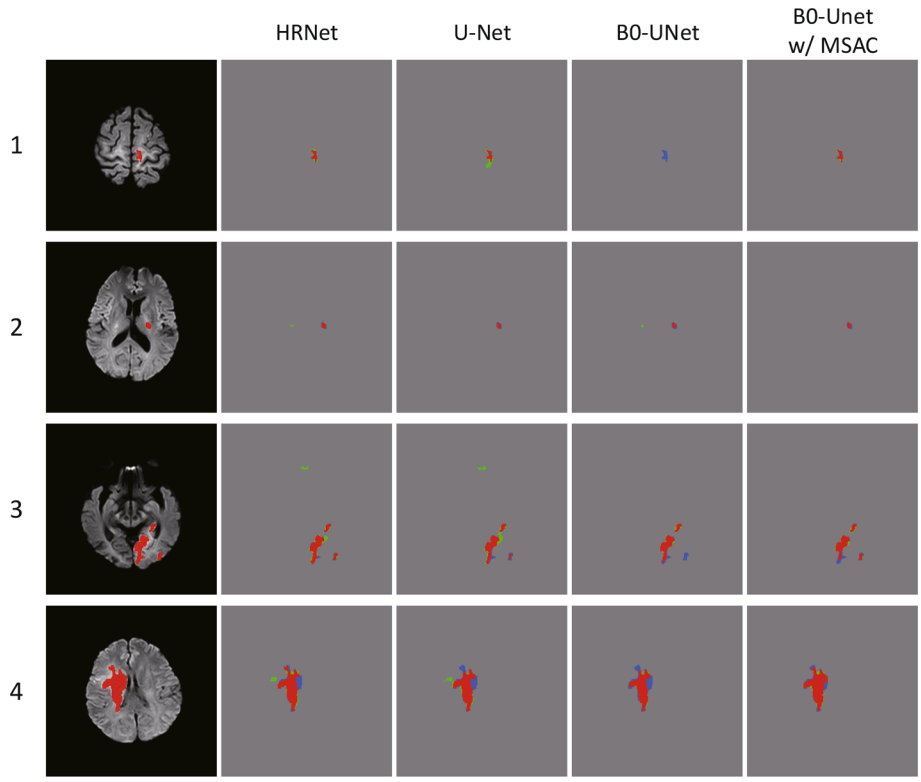

Fig. 5. Visualization of the prediction ischemic core by using the proposed method and other three methods in DWI. True positive pixels are denoted in red, false positive pixels in green and false negative pixels in blue.

Table 3
Evaluation results of FLAIR segmentation. Best results are marked in bold.

	Sensitivity	Specificity	Precision	Dice
U-Net	–	–	–	0.286
B0-UNet	–	–	–	0.299
CM_HRNet	0.494	0.999	0.754	0.597
CM_U-Net	0.489	0.999	0.665	0.564
CM_B0-UNet	0.462	0.999	0.674	0.548
CM_B0-UNet w/ MSAC	**0.561**	**0.999**	**0.763**	**0.647**

0.590, 0.704 respectively. The metrics of different classifiers are listed in Table 4.

The GBDT, RF and ET had relatively low sensitivity (< 0.7) but high specificity (> 0.8), while LR and SVM had relatively low specificity (< 0.8) but high sensitivity (> 0.7). All the accuracies of the five separate models were under 0.77. However, the voting of five models achieved the best specificity (0.840) and accuracy (0.805), which are evidently higher than those of human-derived DWI-FLAIR mismatch. The sensitivity of voting was slightly below human-derived DWI-FALIR mismatch (0.769 vs. 0.817). Table 5 reports the AUCs of the five classifiers using DWI features only, FLAIR features only, and DWI+FLAIR features for TSS identification. For most of the classifiers, using only FLAIR image had the lowest AUCs while using FLAIR + DWI had the best AUCs. We also noticed that the SVM AUC of DWI was higher than DWI+FLAIR to a small extent. This is probably because too many dimensions of image feature led SVM to locally optimal during training. The AUCs of five models using segmentation ROI DWI+FLAIR features in the test set were very close to the AUCs using man-made labeling features, all around 0.8. The sensitivities, specificities, and accuracies of segmentation ROI DWI + FLAIR features were close to those of radiologist's labeling features using the voting of five classifiers, as shown in Fig. 8.

5. Discussion

Determining the unknown TSS from medical imaging is a challenging but meaningful task in acute stroke treatment. As a common method in TSS identification, DWI-FLAIR mismatch was showed safe in rtPA selection (Schwamm et al., 2018). To reduce the radiologist's workload and improve the accuracy of TSS classification, we proposed an automatic TSS identification method using machine learning. The method is composed of 3 steps. First, using convolutional networks to segment stroke lesions from DWI and FLAIR. Second, extracting morphological features of segmentation ROIs in DWI and FLAIR. Third, using the voting of five classifiers to identify TSS. Our method achieved relatively high sensitivity (0.769) and specificity (0.840) in our dataset TSS identification, which are obviously superior to the results of DWI-FLAIR mismatch reported in (Thomalla et al., 2009; Aoki et al., 2010; Kim et al., 2014; Lee et al., 2020).

FLAIR usually needs more time to present infarct lesion hyperintensity than DWI. Most lesions of cases $\leqslant 4.5h$ do not appear hyperintensity in FLAIR. The characteristic will make the lesion ROI and the other tissues more imbalanced. Although using CE loss combined with Dice loss can ease the problem of class imbalance to some extent, the segmentation results using FLAIR only are still unsatisfactory, with Dice less than 0.3. Besides, more factors, e.g. brain tumor and cerebrospinal fluid, are highlighted in FLAIR. It will also make the FLAIR segmentation more difficult. Infarct lesion has high contrast in DWI. The Dice of DWI infarct lesion segmentation can be up to 0.8. Hence, the cross-modal network can provide lesion location information from DWI for FLAIR segmentation. In Fig. 9, case 1 typically presents the single modal FLAIR segmentation failure due to class imbalances, and case 2 exhibits the segmentation failure because of low lesion contrast. However, the lesions in DWI can be easily identified as shown in the second column. Introducing DWI features into the cross-modal networks, the FLAIR segmentation can be more accurate (Column 4 vs. Column 5). We also recognized that the lesion ROIs drawn by one radiologist and then

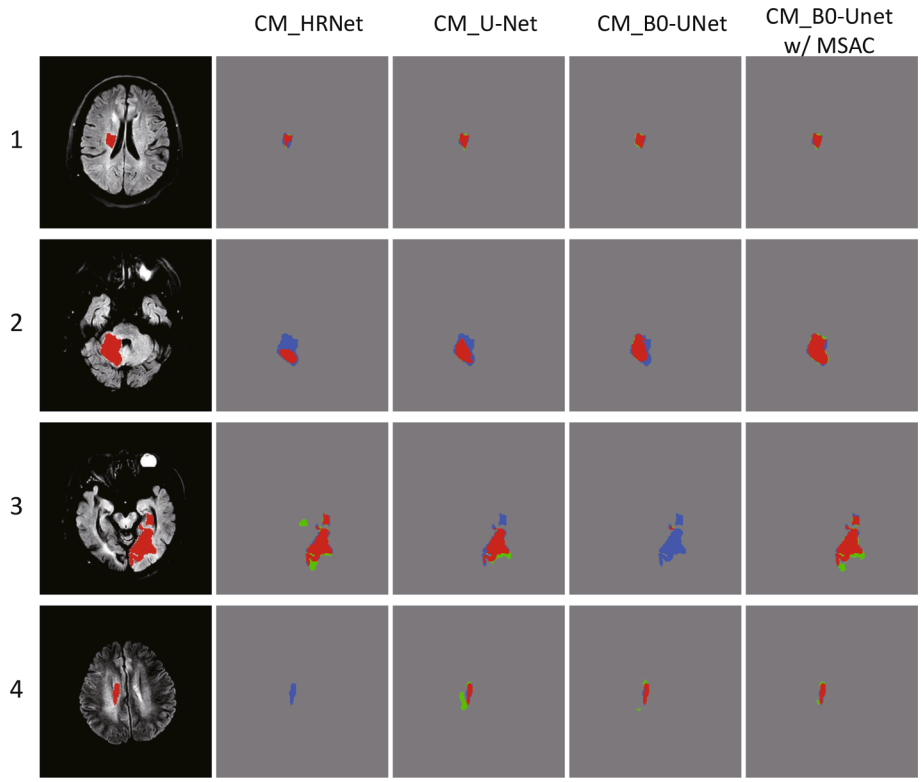

Fig. 6. Visualization of the prediction ischemic core in FLAIR by using cross-modal segmentation networks. True positive pixels are denoted in red, false positive pixels in green and false negative pixels in blue.

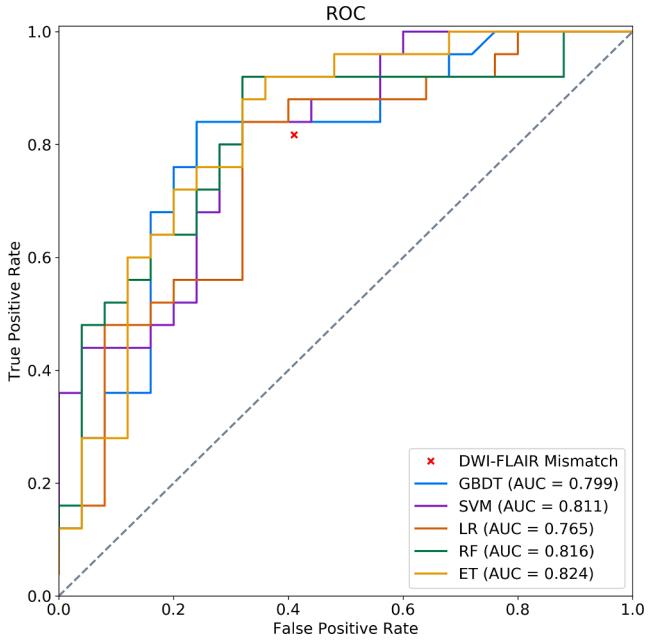

Fig. 7. The ROCs of five classifiers. The red cross marker indicates the false positive rate and true positive rate of DWI-FLAIR mismatch model by radiologists.

checked by another one may be not accurate enough because the ground truth is essentially determined by one person. The shortcoming could influence the model segmentation accuracy.

As indicated in Fig. 7, the machine learning models all outperformed DWI-FLAIR mismatch. Overall, the classification accuracy has been greatly improved by applying our method (Table 4). We also noticed

Table 4
The sensitivities, specificities and accuracies of human-reading DWI-FLAIR mismatch and different models using segmentation ROI DWI-FLAIR features. Best results are marked in bold.

	Sensitivity	Specificity	Accuracy
DWI-FLAIR mismatch	0.817	0.590	0.704
LR	**0.846**	0.643	0.745
SVM	0.769	0.765	0.765
GBDT	0.654	0.841	0.745
RF	0.692	0.800	0.749
ET	0.692	0.840	0.768
Voting	0.769	**0.840**	**0.805**

Table 5
The AUCs of five classifiers using different modal features. Best results are marked in bold.

		LR	SVM	GBDT	RF	ET
Manual labeling	DWI+FLAIR	**0.797**	0.795	**0.802**	**0.789**	**0.807**
	DWI	0.755	**0.840**	0.781	0.782	0.743
	FLAIR	0.682	0.693	0.701	0.665	0.680
Segmentation ROI	DWI+FLAIR	**0.765**	**0.811**	**0.799**	**0.786**	**0.792**
	DWI	0.706	0.693	0.728	0.669	0.668
	FLAIR	0.652	0.698	0.619	0.630	0.674

that the AUCs of five classifiers using DWI features only or FLAIR features only are significantly lower than DWI + FLAIR features in Table 5. Only one single modal image may not be adequate for TSS identification. Leveraging both DWI and FLAIR features is necessary in TSS identification. Although PWI might influence the prediction accuracy of TSS identification (Wouters et al., 2016), it is difficult to obtain PWI images due to the severe condition of stroke patients, the longer scan time or the difficulty of achieving PWI imaging in many primary hospitals, etc. In

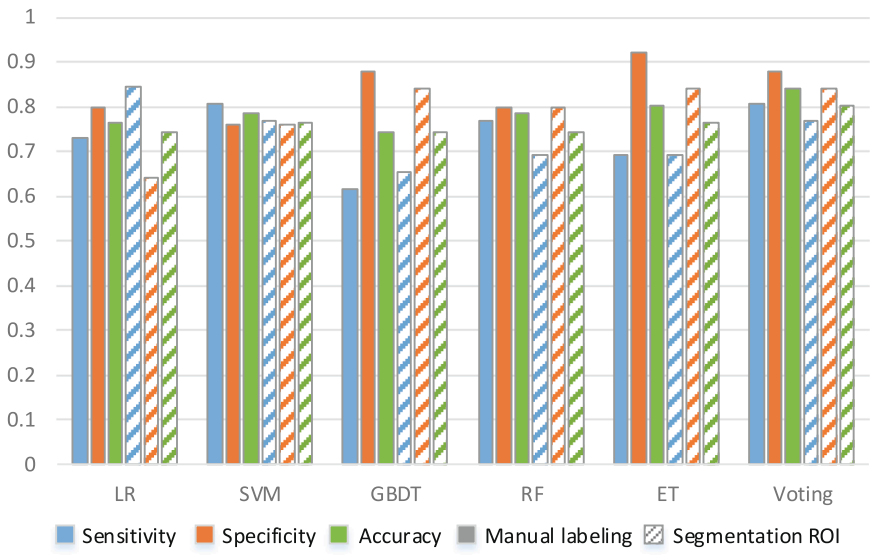

Fig. 8. The sensitivities, specificities and accuracies of different models using DWI+FLAIR features. The columns with different texture are results of manual labeling features and segmentation ROI features.

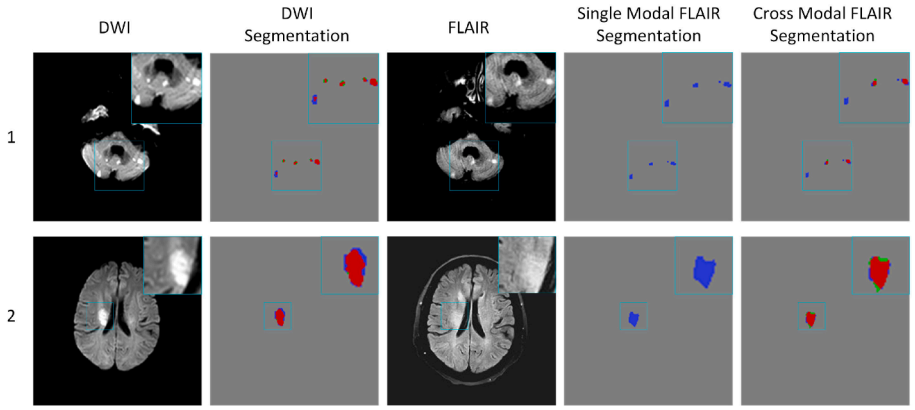

Fig. 9. The effectiveness of cross-modal network in FLAIR ROI segmentation. The true positive pixels are denoted in red, false positive pixels in green and false negative pixels in blue.

this study, we only utilized DWI and FLAIR without PWI images but can achieve higher AUCs than that reported in (King Chung et al., 2019). Our model may have better generalization capability than previous works since PWI imaging is not required. The AUC of SVM using DWI + FLAIR feature is lower than AUC using DWI only. The dimension disaster in SVM training is a point we need to concern about. Different classification models have shown different performance in TSS classification (Fig. 8). The LR and SVM tend to obtain high sensitivity but low specificity, while the GBDT, RF and ET tend to obtain high specificity but low sensitivity. The accuracies of voting are the highest while the sensitivities and specificities are not too low. The voting of five classifiers can achieve robust sensitivity, specificity and accuracy regardless of manual labeling or segmentation ROI feature. The classification accuracies of voting were both higher than that of one single classifier. The sensitivity and specificity can strike a balance instead of too high or too low. The sensitivities, specificities and accuracies of the five classifiers using segmentation ROI DWI + FLAIR features have no significant difference when compared with using radiologist's labeling features (*P-value* = 0.209). It suggests that using the segmentation ROI for TSS identification can achieve comparable performance to manual labeling. Furthermore, after training the segmentation models and the classification models, the average processing time for each test patient was less than 3s (1.5s for segmentation and 0.8s for classification) on our PC. Although the voting of five classifiers will increase the model complexity and take more time for training, the fast testing speed and high classification accuracy of our method may make it practical for computer-aided diagnosis in the future.

Our machine learning model has shown outstanding performance in TSS classification, which is better than human-derived DWI-FLAIR mismatch to some extent, the prediction results of ML model still have some problems. Fig. 10 reveals some false positive and false negative cases of the proposed method. In case 1, there is no significant hyperintensity mismatch between DWI and FLAIR. The TSS of case 1 is 4.4 h, hence the lesion in FLAIR is conspicuous which leads to the misclassification. In case 2, there is only a little match between DWI and FLAIR on slices of left cerebellum. The match is too small to guide the correct classification. Case 3 and 4 reveal the misclassification because of the lesion segmentation failure. The lesion in the FLAIR of case 3 is undetected owing to the low lesion contrast. Thus, the segmentation ROIs on DWI and FLAIR of case 3 are mismatched. Case 4 demonstrates the misclassification due to the segmentation failure on both DWI and FLAIR. The unclear lesions in DWI and FLAIR tend to be neglected by the trained segmentation networks. Besides, the collateral flow will diminish the dependency of FLAIR signal on time from symptom onset. Future work could be concentrated on discriminating the patients whose TSS are close to 4.5 h, improving the accuracy of DWI and FLAIR

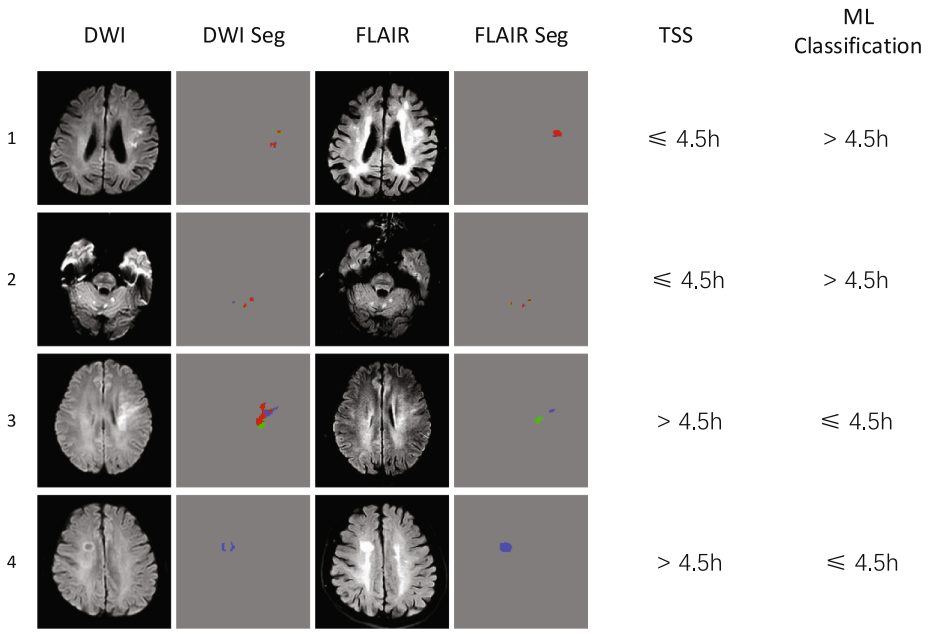

Fig. 10. The failure examples of the proposed ML model. In the segmentation columns, the true positive pixels are denoted in red, false positive pixels in green and false negative pixels in blue.

segmentation, and analyzing the predictive value of collateral flow on TSS identification.

6. Conclusion

In this work, we developed an automatic workflow to identify TSS only using two-modal images (DWI and FLAIR). We first proposed an EfficientNet-B0 based U-Net combined with MSAC block for DWI ROI segmentation. To resolve the barriers in FLAIR segmentation, we also proposed a cross-modal network. Afterwards, the features of DWI and FLAIR were generated in the segmentation ROIs and input to five classification models. In the experiment, we verified that integrating the five classifiers by voting can achieve robust accuracy in TSS identification. We anticipate that such a method will be useful for TSS unknown patients in acute stroke treatment decision-making.

CRediT authorship contribution statement

Haichen Zhu: Methodology, Software, Validation, Writing - original draft. **Liang Jiang:** Conceptualization, Data curation, Formal analysis. **Hong Zhang:** Data curation, Visualization. **Limin Luo:** Supervision, Conceptualization. **Yang Chen:** Resources, Project administration, Funding acquisition. **Yuchen Chen:** Writing - review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the State's Key Project of Research and Development Plan [Grant Nos. 2017YFC0109202, 2017YFA0104302], the National Natural Science Foundation of China [Grant Nos. 61871117], Natural Science Foundation of Jiangsu Province [Grant Nos. BK20201118], and the Science and Technology Program of Guangdong [Grant Nos. 2018B030333001].

References

Aerts, H.J., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., et al., 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature Communications 5 (1), 1–9. https://doi.org/10.1038/ncomms5006.

Aoki, J., Kimura, K., Iguchi, Y., Shibazaki, K., Sakai, K., Iwanaga, T., 2010. Flair can estimate the onset time in acute ischemic stroke patients. Journal of the Neurological Sciences 293 (1–2), 39–44. https://doi.org/10.1016/j.jns.2010.03.011.

Bewick, V., Cheek, L., Ball, J., 2005. Statistics review 14: Logistic regression. Critical Care 9 (1), 1–7. https://doi.org/10.1186/cc3045.

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Broocks, G., Leischner, H., Hanning, U., Flottmann, F., Faizy, T.D., Schön, G., Sporns, P., Thomalla, G., Kamalian, S., Lev, M.H., et al., 2020. Lesion age imaging in acute stroke: Water uptake in ct versus dwi-flair mismatch. Annals of Neurology 88 (6), 1144–1152. https://doi.org/10.1002/ana.25903.

Campbell, B.C., Ma, H., Ringleb, P.A., Parsons, M.W., Churilov, L., Bendszus, M., Levi, C. R., Hsu, C., Kleinig, T.J., Fatar, M., et al., 2019. Extending thrombolysis to 4·5–9 h and wake-up stroke using perfusion imaging: a systematic review and meta-analysis of individual patient data. The Lancet 394 (10193), 139–147. https://doi.org/10.1016/S0140-6736(19)31053-0.

Chen, L., Bentley, P., Rueckert, D., 2017. Fully automatic acute ischemic lesion segmentation in dwi using convolutional neural networks. NeuroImage: Clinical 15, 633–643. https://doi.org/10.1016/j.nicl.2017.06.016.

Coroller, T.P., Grossmann, P., Hou, Y., Velazquez, E.R., Leijenaar, R.T., Hermann, G., Lambin, P., Haibe-Kains, B., Mak, R.H., Aerts, H.J., 2015. Ct-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiotherapy and Oncology 114 (3), 345–350. doi:10.1016/j.radonc.2015.02.015.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20 (3), 273–297. https://doi.org/10.1007/BF00994018.

Ebinger, M., Galinovic, I., Rozanski, M., Brunecker, P., Endres, M., Fiebach, J.B., 2010. Fluid-attenuated inversion recovery evolution within 12 hours from stroke onset: a reliable tissue clock? Stroke 41 (2), 250–255. https://doi.org/10.1161/STROKEAHA.109.568410.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 1189–1232. https://doi.org/10.1214/aos/1013203451.

Furie, K.L., Jayaraman, M.V., 2018. guidelines for the early management of patients with acute ischemic stroke. Stroke 49 (2018), 509–510. https://doi.org/10.1161/STROKEAHA.118.020176.

Ge, R., Yang, G., Chen, Y., Luo, L., Feng, C., Ma, H., Ren, J., Li, S., 2019. K-net: Integrate left ventricle segmentation and direct quantification of paired echo sequence. IEEE Transactions on Medical Imaging 39 (5), 1690–1702. https://doi.org/10.1109/TMI.2019.2955436.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63 (1), 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Isensee, F., Maier-Hein, K.H., 2019. An attempt at beating the 3d u-net, arXiv preprint arXiv:1908.02182.

Jauch, E.C., Saver, J.L., Adams Jr, H.P., Bruno, A., Connors, J., Demaerschalk, B.M., Khatri, P., McMullan Jr, P.W., Qureshi, A.I., Rosenfield, K., et al., 2013. Guidelines for the early management of patients with acute ischemic stroke: a guideline for

healthcare professionals from the american heart association/american stroke association. Stroke 44 (3), 870–947. https://doi.org/10.1161/STR.0b013e318284056a.

Kim, B.J., Kim, Y.-H., Kim, Y.-J., Ahn, S.H., Lee, D.H., Kwon, S.U., Kim, S.J., Kim, J.S., Kang, D.-W., 2014. Color-coded fluid-attenuated inversion recovery images improve inter-rater reliability of fluid-attenuated inversion recovery signal changes within acute diffusion-weighted image lesions. Stroke 45 (9), 2801–2804. https://doi.org/10.1161/STROKEAHA.114.006515.

King Chung, H., Speier, W., Zhang, H., Scalzo, F., El-Saden, S., Arnold, C.W., 2019. A machine learning approach for classifying ischemic stroke onset time from imaging. IEEE Transactions on Medical Imaging 38 (7), 1666–1676. doi:10.1109/TMI.2019.2901445.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

Lambin, P., Leijenaar, R.T., Deist, T.M., Peerlings, J., De Jong, E.E., Van Timmeren, J., Sanduleanu, S., Larue, R.T., Even, A.J., Jochems, A., et al., 2017. Radiomics: the bridge between medical imaging and personalized medicine. Nature Reviews Clinical Oncology 14 (12), 749–762. https://doi.org/10.1038/nrclinonc.2017.141.

Lee, H., Lee, E.-J., Ham, S., Lee, H.-B., Lee, J.S., Kwon, S.U., Kim, J.S., Kim, N., Kang, D.-W., 2020. Machine learning approach to identify stroke within 4.5 hours. Stroke 51 (3), 860–866. doi:10.1161/STROKEAHA.119.027611.

Nielsen, A., Hansen, M.B., Tietze, A., Mouridsen, K., 2018. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. Stroke 49 (6), 1394–1401. https://doi.org/10.1161/STROKEAHA.117.019740.

Pinto, A., Mckinley, R., Alves, V., Wiest, R., Silva, C.A., Reyes, M., 2018. Stroke lesion outcome prediction based on mri imaging combined with clinical information. Frontiers in Neurology 9, 1060. https://doi.org/10.3389/fneur.2018.01060.

Powers, W.J., Rabinstein, A.A., Ackerson, T., Adeoye, O.M., Bambakidis, N.C., Becker, K., Biller, J., Brown, M., Demaerschalk, B.M., Hoh, B., et al., 2019. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the american heart association/american stroke association. Stroke 50 (12), e344–e418. https://doi.org/10.1161/STR.0000000000000211.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Schwamm, L.H., Wu, O., Song, S.S., Latour, L.L., Ford, A.L., Hsia, A.W., Muzikansky, A., Betensky, R.A., Yoo, A.J., Lev, M.H., et al., 2018. Intravenous thrombolysis in unwitnessed stroke onset: Mr witness trial results. Annals of Neurology 83 (5), 980–993. https://doi.org/10.1002/ana.25235.

Shi, Z., Traverso, A., van Soest, J., Dekker, A., Wee, L., 2019. Ontology-guided radiomics analysis workflow (o-raw). Medical Physics 46 (12), 5677–5684. https://doi.org/10.1002/mp.13844.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114.

Thomalla, G., Rossbach, P., Rosenkranz, M., Siemonsen, S., Krützelmann, A., Fiehler, J., Gerloff, C., 2009. Negative fluid-attenuated inversion recovery imaging identifies acute ischemic stroke at 3 hours or less. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society 65 (6), 724–732. https://doi.org/10.1002/ana.21651.

Tiwari, A., Srivastava, S., Pant, M., 2020. Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. Pattern Recognition Letters 131, 244–260. doi:10.1016/j.patrec.2019.11.020.

Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.-C., Pieper, S., Aerts, H.J., 2017. Computational radiomics system to decode the radiographic phenotype. Cancer Research 77 (21), e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339.

Wang, L., 2020. Brief report on stroke prevention and treatment in china 2019. Chinese Journal of Cerebrovascular Diseases 17, 272–280. https://doi.org/10.3969/j.issn.1672-5921.2020.05.008.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al., 2020. Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence doi:10.1109/TPAMI.2020.2983686.

Wouters, A., Dupont, P., Christensen, S., Norrving, B., Laage, R., Thomalla, G., Albers, G., Thijs, V., Lemmens, R., 2016. Association between time from stroke onset and fluid-attenuated inversion recovery lesion intensity is modified by status of collateral circulation. Stroke 47 (4), 1018–1022. https://doi.org/10.1161/STROKEAHA.115.012010.

Yin, X., Zhao, Q., Liu, J., Yang, W., Yang, J., Quan, G., Chen, Y., Shu, H., Luo, L., Coatrieux, J.-L., 2019. Domain progressive 3d residual convolution network to improve low-dose ct imaging. IEEE Transactions on Medical Imaging 38 (12), 2903–2913. https://doi.org/10.1109/TMI.2019.2917258.

Zeineldin, R.A., Karar, M.E., Coburger, J., Wirtz, C.R., Burgert, O., 2020. Deepseg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance flair images. International Journal of Computer Assisted Radiology and Surgery 15 (6), 909–920. https://doi.org/10.1007/s11548-020-02186-z.

Zhang, R., Zhao, L., Lou, W., Abrigo, J.M., Mok, V.C., Chu, W.C., Wang, D., Shi, L., 2018. Automatic segmentation of acute ischemic stroke from dwi using 3-d fully convolutional densenets. IEEE Transactions on Medical Imaging 37 (9), 2149–2160. https://doi.org/10.1109/TMI.2018.2821244.

Zhu, H., Tong, D., Zhang, L., Wang, S., Wu, W., Tang, H., Chen, Y., Luo, L., Zhu, J., Li, B., 2020. Temporally downsampled cerebral ct perfusion image restoration using deep residual learning. International Journal of Computer Assisted Radiology and Surgery 15 (2), 193–201. https://doi.org/10.1007/s11548-019-02082-1.