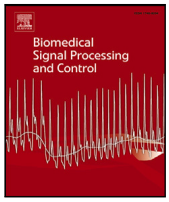
Biomedical Signal Processing and Control 100 (2025) 106897

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



A prior guide feature enhancement network for classifying ischaemic stroke onset time based on DWI and FLAIR imaging

Junjie Ying [a], Yu Xin [a,*], Chaochao Wang [b,*]

[a] Ningbo University, 818 Fenghua Road, Ningbo 315211, China
[b] Department of Radiology, Li Huili Hospital of Ningbo Medical Center, Ningbo, Zhejiang, PR China

ARTICLE INFO

Keywords:
Time since stroke
Deep learning
Domain adaptation
Feature enhancement

ABSTRACT

Thrombolytic therapy for patients with acute ischaemic stroke (AIS) is only effective within 4.5 h after stroke onset. Therefore, accurately predicting the time since stroke (TSS) is crucial in making appropriate treatment plans for AIS patients. With the rapid advancements in computer technology, deep learning techniques have made it possible to automate medical image analysis and diagnosis, improving medical diagnosis efficiency and accuracy. This makes deep learning a promising trend in medical image analysis development. The TSS classification task involves stroke lesion localization and region of interest (ROI) feature extraction. When using deep learning techniques to solve a TSS classification task, the following three issues need to be considered. (1) When using a deep learning network to locate the stroke lesion, domain shift problems exist between datasets due to differences in imaging conditions, making it difficult to obtain reliable results on other datasets using a network trained on a specific dataset. (2) Due to the small size of some stroke lesions, the corresponding stroke lesion ROI area is also small. In the CNN downsampling process, ROI features are seriously lost. (3) In MRI images, the unclear boundary and morphology of stroke lesions and other tissues make it difficult to distinguish stroke lesion features from other tissue features in feature extraction. To address the aforementioned issues, we propose a deep learning-based TSS classification network called PGFENet. In PGFENet, a multilevel adversarial learning-based domain adaptation strategy is used to address domain shift problems and reduce distribution differences between different datasets. Additionally, we design a key slice selection module to select key slices that provide rich features for classification, addressing the feature loss issue. Furthermore, a prior guide feature enhancement module is proposed to enhance stroke lesion feature discriminability, which constructs a mismatch segmentation loss based on prior knowledge of DWI-FLAIR mismatch. Finally, an attention-augmented joint prediction module is proposed to comprehensively consider the features of multiple slices to achieve reliable TSS classification predictions. The experiments demonstrate that PGFENet performs better than other mainstream classification networks. Specifically, the best TSS classification accuracy, precision, F1-score and AUC are obtained by the proposed network.

1. Introduction

Stroke is one of the leading causes of death worldwide, with acute ischaemic stroke (AIS) being the main type [1]. Currently, intravenous thrombolysis with rt-PA is the main treatment for AIS [2], but this treatment has a strict time window (4.5 h), and administering treatment beyond this window increases the risk of intracranial haemorrhage in patients [3]. Therefore, determining the time since stroke (TSS) in AIS patients, whether it is within 4.5 h, is crucial for clinical treatment. Research has found that during stroke, ischaemic tissue can be immediately observed on DWI images, while corresponding areas on FLAIR images may take 3–4 h to show ischaemic tissue [4–7] (as shown in Fig. 1). This mismatch pattern is referred to as DWI-FLAIR mismatch. It is defined as the presence of high signal intensity on DWI images without corresponding high signal intensity on FLAIR images, indicating that the onset of the patient's stroke is within 4.5 h. Conversely, if both DWI and FLAIR images show high signal intensity in corresponding regions, it suggests that the onset of the patient's stroke is more than 4.5 h ago. However, in actual diagnosis, it is difficult for the human eye to perceive the microfeatures in DWI-FLAIR images, and the evaluation process depends on the doctor's subjective experience. Therefore, the accuracy of manual DWI-FLAIR mismatch evaluation in determining TSS in AIS patients is not high in clinical practice [8–11].

* Corresponding authors.
E-mail addresses: xinyu@nbu.edu.cn (Y. Xin), lhlwangchaochao@nbu.edu.cn (C. Wang).

https://doi.org/10.1016/j.bspc.2024.106897
Received 20 June 2023; Received in revised form 28 July 2024; Accepted 7 September 2024
Available online 30 September 2024

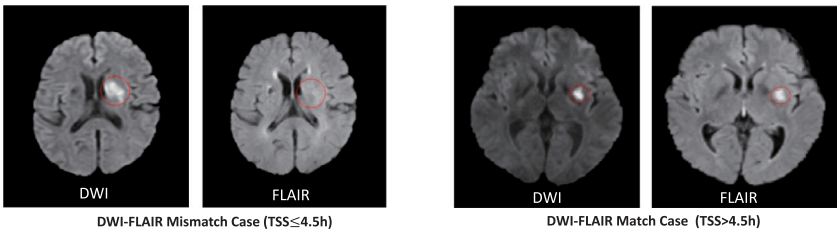


Fig. 1. Two examples of DWI-FLAIR mismatch and DWI-FLAIR match.

With the development of computer vision technology, machine learning and deep learning-based image analysis algorithms have become cutting-edge technologies in medical image analysis. Due to the advantages of strong colour and greyscale resolution and a large sensing range, machine vision can achieve lesion recognition in microseconds in complex environments and solve the problems of subjectivity, inefficiency, and a high misdiagnosis rate in human interpretation. Currently, machine learning-based TSS classification methods use various techniques to obtain ROI masks to locate stroke lesions, extract ROI features using medical imaging tools, and manually select features for modelling. Such algorithms rely on the researcher's knowledge and experience in radiology, and the feature selection process is difficult and time-consuming. In addition, the structure of machine learning models is relatively simple, making it difficult to fully capture deep image features, and machine learning-based TSS classification algorithms are prone to problems such as susceptibility to outlier data and poor robustness. With the development of deep learning technology, deep learning has the advantages of many parameters and strong robustness. It not only avoids the limitations of machine learning's need for manual feature model design but can also better capture the deep features and semantic information in medical images, improving the accuracy and reliability of medical image analysis. Currently, deep learning algorithms have surpassed traditional machine learning algorithms in some challenging medical image tasks, such as brain tumour segmentation [12] and pulmonary vessel segmentation [13].

However, deep learning-based algorithms for solving TSS classification problems require network design and parameter optimization to address the following issues. To locate the stroke lesion, a segmentation network can be used to automatically segment the stroke lesion ROI mask. However, datasets collected in actual diagnosis and treatment (ADT dataset, target domain) often lack the stroke lesion annotations required for segmentation network training. To save time on manual annotation, other datasets (source domain) with stroke lesion annotations can be used to train the segmentation network. However, due to different scan imaging conditions in different medical institutions, there are distribution differences (domain shift) between different datasets, which means that a well-trained segmentation network cannot be applied across domains. In addition, some stroke lesions are distributed in a punctuate manner along the brain sulcus or surface, with small corresponding stroke lesion ROI areas. In the downsampling process in CNN, ROI features are severely lost, making it difficult to provide effective feature information for TSS classification. Furthermore, in FLAIR images, stroke lesions have low contrast with other tissues and are similar in morphology to other high signal forms, such as brain tumours and cerebrospinal fluid, making it difficult for deep learning networks to clearly distinguish the boundaries and morphology of stroke lesions.

To address domain shift, we propose a multilevel adversarial learning-based domain adaptation strategy. By employing adversarial learning between the cross-domain segmentation network and the domain discriminator, source and target domain distribution is aligned in the output space, thus improving the cross-domain and generalization ability of the segmentation network. To address ROI feature loss during downsampling, a key slice selection module is utilized to filter out slices that are prone to feature loss during downsampling, thereby selecting slices that provide rich features for TSS classification. Furthermore, to address stroke lesion feature confusion with other tissue features, a prior guide feature enhancement module is proposed. By using the mismatch segmentation loss to enhance the stroke lesion identification capability of the encoder, the module improves stroke lesion feature discriminability.

The main innovative contributions of our work can be summarized as follows:

(1) We propose a deep learning-based TSS classification network, PGFENet. PGFENet designs a scheme to match the high signal region in DWI with the corresponding region in FLAIR and integrates the stroke lesion spatial information in DWI with the imaging features in FLAIR. This approach achieves reliable TSS classification prediction based on the fused features. Previous TSS classification work often involved simply concatenating features from multiple images and feeding them into the classification model, without fully capturing the core concept of the DWI-FLAIR mismatch method.

(2) We use a multilevel adversarial learning-based domain adaptation strategy to address the domain shift problem caused by different imaging equipment and scanning protocols in medical images between the source and target domains. Without using target domain annotations, the segmentation network can be applied to the target domain, providing more accurate stroke lesion ROIs for feature extraction. Compared to previous methods that required extracting stroke ROI features, our approach avoids the costs associated with manual ROI annotation. It also enhances the network's adaptability to images obtained under different imaging conditions.

(3) We propose a prior guide feature enhancement module that constructs a mismatch segmentation loss based on the prior knowledge of DWI-FLAIR mismatch, guiding the encoder to distinguish the boundaries and morphology of stroke lesions and enhancing the discriminative representation of stroke lesion features. This module solves the problem of feature confusion between stroke lesions and other tissues.

2. Related work

2.1. Traditional medical image classification algorithms

Early image classification algorithms were mainly based on statistical approaches, where the greyscale feature volume distribution in the lesion region was calculated and analysed to predict the classification probability. Chabat et al. [14] used the cooccurrence matrix statistical analysis algorithm to distinguish obstructive lung diseases in CT images. Rajaei et al. [15] extracted medical image features using wavelet transform and discrete cosine transform techniques. To complete tumour lesion and normal breast parenchyma classification, Guo et al. [16] used five fractal dimension estimation algorithms to analyse images and introduced the concept of lacunarity to describe the spatial distribution of pixel intensity in mammographic images. Similarly, some medical image classification algorithms use feature set combinations [17], principal component analysis (PCA) [18], and local binary patterns [19] to extract and analyse image features. However, these early algorithms based on mathematical analysis and statistics

usually have limited performance and rarely surpass the recognition level of professional doctors.

As the volume, diversity, and complexity of medical data continue to increase, machine learning has become more efficient, accurate, and comprehensive than statistical algorithms in analysing and utilizing medical images, playing an increasingly important role in the medical field. In the task of classifying cognitive visual objects, Bhalerao et al. [20] proposed the MSSDM-RCFF algorithm by combining the advantages of multivariate SSDM and RCFF to enhance the visual object recognition capabilities of EEG-MEG signals. Chaudhary et al. [21] introduced a novel multiresolution decomposition technique — the iterative Fourier–Bessel decomposition method (IFBDM) — to decompose ultrasound images into meaningful sub-band images. Convolutional neural network kernel features were then extracted from these sub-band images, utilizing these features to classify benign and malignant categories. Furthermore, Chaudhary et al. [22] proposed an empirical wavelet transform based on the two-dimensional Fourier–Bessel series expansion (2D-FBSE-EWT) method. This method uses the FBSE spectrum for boundary detection and decomposes fundus images into sub-images at multiple frequency scales, employing two methods for glaucoma detection.

Some machine learning-based medical image classification algorithms first obtain the ROI of the lesion by various techniques, then extract the ROI features in the image using medical imaging tools, and finally manually filter some imaging features to input into the machine learning model for training. Zhang et al. [23] manually labelled stroke lesions in DWI images as an ROI mask, then extracted high-dimensional features for each type of image using medical imaging tools. Finally, they used a t test, random forest and univariate logistics with a least absolute shrinkage selection operator (LASSO) to determine the final features for model training. Lee et al. [24] segmented the stroke lesion ROI mask of ADC images using normalized absolute thresholding. Zhu et al. [25] designed a cross-modal segmentation network to automatically segment DWI and FLAIR stroke lesion ROI masks, saving considerable time in manual ROI mask labelling. Ho et al. [26] utilized the PWI image region with Tmax>6 s as the stroke lesion ROI masks, extracted and screened ROI features from DWI, ADC, and FLAIR as the baseline features, and then fused the deep features extracted from the PWI images using an autoencoder for classification training.

Machine learning-based medical image classification methods can uncover information that cannot be detected by the naked eye. Their classification performance is significantly better than that of human reading, but they still require manually designed feature models and costly ROI mask generation.

2.2. Deep learning-based medical image classification algorithms

Medical image classification algorithms based on deep learning are an important application direction in medical imaging. Deep learning technology can automatically extract image features, use a large quantity of data for training, and continuously optimize the model to achieve efficient and accurate medical image classification, thereby providing strong support for medical diagnosis and treatment. Asl et al. [27] used a 3D convolutional neural network-based autoencoder to extract brain features from sMRI images to complete disease screening for Alzheimer's disease. Yan et al. [28] proposed the Full-BiLSTM model to classify mild cognitive impairment. They first used Full-LSTM to capture the time-varying information in RS-fMRI scans and then mined the long-range context in both directions using BiLSTM. Gao et al. [29] designed a 2D- and 3D-based CNN classification network to extract features from CT images and fused the judgements of both dimensions to classify Alzheimer's disease. Since unimodal medical images can only provide limited disease information and there are noises and artifacts, the accuracy and robustness of these classification algorithms based on unimodal medical images are limited.

With the development of medical image acquisition technology, multimodal medical image classification has become an important means for improving the accuracy and reliability of auxiliary diagnosis. Zhang et al. [30] used two independent branches to extract MRI and PET feature images and proposed a novel attention complementary strategy to adaptively fuse the features of the two modalities in a certain ratio to improve Alzheimer's disease prediction accuracy. Dai et al. [31] combined the advantages of CNNs and transformers, which can not only effectively extract low-level image features but also establish long-range dependencies between modalities. Bhalerao et al. [32] extracted a set of channel-aligned oscillatory components (CAOC) using an improved multivariate swarm filtering and sparse spectral technique. They then constructed multivariate joint time–frequency (JTF) images from the cross-channel joint instantaneous frequency and amplitude to develop an end-to-end framework for imagined speech detection. To enhance the performance of upper limb movement recognition, Bhalerao et al. [33] proposed the clustering sparse swarm decomposition method (CSSDM), which segments EEG signals into homogeneous clusters and employs swarm filtering and sparse spectrum to analyse the mutual characteristics across channels.

TSS classification is also a multimodal medical image analysis task. Inspired by previous work, deep learning techniques are also gradually being used in TSS classification. Zhang et al. [34] proposed an intradomain task-adaptive transfer learning method for TSS classification. This method uses a multistage training mode to improve the model performance in a more difficult task (stroke classification) by learning features from a simpler task (stroke detection). Polson et al. [35] designed multiple weight-sharing subnetworks to extract spatial information from neighbouring slices and used a trainable aggregation layer to fuse the feature output from the subnetworks to accomplish TSS classification. Gao et al. [36] developed a 3D Convolutional Neural Network (3DCNN) for TSS classification prediction. The network's input consists of concatenated channels of CBF, Tmax, and an ROI mask (Tmax > 6s). The CBF and Tmax channels provide detailed features on the severity of CBF reduction and Tmax delay, while the ROI mask channel offers a weight map indicating the extent of CBF reduction and Tmax delay. Yoon et al. [37] proposed the MM-UNET multimodal segmentation network, which captures complementary and interdependent features from cross-modal images to segment DWI and ADC images for stroke, generating an ROI mask. This ROI mask is then used to extract radiological features and deep features for TSS classification. Sun et al. [38] introduced a two-stage coarse-to-fine contrastive learning framework to enhance the localization of ischemic lesions and the perception of semantic relationships between image regions. Finally, they incorporated a multimodal region-related feature fusion module to capture feature relationships among various images for TSS classification.

These TSS classification algorithms based on deep learning simply concatenate DWI and FLAIR images and input them into a deep learning model for classification. However, they not only fail to reflect the process of matching high signal areas in DWI images with corresponding regions in FLAIR but also do not use an ROI mask to locate stroke lesions, resulting in the introduction of irrelevant information that affects classification accuracy.

3. Method

3.1. Network overview

According to the DWI-FLAIR mismatch, TSS can be determined based on whether the DWI stroke lesion high signal area and the corresponding area of FLAIR are mismatched. Therefore, as shown in Fig. 2, we first use the cross-domain segmentation network to segment the stroke lesion in DWI images and use the segmentation output as the ROI mask. Then, the corresponding region features of FLAIR images are extracted using the ROI mask, which fuses the stroke lesion spatial

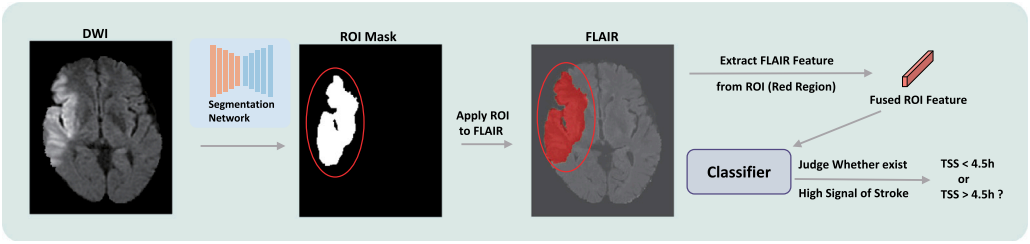


Fig. 2. The conceptual diagram of a classification network.

features of DWI with the imaging features of FLAIR. Based on the fused ROI features, the mismatch relationship between DWI and FLAIR can be determined, and TSS classification prediction can be completed.

However, there are several issues in the network design process. In terms of stroke lesion ROI mask generation, the dataset collected in actual diagnosis and treatment (ADT dataset,defined as the target domain) lacks annotations for stroke lesions. Therefore, it is necessary to use other datasets with stroke annotations (such as ISLES 2022 [39], defined as the source domain) to train the cross-domain segmentation network. However, in practical applications, due to different imaging conditions, there is a distribution difference (domain shift) between the DWI images from the source and target domains, making it difficult for the network to be cross-domain applied. In terms of fused ROI feature extraction, due to the small size of the ROI, there is a feature loss phenomenon during downsampling. Furthermore, in the fused ROI feature extraction process, there is an issue of unclear boundaries and morphology of the stroke lesion and other tissues, which leads to feature confusion between stroke lesions and other tissues.

To address the above issues, we propose a TSS classification network called PGFENet, as shown in Fig. 3, which consists of four key modules:

(1) Multilevel Adversarial Learning-Based Domain Adaptation: This module adopts a multilevel adversarial learning-based domain adaptation strategy, encouraging the cross-domain segmentation network to fool the domain discriminator and generate a segmentation distribution similar to the source domain prediction in the target domain. This process reduces the distribution difference between the source and target domains in the output space and solves the domain shift problem between the two domains.

(2) Key Slice Selection: This module selects key slices that contain rich ROI features from a set of patient slices, reducing the feature loss caused by CNN downsampling and providing a reliable judgement basis for TSS classification.

(3) Prior Guided Feature Enhancement: This module constructs a mismatch segmentation loss to enhance the encoder's discrimination ability for stroke lesion features, solving the problem of feature confusion between stroke lesions and other tissues.

(4) Attention Augment Joint Prediction: This module uses lesion attention to enhance the contribution of ROI feature vectors with significant stroke hyperintensities in the classification decision, effectively improving TSS classification performance.

3.2. Multilevel adversarial learning-based domain adaptation

Due to differences in imaging equipment, scanning protocols, and other imaging conditions, there are distribution differences in the DWI images of different datasets (such as the ISLES 2022 dataset and the ADT dataset), leading to domain shift problems. Although DWI images from the source and target domains have different imaging effects, they have many similarities in the segmentation prediction output by the network, such as the shape and spatial distribution of stroke lesions. Therefore, we consider reducing the distribution difference between the source and target domains in the output (segmentation) space to solve the domain shift problem. We adopt an adversarial learning-based domain adaptation algorithm between the domain discriminator and the cross-domain segmentation network, where the cross-domain segmentation network generates a segmentation distribution similar to the source domain prediction in the target domain, aligning the source and target domain distributions in the output space.

If the domain adaptation algorithm is only applied to the features of the final output layer, when the error gradient is backpropagated from the output layer to the lower-level features, the gradient will decrease during the propagation process (gradient vanishing), leading to problems such as semantic bias and insufficient semantic information extraction in the lower-level features. To address this issue, we adopt a multilevel adversarial learning-based domain adaptation(MALDA) strategy to conduct adversarial learning on the lower-level features, alleviating the vanishing gradient problem and improving the cross-domain segmentation network performance.

As shown in Fig. 4, the domain adaptation algorithm includes the cross-domain segmentation network and the domain discriminator. The segmentation outputs of the source and target domains are input into the domain discriminator D to distinguish whether the segmentation outputs belong to the source or target domain. A binary cross-entropy loss is used to supervise the domain discriminator, and its formula is as follows:

$$\mathcal{L}_d(P) = -\frac{1}{H'W'}\sum_{h=1}^{H'}\sum_{w=1}^{W'}(1-z)\log\left(1-D(P)^{(h,w)}\right) + z\log\left(D(P)^{(h,w)}\right) \tag{1}$$

where $P = G(I) \in \mathbb{R}^{H\times W\times C}$ is the segmentation outputted by the cross-domain segmentation network G for image I, and C is the number of classes. When P is the segmentation output from the target domain, $z = 0$; when P is the segmentation output from the source domain, $z = 1$. $D(P) \in \mathbb{R}^{H'\times W'}$ is the discrimination result outputted by the discriminator, where $H' = H/S$, $W' = W/S$, S is the downsampling factor, and HW denotes the height and width of the DWI image, respectively.

During the training of the cross-domain segmentation network, we use cross-entropy loss to supervise the segmentation outputs of the source domain:

$$\mathcal{L}_{seg}\left(I_s, Y_s\right) = -\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=0}^{C-1} Y_s^{(h,w,c)}\log P_s^{(h,w,c)} \tag{2}$$

where I_s and Y_s refer to the source domain image and its segmentation annotation, respectively. $P_s = G(I_s)$ is the segmentation output of the source domain image I_s by the cross-domain segmentation network.

We utilize adversarial loss to encourage the cross-domain segmentation network to fool the domain discriminator, resulting in segmentation outputs generated in the target domain that are similar to the distribution of segmentation outputs in the source domain. The formula for adversarial loss is presented below:

$$\mathcal{L}_{adv}\left(I_t\right) = -\frac{1}{H'W'}\sum_{h=1}^{H'}\sum_{w=1}^{W'}\log\left(D\left(P_t\right)^{(h,w)}\right) \tag{3}$$

where I_t represents the target domain image, and $P_t = G(I_t)$ is the segmentation output of the target domain image I_t by the cross-domain segmentation network.

To alleviate the vanishing gradient problem, we employ a multilevel adversarial learning strategy. In Fig. 4, in addition to the output layer

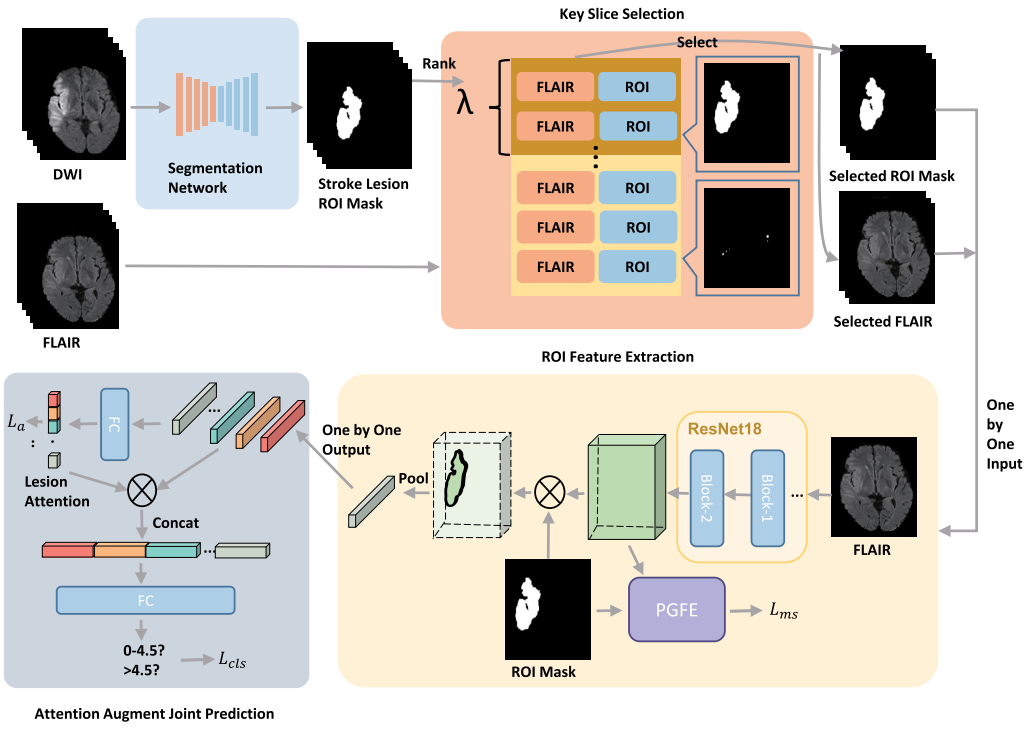


Fig. 3. The PGFENet architecture.

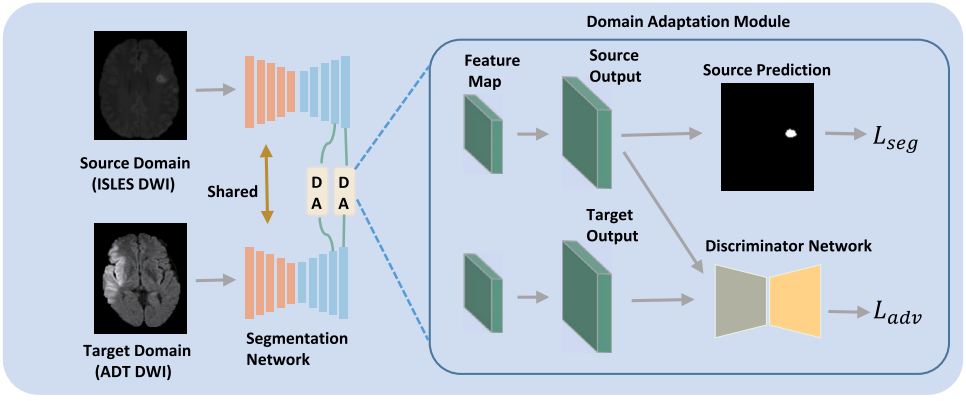


Fig. 4. Multilevel adversarial learning-based domain adaptation.

features, the lower-level features (features in the penultimate layer) are also input into the domain adaptation module for adversarial learning. This allows for more comprehensive semantic information extraction from the lower-level features and reduces the impact of semantic bias. The loss function of the cross-domain segmentation network during multilevel adversarial learning is presented as follows:

$$\mathcal{L}_{da}\left(I_s, I_t\right) = \sum_i \lambda_{\text{seg}}^i L_{seg}^i\left(I_s\right) + \sum_i \lambda_{adv}^i L_{adv}^i\left(I_t\right) \tag{4}$$

where i represents the feature of the last ith layer, and this feature is input to the domain adaptation module. Specifically, $i = \{1,2\}$, $\lambda_{seg}^i = \{1, 0.1\}$, $\lambda_{adv}^i = \{0.001, 0.0002\}$.

3.3. Key slice selection

In MRI images, the same stroke lesion may appear in multiple consecutive slices. In each slice, the area of the stroke lesion ROI varies. Compared to smaller ROIs, larger ROIs are less likely to lose features during downsampling and can provide more complete and robust feature information. To address this issue, we design a key slice selection module to select key slices that can provide rich features for classification based on the total area of the ROIs. As depicted in Fig. 3, the cross-domain segmentation network performs stroke lesion segmentation on each DWI image of a patient in the target domain. The segmentation output is binarized and used as the stroke lesion ROI mask. The ROI mask is defined as $M = \left\{M^{(i,j)}\right\}(1 \le i \le H, 1 \le j \le W)$, where:

$$M^{(i,j)} = \left[P_t^{(i,j,1)} > \tau\right] \tag{5}$$

If the stroke lesion predicted confidence of a certain pixel in $P_t = G\left(I_t\right)$ is greater than τ, it is determined that the predicted class of this pixel is a stroke lesion. We set τ to 0.5.

Next, the total area of each ROI mask is calculated using the following formula:

$$R(I_t) = \sum_{i=1}^{H}\sum_{j=1}^{W} M^{(i,j)} \tag{6}$$

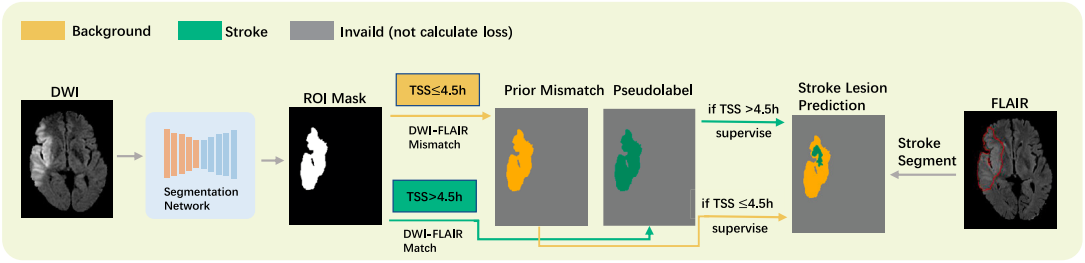


Fig. 5. Construction scheme for the prior mismatch pseudolabel.

The total area of the ROI for each slice is calculated using $R(I_t)$, and the α slices with the largest total ROI area are selected as the key slices to participate in the subsequent classification task.

We experimentally verify that the optimal value of α is 5. I_{t_p} is defined as a set of slices for a patient, and I_{key} is the set of key slices that participate in the classification. Specifically, $\left|I_{t\_p}\right| = 18$, $\left|I_{key}\right| = \alpha = 5$. The hyperparameter may be adjusted accordingly for different datasets.

3.4. Prior guide feature enhancement

In FLAIR images, the contrast between the stroke lesion and other tissues is low, and stroke lesion is similar to other high signal regions, such as brain tumours and cerebrospinal fluid. This makes it difficult for the encoder to distinguish stroke lesion from other tissues. To address this issue, we contrive a mismatch segmentation loss to guide the encoder to enhance its training on the boundaries and morphology of the stroke lesion, thereby strengthening the encoder's feature extraction ability for the stroke lesion.

According to the characteristics of DWI-FLAIR mismatch method, the cross-domain segmentation network is used to segment the stroke lesion high signal area in the DWI image as a stroke lesion ROI mask. Based on this, we construct a prior mismatch pseudolabel by combining prior knowledge of DWI-FLAIR mismatch with TSS classification labels. As shown in Fig. 5, the construction scheme is as follows:

(1) When TSS is less than or equal to 4.5 h, the ROI label is set to 0, indicating that there should be no stroke lesion high signal in the corresponding FLAIR area.

(2) When TSS is greater than 4.5 h, the ROI label is set to 1, indicating that there should be stroke lesion high signal in the corresponding FLAIR area.

The definition of the prior mismatch pseudolabel is given by the formula $Y_m = \left\{ Y_m^{(i,j)} \right\} (1 \le i \le H, 1 \le j \le W)$, where:

$$Y_m^{(i,j)} = \begin{cases} 1, \text{ if } M^{(i,j)} = 1 \text{ and } Y_{cls} = 1 \\ 0, \text{ if } M^{(i,j)} = 1 \text{ and } Y_{cls} = 0 \\ 255, \text{ else} \end{cases} \tag{7}$$

Y_{cls} is the TSS label of patient. When the TSS is less than or equal to 4.5 h, $Y_{cls} = 0$. When the TSS is greater than 4.5 h, $Y_{cls} = 1$.

As shown in Fig. 6, the FLAIR image features extracted by the encoder are passed through the pyramid pooling module and segmentation head to predict the segmentation output of the stroke lesion. Finally, we construct a mismatch segmentation loss to enhance the encoder's ability to distinguish the stroke lesion from other tissues, which is conducive to extracting discriminative fused ROI features. The segmentation loss within the ROI is used as the mismatch segmentation loss; its formula is as follows:

$$\mathcal{L}_{ms}\left(I_f\right) = -\frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \mathbb{1}_{\left[Y_m^{(h,w)} < 255\right]} \sum_{c=0}^{C-1} \mathbb{1}_{\left[Y_m^{(h,w)} = c\right]} \log P_f^{(i,j,c)} \tag{8}$$

where $P_f = G(I_f)$ is the segmentation output of the cross-domain segmentation network G applied to FLAIR image I_f.

3.5. Attention augment joint prediction

For the same patient, stroke hyperintensities appear in multiple FLAIR slices, with varying degrees of significance across different slices. The ROI feature vectors from different slices contain different amounts of stroke-related information. In some slices, the stroke hyperintensities are not prominent, and the stroke features are overshadowed by other features, making it difficult for the classifier to make clear classification predictions.

To address the aforementioned issue, we propose an attention-augmented joint prediction classifier to enhance the contribution of ROI feature vectors with significant stroke hyperintensities in the classification decision. The structure is shown in Fig. 3. The feature extraction module extracts key slice features and pools them into a set of ROI feature vectors $V = \{V_i\}(1 \le i \le \alpha)$. To ensure that the fully connected layer classifier focuses more on ROI feature vectors containing substantial stroke information during the classification process, we introduce lesion attention to weight the ROI feature vectors separately. This reassigns the contribution weights of each ROI feature vector in the classification decision, which can be defined as $A = \{A^i\}(1 \le i \le \alpha)$, where:

$$A^i = sigmoid(ReLU(W_\theta V_i)) \tag{9}$$

$W_\theta \in \mathbb{R}^{L \times 1}$ represents the weights of the fully connected layer, and L represents the dimension of the ROI feature vector.

The key slice selection module has already filtered out the DWI images that do not contain strokes, allowing us to use the TSS classification labels to supervise A^i. The loss function is as follows:

$$\mathcal{L}_a(A) = -\frac{1}{\alpha} \sum_{i=1}^{\alpha} Y_{cls} \log A^i \tag{10}$$

Lesion attention is multiplied by the ROI feature vector to enhance the stroke lesion feature representation in the ROI feature vector. The enhanced ROI feature vector is defined as follows:

$$\bar{V}^i = A^i \times V^i, (1 \le i \le \alpha) \tag{11}$$

Finally, the vectors from $\bar{V}$ are concatenated and input into the fully connected classifier for joint TSS prediction. TSS classification labels are used to supervise the prediction; the loss function is as follows:

$$\mathcal{L}_{cls}\left(\bar{V}, Y_{cls}\right) = -\frac{1}{N} \sum_{i=1}^{N} Y_{cls}^i \log\left(sigmoid\left(ReLU\left(W_{fc} ConCat(\bar{V})\right)\right)\right) \tag{12}$$

where N is the number of patient samples and $W_{fc} \in \mathbb{R}^{\alpha L \times 1}$ is the parameter of the fully connected classifier.

The total loss for TSS classification training of PGFENet is defined as follows:

$$\mathcal{L}_{cls\_total} = \lambda_{cls}\mathcal{L}_{cls} + \lambda_a \mathcal{L}_a + \lambda_{ms}\mathcal{L}_{ms} \tag{13}$$

where $\lambda_{cls} = 1$, $\lambda_a = 0.1$, $\lambda_{ms} = 0.05$.

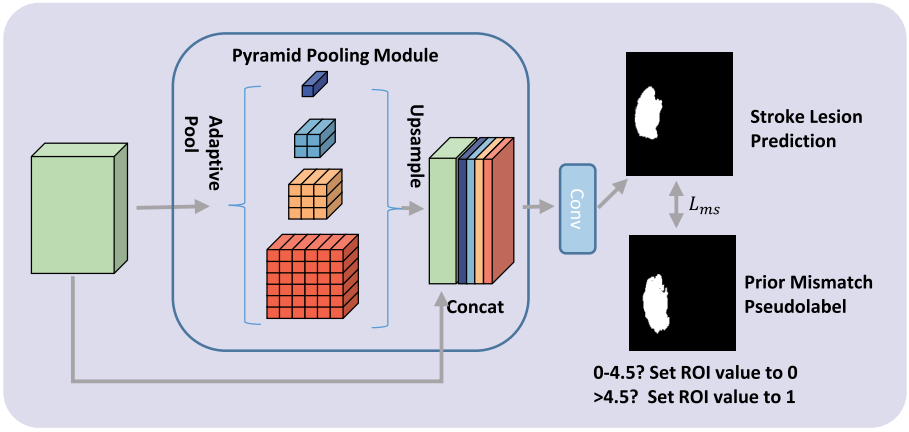


Fig. 6. Prior guide feature enhancement module.

4. Experiments

4.1. Datasets and preprocessing

The datasets used in this paper include the ISLES 2022 dataset and the ADT dataset. ISLES 2022 is a magnetic resonance imaging stroke lesion segmentation dataset that publicly releases MRI images of 250 patients, including DWI, FLAIR and other images, and the stroke lesions were annotated by experts. The DWI images are obtained from a TR range of 3,175–16,439 ms and a TE range of 55–91 ms. The pixel dimensions range from $0.57 \times 0.57 \times 2.00$ mm to $2.0 \times 2.0 \times 6.5$ mm, and the number of images scanned for each patient ranges from 24 to 74. The image resolution in the dataset ranges from 112×112 to 256×256, with 112×112 resolution being the majority. In terms of data preprocessing, we use the functions of MONAI [40] to process the image data. First, the DWI images and their stroke lesion annotations are uniformly resized to a size of 112×112. Then, the images and their stroke lesion annotations are padded to a resolution of 128×128 using the SpatialPad function, and a random selection of 35 consecutive images is made (with padding of 0 for less than 35 images). Finally, these images are subjected to data augmentation operations such as RandFilp, RandScaleIntensity, RandShiftIntensity, and NormalizeIntensity and are used as network inputs.

The ADT dataset collected AIS patient data from a hospital in Zhejiang Province, China, from December 2019 to November 2022, including MRI images of 187 patients, including DWI and FLAIR images. According to the stroke onset time, they were divided into negative (TSS) and positive (TSS) as labels for TSS classification. The ADT dataset was obtained using a Philips or GE echo planar scanner, with the number of slices per patient ranging from 18 to 23. The pixel dimensions of DWI images range from $0.78 \times 0.78 \times 6.5$ mm to $0.94 \times 0.94 \times 7$ mm, and the pixel dimensions of FLAIR images range from $0.76 \times 0.76 \times 5.5$ mm to $0.94 \times 0.94 \times 7.3$ mm. The patient characteristics are shown in Table 1, and the distribution of clinical factors among patients is similar. The dataset was divided into training, validation, and test sets at a ratio of 3:1:1. In terms of data preprocessing, the DWI and FLAIR images were registered and skull-stripped using 3D Slicer, and data augmentation was also performed using MONAI. Eighteen consecutive images were randomly selected as network inputs.

From the detailed descriptions of the two datasets above, it can be seen that the ISLES 2022 dataset contains DWI images and their stroke lesion annotations, while the ADT dataset contains DWI and FLAIR images as well as TSS classification labels. In this paper, the DWI images and stroke lesion annotations from the ISLES 2022 dataset were used as the source domain, and the DWI images from the ADT dataset were used as the target domain. The cross-domain segmentation network was trained using the MALDA strategy. For the TSS classification task, DWI-FLAIR images and the TSS classification labels from the ADT dataset and the ROI mask provided by the cross-domain segmentation network were used for classification training.

Table 1
Demographic characteristics of AIS patients in the ADT dataset. Numbers are n (%) or median (interquartile ranges). MRI indicates magnetic resonance imaging; and NIHSS, National Institutes of Health Stroke Scale.

	Training set (n=112)	Test set (n=37)
Age(years)	66(57-74)	66(54-75)
Female	42(37.5%)	13(35.1%)
NIHSS on admission	5(3-12)	6(2-11)
Onset to MRI(min)	254(126-698)	243(132-656)
Within 4.5-hour window	42(37.5%)	14(37.8%)

4.2. Implementation details

Network framework. The network in MALDA consists of a cross-domain segmentation network and a domain discriminator. In this paper, the classic medical image segmentation network U-Net [41] is used as the cross-domain segmentation network, with the encoder using VGG-16 [42]. For the discriminator design, a fully convolutional network is used to perform domain discrimination on the segmentation outputs. The discriminator is composed of 5 convolutional layers with a 4×4 kernel and a stride of 2, and the number of channels in each layer is $\{64, 128, 256, 512, 1\}$. The first 4 convolutional layers are coupled with Leaky-ReLU [43] with a negative slope of 0.2. As shown in Fig. 4, the last two feature layers output by the decoder are input into two discriminators with the same structure for multilevel adversarial learning, improving the network's ability to extract semantic information from lower-level features and reducing the impact of semantic bias.

The overall PGFENet architecture is shown in Fig. 3 . The cross-domain segmentation network trained by MALDA is used to segment DWI images, providing accurate stroke lesion ROI masks for TSS classification. The ROI masks and FLAIR images selected by the key slice selection module are input into the ROI feature extraction module to extract fused ROI features. Because the texture information in FLAIR images is an important basis for TSS classification, the low-level texture features output by block2 in ResNet-18 are input into the prior guide feature enhancement (PGFE) module to enhance stroke lesion feature discriminability. The PGFE module uses a pooling pyramid module to improve segmentation performance by utilizing context information at different scales. The pyramid module consists of 4 parallel adaptive pooling layers with output sizes of 1×1, 2×2, 3×3, and 6×6, and feature extraction is performed using the "pooling layer-upsampling layer-1×1 convolution layer". Finally, the feature maps of the 4

branches are concatenated and input into the segmentation head to predict segmentation outputs.

Pretraining strategy. Due to the limited sample size of the ADT dataset, the deep learning network was not adequately trained, which limited the network's performance and generalization ability. Therefore, we used MoCo [44] to pretrain the encoder (ResNet-18) to improve the downstream classification task performance. MoCo is a self-supervised method based on contrastive learning, so positive and negative sample pairs need to be constructed. In MoCo, a positive sample pair consists of an image and its augmented image, while a negative sample pair consists of an image and other images in a queue. For TSS classification, DWI and FLAIR are two views of the same patient's brain imaging, which can be directly used as positive sample pairs. In the ablation study, it was verified that the PGFENet classification performance was greatly improved after using weights pretrained with MoCo for the encoder.

Network training and inference. All networks were implemented using PyTorch [45]. All experiments were trained on an NVIDIA 3090 GPU with 24 GB of memory. During the MALDA training process, both the cross-domain segmentation network and the domain discriminator used the Adam optimizer [46]. The optimizer for the cross-domain segmentation network had an initial learning rate of 2.5×10^{-4}, while the optimizer for the domain discriminator had an initial learning rate of 10^{-4}. A polynomial decay curve with a power of 0.9 was used to adjust the learning rate, and the training was performed for 500 epochs. During the classification network training process, the cross-domain segmentation network trained by MALDA was frozen, and the Adam optimizer was used with an initial learning rate of 10^{-4}. A polynomial decay curve with a power of 0.9 was also used to adjust the learning rate, and the training was performed for 200 epochs. The network weights with the highest accuracy on the validation set were saved, and the network was tested on the test set.

When inferring with PGFENet, DWI and FLAIR image sets from an AIS patient are used as input, with each set containing 18 slices. The network outputs a prediction of the patient's TSS. Inference is performed on an RTX 3090 GPU, with a parameter count of 36.3M. The inference runs with 10.48 person/s, 1.02 TFLOPs, and 3.6 GB GPU memory. The inference performances indicate that PGFENet operates at a high speed, achieving an inference time of 95.32 ms per patient on an RTX 3090 GPU. Additionally, the low GPU memory usage allows it to run smoothly even on GPUs with smaller memory capacities, which is particularly important for deployment on resource-constrained devices.

4.3. Evaluation metrics

We used several metrics to evaluate the classification performance of PGFENet. These metrics include accuracy (ACC), precision (PR), F1-score (F1), and AUC. Accuracy is the proportion of correctly classified samples to the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

Precision is the proportion of true positive samples to the total number of samples classified as positive by the classifier.

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

The F1-score is a harmonic mean of precision and recall, which considers both precision and recall.

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{16}$$

AUC is a metric used to evaluate binary classification network performance, which represents the area under the ROC curve.

Table 2
Comparisons of classification performance on the ADT dataset.

Network	Type	ACC(%)	PR(%)	F1(%)	AUC(%)
ShuffleNetV2(1.0X)[47]	CNN	48.6	56.7	64.2	40.5
InceptionV3[48]	CNN	54.1	59.4	69.1	44.9
DenseNet-121[49]	CNN	56.8	64.0	66.7	52.6
VGG-16[42]	CNN	56.8	66.7	63.6	55.4
ResNet-50[50]	CNN	59.5	64.3	70.6	53.4
GC ViT-T [51]	Transformer	59.5	62.5	72.7	50.6
TNT-S [52]	Transformer	62.2	65.5	73.1	55.6
Swin-T [53]	Transformer	62.2	71.4	68.2	61.2
MViTv2-T [54]	Transformer	64.9	69.2	73.5	60.6
UNETR [55]	MSN	59.5	63.3	71.7	52
AD-Net [56]	MSN	62.2	69.6	69.6	59.8
NestFormer [12]	MSN	67.6	66.7	54.5	43.9
Ours	–	**81.1**	**76.7**	**86.8**	**75**

4.4. Performance comparison

To better evaluate the performance and effectiveness of PGFENet, we conducted comparative experiments on the ADT dataset using currently mainstream CNN networks, transformer networks, and multimodality segmentation networks (MSNs). In the experiment, a CNN or transformer network was used as the feature extraction network. The DWI and FLAIR images of a patient were concatenated by channel as the input of the feature extraction network. Finally, the DWI and FLAIR features were concatenated along the channel dimension to predict TSS for the patient. In addition, we modified the structures of medical image multimodality segmentation networks, to adapt to the classification prediction and included them in the comparative experiments.

In Table 2, it can be observed that the classification performance of mainstream CNN networks for TSS is poor. Networks based on the transformer architecture, which use a self-attention mechanism to capture long-range dependencies between different image regions, have better overall TSS classification performance than CNN-based networks. However, networks based on both CNN and transformer architectures did not consider a solution for multimodal data fusion, resulting in poor TSS classification performance that involves DWI-FLAIR bimodal data. Although the multimodal networks considered DWI and FLAIR feature fusion, they did not address the problem of stroke lesion localization or the matching scheme for DWI-FLAIR stroke lesion regions, which affected their TSS classification performance. The TSS classification network proposed in this paper uses a cross-domain segmentation network to generate ROI masks for stroke lesion localization. By utilizing the DWI-FLAIR matching scheme, the stroke lesion spatial features of DWI and imaging features of FLAIR are effectively fused, thus improving TSS classification performance. The classification accuracy (81.1%), precision (76.7%), F1-score (86.8%), and AUC (75%) achieved by PGFENet reached the best level among the comparative experiments.

4.5. Ablation study

Validate the effectiveness of each module. We conducted ablation experiments to verify the effectiveness of each module in PGFENet, and the results are shown in Table 3 . The complete PGFENet achieved an accuracy of 81.1% on the ADT dataset (Row 7), which was 21.6% higher than the encoder without any modules (Row 1). When the encoder was initialized with random weights, the classification accuracy of the network decreased by 11% (cf. Row 2 and Row 7). This indicates that when training samples are limited, using self-supervised methods to pretrain the encoder can effectively enhance the feature representation of medical images and improve TSS classification performance. In addition, when the cross-domain segmentation network was not trained with the MALDA strategy, the quality of the generated stroke lesion ROI mask was poor (detailed visualization results in Section 4.6), and the TSS classification accuracy decreased by approximately 16% (cf. Row

Table 3
Network performance with different modules.

	Mo.pre	MALDA	LA	PGFE	KSS	ACC(%)	PR(%)	F1(%)	AUC(%)
1	–	–	–	–	–	59.5	62.5	72.7	50.6
2	–	✓	✓	✓	✓	70.3	73.1	77.6	66.3
3	✓	–	✓	✓	✓	64.9	69.2	73.5	60.6
4	✓	✓	–	✓	✓	70.3	67.6	80.7	60.7
5	✓	✓	✓	–	✓	75.7	75.0	82.4	70.7
6	✓	✓	✓	✓	–	67.6	67.7	77.8	59.9
7	✓	✓	✓	✓	✓	**81.1**	**76.7**	**86.8**	**75**

3 and Row 7). When the lesion attention (LA) was removed from the joint prediction classifier, the classifier could not focus on the stroke lesion feature representation in the ROI feature vector, leading to a decrease in classification accuracy of approximately 10.7% (cf. Row 4 and Row 7). When the encoder lost the PGFE module for extracting stroke discriminative features, the classification accuracy decreased by approximately 5.4%(cf. Row 5 and Row 7). When all slices of a patient were included in TSS classification, the classification accuracy decreased by 13.5% (cf. Row 6 and Row 7). Therefore, using the key slice selection (KSS) module to select key slices that can provide rich features for classification can avoid introducing unnecessary noise and interference during network training.

Validate the impact of individual features and fusion features. To explore the importance of fusion features in the classification decision process, we designed ablation experiments where the network uses single features and fusion features for TSS classification prediction. Since the inputs to the KSS and PGFE modules are related to the DWI stroke lesion spatial features (DWI-SLSF) and FLAIR image features (FLAIR-IF), in designing the ablation experiments for single features, we directly input the DWI stroke lesion ROI mask and FLAIR images separately into the feature extraction module. The results are shown in Table 4.

When using only the DWI stroke lesion features for TSS classification prediction, the network cannot determine whether the corresponding region in the FLAIR image contains stroke features, and thus cannot assess the mismatch between DWI and FLAIR, resulting in poor performance. When the network uses only FLAIR image features for TSS classification prediction, the network loses the guidance of the PGFE module, making it difficult to distinguish stroke hyperintensities from other hyperintensities in the FLAIR images, and it also cannot judge the mismatch between DWI and FLAIR, resulting in poor performance with only 54.1% accuracy.

Since the network can simply judge TSS based on the presence of hyperintensities in FLAIR images, it can still make correct predictions in images without other hyperintensity interferences. Therefore, the performance of the network using only FLAIR image features for TSS classification prediction is better than using only DWI stroke spatial features, with an accuracy improvement of about 5.4% (cf. row 1 and row 2).

Based on the DWI-FLAIR mismatch, TSS can be determined by whether there is a mismatch between the stroke hyperintensity region in DWI and the corresponding region in FLAIR. Therefore, we fuse the stroke spatial features of DWI with the image features of FLAIR. The network uses the fusion ROI features to determine the mismatch relationship between DWI and FLAIR and complete the TSS classification prediction. Thus, using fusion features for TSS classification prediction fully aligns with the DWI-FLAIR mismatch concept. The performance shows a significant improvement compared to using single-feature classification schemes, with the accuracy increasing from 59.5% to 81.1%.

Validate the impact of architectural differences. To explore the performance differences brought by using different architectures in PGFENet, we conducted ablation experiments by replacing different architectures in the ROI feature extraction module. As shown in Table 5,

Table 4
Network performance with individual features and fusion features.

	DWI-SLSF	FLAIR-IF	ACC(%)	PR(%)	F1(%)	AUC(%)
1	✓	–	54.1	62.5	63.8	50.5
2	–	✓	59.5	64.3	70.6	53.4
3	✓	✓	**81.1**	**76.7**	**86.8**	**75**

Table 5
Network performance with different architectures.

Architecture	Params	ACC(%)	PR(%)	F1(%)	AUC(%)
MobileViT-S [57]	4.9M	64.9	63.9	78	53.6
ResNet-18[50]	11.2M	**81.1**	**76.7**	**86.8**	**75**
TNT-S [52]	23.4M	67.6	69	76.9	61.3
ResNet-50[50]	23.5M	62.2	62.2	76.7	50
VGG-16[42]	134.3M	59.5	64.3	70.6	53.4

the larger the parameters of the architecture, the more likely the network is to experience overfitting, resulting in poorer TSS classification prediction performance. Additionally, given that the dataset used in this study is relatively small (a total of 187 cases), it is even more critical to control the network's parameters. When the network's parameters is too small, the network's learning capacity is limited, and it cannot fully learn the features from the training data, leading to underfitting. Therefore, we ultimately selected ResNet-18, which has a relatively appropriate parameters, as the architecture for PGFENet.

4.6. Visualization of segmentation outputs before and after domain adaptation

Fig. 8 presents the visualization of the DWI image segmentation outputs on the ADT dataset by the cross-domain segmentation network when trained without (seg w/o MALDA, Row 2) and with (seg with MALDA, Row 3) the MALDA strategy.

We randomly selected 5 samples, and the visualization of the segmentation outputs from Fig. 8 shows that seg with MALDA generates segmentation outputs closer to the ground truth(marked with a blue line) than seg w/o MALDA, indicating that the former has better segmentation performance. The reason is that there is a domain shift between the ISLES 2022 dataset and the ADT dataset due to different imaging conditions, which makes the segmentation network trained without the MALDA strategy unable to be cross-domain applied. The distribution difference between the two datasets, such as the difference in grey-level distribution (as shown in Fig. 7), leads to segmentation inaccuracies in seg w/o MALDA, such as oversegmentation and undersegmentation in sample 1 and sample 2, respectively. However, in seg with MALDA, the segmentation network trained with the MALDA strategy can effectively reduce the distribution difference between the two datasets in the output space, accurately segmenting the stroke lesion. In addition, compared to the ISLES 2022 dataset, the ADT dataset has lower image quality and more artefacts and noise. For example, magnetic susceptibility artefacts are present in sample 3 and sample 4. In seg w/o MALDA, the segmentation network incorrectly identifies these artefacts as stroke lesions. However, in combination with MALDA, the MALDA strategy enables the segmentation network to generate a stroke lesion spatial distribution similar to that of the ISLES dataset, helping to distinguish artefacts from stroke lesions and improving segmentation reliability.

From the visualization of the segmentation outputs before and after domain adaptation, it can be observed that the stroke lesion segmentation performance of the segmentation network improved after being trained with the MALDA strategy. In the ablation study, we verified that as the stroke lesion ROI is segmented more accurately, the TSS classification performance is significantly improved.

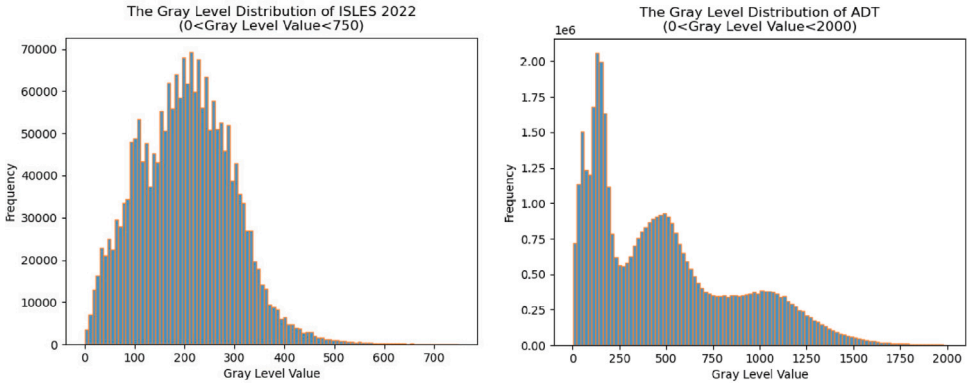


Fig. 7. The grey-level distribution of the ISLES 2022 and ADT datasets.

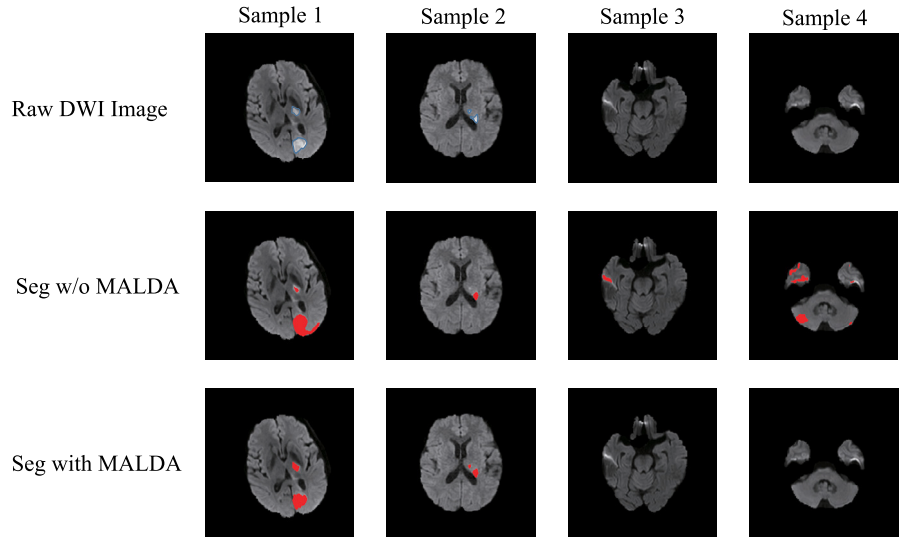


Fig. 8. Visualization of segmentation outputs before and after domain adaptation.

5. Conclusion

To address the TSS classification problem, we propose a deep learning-based classification network called PGFENet. To solve the domain shift problem between the source domain (such as the ISLES 2022 dataset) and the target domain (such as the ADT dataset), we use a domain adaptation strategy based on multilevel adversarial learning to train a cross-domain segmentation network, aligning the distribution differences of the source and target domains in the output space, and providing accurate stroke lesion ROI masks for TSS classification. Based on this, a key slice selection module is proposed to filter out key slices that provide rich features for reliable TSS classification. Furthermore, to address the problem of feature confusion between stroke lesions and other tissues, a prior guide feature enhancement module is proposed, which enhances the encoder's stroke discrimination ability by using the mismatch segmentation loss. Finally, an attention augment joint prediction module is proposed, which uses lesion attention to enhance the contribution of ROI feature vectors with significant stroke hyperintensities in the classification decision and considers the features of multiple slices to achieve reliable TSS classification prediction. In experiments, PGFENet is compared with mainstream classification networks; it achieves the best classification performance, with accuracy, precision, F1-score, and AUC all being the highest, which are 81.1%, 76.7%, 86.8%, and 75%, respectively.

When deploying PGFENet in clinical settings, several practical challenges remain. First, the output of the network only determines whether the stroke onset time for AIS patients is within 4.5 h, without providing insight into the decision-making process of the network. In clinical environments, healthcare professionals need to understand the network's decision-making process and rationale to trust and rely on its results. Second, all imaging data for the patients comes from a single centre, and the network parameters are optimized and adjusted based on this centre's data. The trained network lacks validation with multi-centre data, as images obtained from different centres using various scanning equipment may differ. Consequently, the generalization performance of a network trained on single-centre data may not be as robust when applied to other centres.

During the course of our research, we encountered several limitations. First, the imaging sample size used in this study was relatively small. Deep learning models require a large amount of high-quality data for training to achieve stable performance. Therefore, future work will involve collecting more imaging samples to improve the network's performance. Second, the network design did not take into account other clinical features and past medical history of the patients. There are individual differences among patients, and ignoring personalized characteristics such as age, blood pressure, blood sugar levels, and past medical history can affect the network's recognition performance. Hence, future work will consider using a multimodal framework like CLIP to jointly analyse patients' past medical history in text form, clinical features, and images. The global-level image-text alignment characteristics of CLIP can help the network provide a more comprehensive evaluation for classification tasks. Lastly, this study did not

include all AIS patients, excluding those with trauma, hemorrhage, and extensive cerebral white matter lesions, as their images might provide misleading information for TSS classification. To analyse more types of AIS patients in future research, we will consider incorporating other types of images (such as PWI images) to provide more complementary decision-making information for TSS classification.

CRediT authorship contribution statement

Junjie Ying: Investigation, Methodology, Writing – original draft, Writing – review & editing. **Yu Xin:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing. **Chaochao Wang:** Data curation, Validation, Writing – review & editing.

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.

Data availability

The authors do not have permission to share data.

Acknowledgements

We acknowledge the support of the National Natural Science Foundation of China under Grant No. 61602133; Natural Science Foundation of Ningbo, China (2019A610093); The 3315 Plan Foundation of Ningbo No. 2019B-18-G; The Major Project of Ningbo Technological Innovation 2025 No. 2019B10125; China NSF under Grant 62271274, in part by the Zhejiang NSF under Grant LZ20F020001.

References

[1] M.S. Phipps, C.A. Cronin, Management of acute ischemic stroke, BMJ 368 (2020).
[2] N.T. Cheng, A.S. Kim, Intravenous thrombolysis for acute ischemic stroke within 3 h versus between 3 and 4.5 h of symptom onset, Neurohospitalist 5 (3) (2015) 101–109.
[3] B.C. Campbell, H. Ma, P.A. Ringleb, M.W. Parsons, L. Churilov, M. Bendszus, C.R. Levi, C. Hsu, T.J. Kleinig, M. Fatar, Extending thrombolysis to 4 · 5–9 h and wake-up stroke using perfusion imaging: a systematic review and meta-analysis of individual patient data, Lancet 394 (10193) (2019) 139–147.
[4] S. Emeriau, I. Serre, O. Toubas, F. Pombourcq, C. Oppenheim, L. Pierot, Can diffusion-weighted imaging–fluid-attenuated inversion recovery mismatch (positive diffusion-weighted imaging/negative fluid-attenuated inversion recovery) at 3 t identify patients with stroke at< 4.5 h? Stroke 44 (6) (2013) 1647–1651.
[5] G. Thomalla, B. Cheng, M. Ebinger, Q. Hao, T. Tourdias, O. Wu, J.S. Kim, L. Breuer, O.C. Singer, S. Warach, DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4 · 5 h of symptom onset (PRE-FLAIR): a multicentre observational study, Lancet Neurol. 10 (11) (2011) 978–986.
[6] G. Thomalla, P. Rossbach, M. Rosenkranz, S. Siemonsen, A. Krützelmann, J. Fiehler, C. Gerloff, Negative fluid-attenuated inversion recovery imaging identifies acute ischemic stroke at 3 h or less, Ann. Neurol. 65 (6) (2009) 724–732.
[7] M. Ebinger, I. Galinovic, M. Rozanski, P. Brunecker, M. Endres, J.B. Fiebach, Fluid-attenuated inversion recovery evolution within 12 h from stroke onset: a reliable tissue clock? Stroke 41 (2) (2010) 250–255.
[8] A. Ziegler, M. Ebinger, J.B. Fiebach, H.J. Audebert, S. Leistner, Judgment of FLAIR signal change in DWI–FLAIR mismatch determination is a challenge to clinicians, J. Neurol. 259 (2012) 971–973.
[9] I. Galinovic, J. Puig, L. Neeb, J. Guibernau, A. Kemmling, S. Siemonsen, S. Pedraza, B. Cheng, G. Thomalla, J. Fiehler, Visual and region of interest–based inter-rater agreement in the assessment of the diffusion-weighted imaging–fluid-attenuated inversion recovery mismatch, Stroke 45 (4) (2014) 1170–1172.
[10] J. Aoki, K. Kimura, Y. Iguchi, K. Shibazaki, K. Sakai, T. Iwanaga, FLAIR can estimate the onset time in acute ischemic stroke patients, J. Neurol. Sci. 293 (1–2) (2010) 39–44.
[11] B.J. Kim, Y.-H. Kim, Y.-J. Kim, S.H. Ahn, D.H. Lee, S.U. Kwon, S.J. Kim, J.S. Kim, D.-W. Kang, 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) (2014) 2801–2804.
[12] Z. Xing, L. Yu, L. Wan, T. Han, L. Zhu, Nestedformer: Nested modality-aware transformer for brain tumor segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V, Springer, pp. 140–150.
[13] R. Wu, Y. Xin, J. Qian, Y. Dong, A multi-scale interactive U-net for pulmonary vessel segmentation method based on transfer learning, Biomed. Signal Process. Control 80 (2023) 104407.
[14] F. Chabat, G.-Z. Yang, D.M. Hansell, Obstructive lung diseases: texture classification for differentiation at CT, Radiology 228 (3) (2003) 871–877.
[15] A. Rajaei, L. Rangarajan, Wavelet features extraction for medical image classification, Int. J. Eng. Sci. 4 (Sept 2011) (2011).
[16] Q. Guo, J. Shao, V.F. Ruiz, Characterization and classification of tumor lesions using computerized fractal-based texture analysis and support vector machines in digital mammograms, Int. J. Comput. Assist. Radiol. Surg. 4 (2009) 11–25.
[17] M. Radhakrishnan, T. Kuttiannan, N. Tiruchengode, Comparative analysis of feature extraction methods for the classification of prostate cancer from TRUS medical images, IJCSI Int. J. Comput. Sci. Issues 9 (1) (2012) 171–179.
[18] N. Abdullah, L.W. Chuen, U.K. Ngah, K.A. Ahmad, Improvement of MRI brain classification using principal component analysis, in: 2011 IEEE International Conference on Control System, Computing and Engineering, IEEE, 2011, pp. 557–561.
[19] D. Liu, S. Wang, D. Huang, G. Deng, F. Zeng, H. Chen, Medical image classification using spatial adjacent histogram based on adaptive local binary patterns, Comput. Biol. Med. 72 (2016) 185–200.
[20] S.V. Bhalerao, R.B. Pachori, Automated classification of cognitive visual objects using multivariate swarm sparse decomposition from multichannel EEG-MEG signals, IEEE Trans. Hum.-Mach. Syst. (2024).
[21] P.K. Chaudhary, K. Das, R.B. Pachori, Breast Cancer Diagnosis Using Iterative Fourier-Bessel Decomposition Method Based CNN-Kernel Features, Authorea, 2023, Authorea Preprints.
[22] P.K. Chaudhary, R.B. Pachori, Automatic diagnosis of glaucoma using two-dimensional Fourier-bessel series expansion based empirical wavelet transform, Biomed. Signal Process. Control 64 (2021) 102237.
[23] Y.-Q. Zhang, A.-F. Liu, F.-Y. Man, Y.-Y. Zhang, C. Li, Y.-E. Liu, J. Zhou, A.-P. Zhang, Y.-D. Zhang, J. Lv, MRI radiomic features-based machine learning approach to classify ischemic stroke onset time, J. Neurol. (2022) 1–11.
[24] H. Lee, E.-J. Lee, S. Ham, H.-B. Lee, J.S. Lee, S.U. Kwon, J.S. Kim, N. Kim, D.-W. Kang, Machine learning approach to identify stroke within 4.5 h, Stroke 51 (3) (2020) 860–866.
[25] H. Zhu, L. Jiang, H. Zhang, L. Luo, Y. Chen, Y. Chen, An automatic machine learning approach for ischemic stroke onset time identification based on DWI and FLAIR imaging, Neuroimage 31 (2021) 102744.
[26] K.C. Ho, W. Speier, H. Zhang, F. Scalzo, S. El-Saden, C.W. Arnold, A machine learning approach for classifying ischemic stroke onset time from imaging, IEEE Trans. Med. Imaging 38 (7) (2019) 1666–1676.
[27] E. Hosseini-Asl, G. Gimel'farb, A. El-Baz, Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network, 2016, arXiv preprint arXiv:1607.00556.
[28] W. Yan, H. Zhang, J. Sui, D. Shen, Deep chronnectome learning via full bidirectional long short-term memory networks for MCI diagnosis, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11, Springer, pp. 249–257.
[29] X.W. Gao, R. Hui, Z. Tian, Classification of CT brain images based on deep learning networks, Comput. Methods Programs Biomed. 138 (2017) 49–56.
[30] T. Zhang, M. Shi, Multi-modal neuroimaging feature fusion for diagnosis of Alzheimer's disease, J. Neurosci. Methods 341 (2020) 108795.
[31] Y. Dai, Y. Gao, F. Liu, Transmed: Transformers advance multi-modal medical image classification, Diagnostics 11 (8) (2021) 1384.
[32] S.V. Bhalerao, R.B. Pachori, Imagined Speech-EEG Detection Using Multivariate Swarm Sparse Decomposition-Based Joint Time-Frequency Analysis for Intuitive BCI, Authorea, 2024, Authorea Preprints.
[33] S.V. Bhalerao, R.B. Pachori, Clustering sparse swarm decomposition for automated recognition of upper limb movements from non-homogeneous cross-channel EEG signals, IEEE Sens. Lett. (2023).
[34] H. Zhang, J.S. Polson, K. Nael, N. Salamon, B. Yoo, S. El-Saden, F. Scalzo, W. Speier, C.W. Arnold, Intra-domain task-adaptive transfer learning to determine acute ischemic stroke onset time, Comput. Med. Imaging Graph. 90 (2021) 101926.
[35] J.S. Polson, H. Zhang, K. Nael, N. Salamon, B.Y. Yoo, S. El-Saden, S. Starkman, N. Kim, D.-W. Kang, W.F. Speier IV, Identifying acute ischemic stroke patients within the thrombolytic treatment window using deep learning, J. Neuroimaging 32 (6) (2022) 1153–1160.
[36] H. Gao, Y. Bian, G. Cheng, H. Yu, Y. Cao, H. Zhang, J. Wang, Q. Li, Q. Yang, L. Wang, Identifying patients with acute ischemic stroke within a 6-h window for the treatment of endovascular thrombectomy using deep learning and perfusion imaging, Front. Med. 10 (2023) 1085437.

[37] C. Yoon, S. Misra, K.-J. Kim, C. Kim, B.J. Kim, Collaborative multi-modal deep learning and radiomic features for classification of strokes within 6 h, Expert Syst. Appl. 228 (2023) 120473.
[38] J. Sun, Y. Liu, Y. Xi, G. Coatrieux, J.-L. Coatrieux, X. Ji, L. Jiang, Y. Chen, Multi-grained contrastive representation learning for label-efficient lesion segmentation and onset time classification of acute ischemic stroke, Med. Image Anal. (2024) 103250.
[39] M.R. Hernandez Petzsche, E. de la Rosa, U. Hanning, R. Wiest, W. Valenzuela, M. Reyes, M. Meyer, S.-L. Liew, F. Kofler, I. Ezhov, ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset, Sci. Data 9 (1) (2022) 762.
[40] M. Consortium, MONAI: Medical open network for AI, 5525502, 2020, Online at https://doi.org/10.5281/zenodo.
[41] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, pp. 234–241.
[42] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[43] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, in: Proc. Icml, Vol. 30, Atlanta, Georgia, USA, p. 3.
[44] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9729–9738.
[45] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in pytorch, 2017.
[46] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
[47] N. Ma, X. Zhang, H.-T. Zheng, J. Sun, Shufflenet v2: Practical guidelines for efficient cnn architecture design, in: Proceedings of the European Conference on Computer Vision, ECCV, pp. 116–131.
[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.
[49] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
[50] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
[51] A. Hatamizadeh, H. Yin, J. Kautz, P. Molchanov, Global context vision transformers, 2022, arXiv preprint arXiv:2206.09959.
[52] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, Y. Wang, Transformer in transformer, Adv. Neural Inf. Process. Syst. 34 (2021) 15908–15919.
[53] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022.
[54] Y. Li, C.-Y. Wu, H. Fan, K. Mangalam, B. Xiong, J. Malik, C. Feichtenhofer, Mvitv2: Improved multiscale vision transformers for classification and detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4804–4814.
[55] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 574–584.
[56] Y. Peng, J. Sun, The multimodal MRI brain tumor segmentation based on AD-Net, Biomed. Signal Process. Control 80 (2023) 104336.
[57] S. Mehta, M. Rastegari, Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer, 2021, arXiv preprint arXiv:2110.02178.